

Lecture Notes of Signals, Systems and Control

Duarte Valério

November 21, 2023

Overview of contents

I	Modelling	3
II	Systems theory	105
III	Sensors and actuators	209
IV	Control systems	265
V	Practical implementation of control systems	385
VI	System identification	435
VII	Fractional order systems	481
VIII	Stochastic systems	527
	Epilogue	613

Acknowledgements

Some of the exercises in these lecture notes have been taken or adapted from collections of exercises or exams conceived by colleagues to whom I am indebted for authorising their reproduction here. In particular, I thank professors Alexandra Moutinho, Carlos Cardeira, Jorge Martins, José Raul Azinheira, Mário Ramalho, Miguel Ayala Botto, Paulo Oliveira, Pedro Lourtie and Rodrigo Bernardo for this.

Contents

1	The name of the game	1
I	Modelling	3
2	The Laplace transform	7
	2.1. Definition, 7. — 2.2. Finding Laplace transforms, 8. — 2.3. Finding inverse Laplace transforms, 10. — 2.4. Important properties: derivatives and integrals, 14. — 2.5. What do we need this for?, 15. — 2.6. More important properties: initial and final values, convolution, 17. — 2.7. The Fourier transform, 19. — Glossary, 21. — Exercises, 22.	
3	Examples of mechatronic systems and signals	23
	3.1. Systems, 23. — 3.2. Signals, 28. — 3.3. Models, 35. — Glossary, 36. — Exercises, 38.	
4	Modelling mechanical systems	41
	4.1. Modelling the translation movement, 41. — 4.2. Simulating transfer functions in MATLAB, 47. — 4.3. Modelling the rotational movement, 49. — 4.4. Energy, effort and flow, 50. — 4.5. Other components, 51. — Glossary, 56. — Exercises, 56.	
5	Modelling electrical systems	61
	5.1. Passive components, 61. — 5.2. Energy, effort and flow, 64. — 5.3. The operational amplifier (OpAmp), an active component, 68. — 5.4. Other components, 73. — Glossary, 74. — Exercises, 75.	
6	Modelling fluidic systems	79
	6.1. Energy, effort and flow, 79. — 6.2. Basic components of a fluidic system, 80. — 6.3. Other components, 83. — Glossary, 83. — Exercises, 85.	
7	Modelling thermal systems	87
	7.1. Energy, effort and flow, 87. — 7.2. Basic components of a thermal system, 87. — Glossary, 91. — Exercises, 91.	
8	Modelling interconnected and non-linear systems	93
	8.1. Energy, effort and flow, 93. — 8.2. System interconnection, 95. — 8.3. Dealing with non-linearities, 97. — Glossary, 100. — Exercises, 100.	
II	Systems theory	105
9	Transfer functions and block diagrams	109
	9.1. More on transfer functions, 109. — 9.2. Block diagrams, 113. — 9.3. Control in open-loop and in closed-loop, 122. — Glossary, 124. — Exercises, 125.	
10	Time and frequency responses	129
	10.1. Time responses: steps and impulses as inputs, 129. — 10.2. Steady-state response and transient response, 134. — 10.3. Stability, 138. — 10.4. Time responses: periodic inputs, 143. — 10.5. Frequency responses and the Bode diagram, 146. — Glossary, 155. — Exercises, 156.	

11 Finding time and frequency responses	161
11.1. Time and frequency responses of a pole at the origin, 161. — 11.2. Time and frequency responses of a first-order system without zeros, 163. — 11.3. Time and frequency responses of a second-order system without zeros, 168. — 11.4. Systems with more zeros and poles: frequency responses, 181. — 11.5. Systems with more zeros and poles: stability, 190. — 11.6. Systems with more zeros and poles: time responses, 194. — Glossary, 200. — Exercises, 201.	
III Sensors and actuators	209
12 Measuring chains and actuation chains	213
12.1. What are measuring chains and actuation chains, 214. — 12.2. Filters, 216. — 12.3. Bandwidth, 222. — 12.4. Signal characteristics, 223. — 12.5. Op-amp implementation of signal conditioning, 232. — Glossary, 234. — Exercises, 235.	
13 Sensors	239
13.1. Position, proximity, velocity and acceleration sensors, 240. — 13.2. Working principles of position sensors, 242. — 13.3. Working principles of velocity and acceleration sensors, 252. — 13.4. Sensors for force, binary, pressure and level, 254. — 13.5. Sensors for flow and pressure in flows, 254. — 13.6. Sensors for temperature and luminosity, 254. — 13.7. Sensors for pH and concentration, 254. — Glossary, 254. — Exercises, 254.	
14 Actuators	259
14.1. Generalities about electric motors, 259. — 14.2. DC motors, 259. — 14.3. AC motors, 259. — 14.4. Generalities about pneumatic and hydraulic actuators, 259. — 14.5. Pneumatic and hydraulic compressors and motors, 259. — 14.6. Cylinders and valves, 259. — 14.7. Dimensioning of pneumatic and hydraulic circuits, 260. — Glossary, 260. — Exercises, 260.	
IV Control systems	265
15 Control strategies and controller structures	269
15.1. Open loop control, 269. — 15.2. Closed loop control, 272. — 15.3. Design of open loop controllers, 276. — 15.4. Closed loop controllers, 276. — Glossary, 281. — Exercises, 282.	
16 Root locus	283
16.1. Simple examples, 283. — 16.2. Rules for the root locus, 288. — 16.3. Proofs of rules for the root locus, 294. — 16.4. Finding desired poles from specifications, 299. — Glossary, 302. — Exercises, 302.	
17 The Nyquist stability criterion	305
17.1. The polar diagram, 305. — 17.2. The Nyquist diagram when there are no poles on the imaginary axis, 307. — 17.3. The Nyquist criterion, 309. — 17.4. The Nyquist diagram when there are poles on the imaginary axis, 314. — Glossary, 317. — Exercises, 318.	
18 Stability margins	321
18.1. Stability margins in the Nyquist diagram, 321. — 18.2. Stability margins in the Bode diagram, 326. — 18.3. Stability margins with opposite signs, 329. — 18.4. Stability margins and the root locus diagram, 330. — Glossary, 332. — Exercises, 333.	
19 The Nichols diagram	335
19.1. Examples, 335. — 19.2. Stability margins, 336. — 19.3. The N and M curves, 336. — Glossary, 339. — Exercises, 340.	
20 Steady-state errors	341
20.1. Steps as references, 342. — 20.2. Ramps as references, 343. — 20.3. Parabolas as references, 346. — 20.4. Summing up the results, 347. — Glossary, 347. — Exercises, 348.	

21 Design of PID controllers	351
21.1. Root locus and Bode diagrams, 351. — 21.2. Tuning rules, 353. — 21.3. PID design by pole-placement, 359. — Glossary, 362. — Exercises, 362.	
22 Design of lead-lag controllers	367
Glossary, 367. — Exercises, 368.	
23 Internal Model Control	373
23.1. IMC as a variation of closed-loop control, 373. — 23.2. IMC as a design method for closed-loop controllers, 376. — Glossary, 376. — Exercises, 377.	
24 Delay systems	379
24.1. Pure delays, 379. — 24.2. Padé approximations, 380. — 24.3. Smith predictor, 380. — 24.4. Control of systems similar to delay systems, 380. — Glossary, 381. — Exercises, 381.	
V Practical implementation of control systems	385
25 Digital signals and systems	389
25.1. The \mathcal{Z} transform, 391. — 25.2. Discrete transfer functions, 394. — 25.3. Zero order hold, 395. — 25.4. Choosing the sampling time, 396. — 25.5. Stability and causality of discrete transfer functions, 399. — 25.6. Primary and complementary strips in s , 401. — Glossary, 402. — Exercises, 402.	
26 Digital approximations of continuous systems	405
26.1. Using the \mathcal{Z} transform, 405. — 26.2. Mapping poles and zeros, 405. — 26.3. Backward first order approximation, 405. — 26.4. Forward first order approximation, 405. — 26.5. The Tustin approximation, 405. — 26.6. Approximating controllers, 405. — Glossary, 406. — Exercises, 406.	
27 Study and control of digital systems	407
27.1. Block diagrams, 407. — 27.2. Pole placement, 407. — 27.3. Steady state errors, 407. — 27.4. Root locus, 408. — 27.5. Jury and Routh-Hurwitz criteria, 408. — 27.6. Frequency responses, 408. — 27.7. Studying stability from frequency responses, 408. — Glossary, 408. — Exercises, 409.	
28 Non-linearities in control systems	413
28.1. Non-linearities and responses in time, 413. — 28.2. Describing function, 418. — 28.3. Predicting limit cycles, 424. — Glossary, 429. — Exercises, 429.	
29 Other aspects of controller design and implementation	433
Glossary, 433. — Exercises, 434.	
VI System identification	435
30 Overview and general issues	439
30.1. Types of data, types of identification methods, and types of models, 439. — 30.2. Comparing model performance, 441. — 30.3. Noise, 444. — 30.4. Interpolation vs. curve fitting, 446. — Glossary, 449. — Exercises, 449.	
31 Identification from time responses	451
31.1. Identification of the order of the model, 451. — 31.2. Identification from analytical characteristics of the response, 452. — 31.3. Identification by minimisation of error, 453. — 31.4. Deconvolution, 456. — 31.5. Real time identification, 459. — 31.6. Digital model validation, 462. — Glossary, 463. — Exercises, 463.	
32 Identification from frequency responses	465
32.1. Finding frequency response data, 465. — 32.2. Identification from the Bode diagram, 467. — 32.3. Levy's method, 467. — 32.4. Matsuda's method, 471. — 32.5. Oustaloup's method, 473. — Glossary, 474. — Exercises, 474.	

33 Identification of non-linearities	477
33.1. Identifying the presence of a non-linearity, 477. — 33.2. Identification from a time response, 477. — 33.3. Identification from a limit cycle, 478. — 33.4. Identification of a pure delay, 479. — Exercises, 480.	
VII Fractional order systems	481
34 Fractional order systems and their frequency responses	485
34.1. Frequency responses that require fractional order systems, 485. — 34.2. Frequency responses of fractional order systems, 487. — 34.3. Identification from the Bode diagram, 492. — 34.4. Levy's method extended, 493. — Glossary, 496. — Exercises, 496.	
35 Fractional order derivatives	501
35.1. Gamma function, 501. — 35.2. Two apparently simple examples, 505. — 35.3. The Grünwald-Letnikov definition of fractional derivatives, 507. — 35.4. The Riemann-Liouville definition of fractional derivatives, 509. — 35.5. Properties of fractional derivatives, 512. — 35.6. Applications of fractional derivatives, 513. — Glossary, 514. — Exercises, 514.	
36 Time responses of fractional order systems	517
36.1. The Mittag-Leffler function, 517. — 36.2. Time responses of simple fractional order transfer functions, 519. — 36.3. Stability of fractional systems, 520. — 36.4. Identification from time responses, 522. — 36.5. Final comments about models of fractional order, 525. — Exercises, 525.	
VIII Stochastic systems	527
37 Stochastic processes and systems	531
37.1. Stochastic processes, 531. — 37.2. Characterisation of stochastic processes, 534. — 37.3. Relations between stochastic processes, 545. — 37.4. Operations with stochastic processes, 550. — Glossary, 552. — Exercises, 553.	
38 Spectral density	555
38.1. The bilateral Fourier transform, 555. — 38.2. Definition and properties of the spectral density, 558. — 38.3. Numerical computation of the PSD and CSD, 563. — 38.4. White noise, 568. — Glossary, 569. — Exercises, 569.	
39 Identification of continuous stochastic models	571
39.1. Identification in time, 571. — 39.2. Identification in frequency, 575. — Glossary, 580. — Exercises, 580.	
40 Filter design	581
40.1. Wiener filters, 581. — 40.2. Whitening filters, 584. — Glossary, 585. — Exercises, 585.	
41 Digital stochastic models	587
41.1. Types of digital stochastic models, 587. — 41.2. Autocorrelation of a MA, 590. — 41.3. Partial autocorrelation of an AR, 592. — 41.4. Finding the orders of an ARMA, 596. — Glossary, 596. — Exercises, 597.	
42 Control of stochastic systems	599
42.1. Minimum variance control, 599. — 42.2. Pole-assignment control, 606. — 42.3. Control design for time-varying references, 608. — 42.4. Adaptive control, 609. — Glossary, 609. — Exercises, 610.	
Epilogue	613
43 What next?	613
43.1. Discrete events and automation, 613. — 43.2. State-space representations of systems continuous in time, 613. — 43.3. State-space representations of systems discrete in time, 613. — 43.4. MIMO systems and MIMO control, 613. — 43.5. Optimal control, 614. — 43.6. Fractional order control, 614. — 43.7. Other areas, 614. — Glossary, 614.	

Chapter 1

The name of the game

Signum est quod et se ipsum sensui et praeter se aliquid animo ostendit.

Saint AUGUSTINE of Hippo (354 — †430), *De dialectica* (c. 387), V

System is the part of the Universe we want to study.

System

A **signal** is a function of time or space that conveys information about a system.

Signal

A **control system** is, in simple terms, a system designed to behave as the designer desires.

Control system

There are systems of many types: biological systems, economic systems, chemical systems, social systems. . . These lecture notes are concerned in particular with **mechatronic systems**, i.e. those combining both mechanical and electronic components.

Mechatronic system

Example 1.1. A Wave Energy Converter (WEC) is a mechatronic system that extracts the energy of sea waves, usually to produce electricity. The power injected to the electric grid is a signal that depends on time. The elevation of the sea waves is a signal that depends on both time and space. Figure 1.1 illustrates these signals. \square

We will study the following subjects:

Part I addresses the development of mathematical models to describe the behaviour of mechatronic systems.

Part II uses the models from Part I to study how systems behave.

Part III presents the technology used to measure signals and to control mechatronic systems.

Part IV shows how controllers can be designed to make control systems behave as desired.

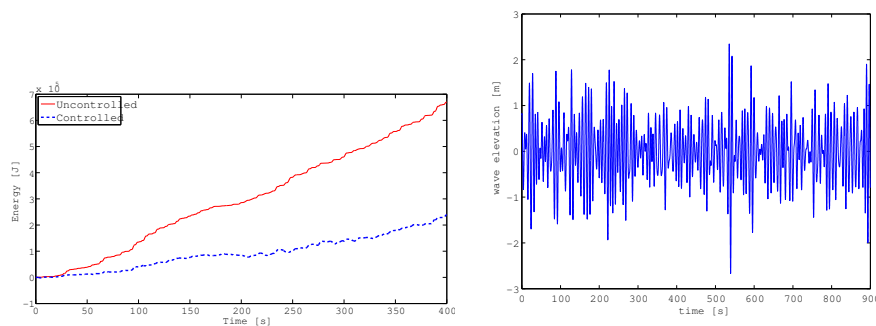


Figure 1.1: Left: electrical energy produced by a WEC as a function of time. Right: wave elevation as a function of time at a given point; at a different location, the wave elevation will be different.

Part V is about practical aspects of the implementation of control systems.

Part VI explores system identification, i.e. ways of finding models from experimental data.

Part VII studies systems described by models with derivatives and integrals with orders that are not integer numbers.

Part VIII covers systems with outputs that are not deterministic.

Chapter 43 concludes these lecture notes with an overview of related subjects.

At IST, ULisboa, these subjects are covered in the following courses:

- **Signals and Mechatronic Systems** covers Parts I, II and III.
- **Control Systems** covers Parts IV and V.
- **System Identification** covers Parts VI, VII and VIII.

MATLAB
SIMULINK

Octave

Scilab

We will often use a software called MATLAB, as well as MATLAB's graphical environment for working with block diagrams SIMULINK. MATLAB is not a free software. On the other hand, some of the functionalities of MATLAB can be supplied by free software such as Octave (which you can install from <https://www.gnu.org/software/octave> or run online from <https://octave-online.net/>) or Scilab (which you can install from <https://www.scilab.org/>). Notice that several functionalities of MATLAB we will need are missing from Octave, which also lacks anything parallel to SIMULINK. Scilab is more complete, and has Xcos, which is similar to SIMULINK; but is slightly less compatible with Matlab.

In what follows it is presumed that you are acquainted with the most basic features of MATLAB (or Octave, or Scilab), which you can learn with the "Getting started with MATLAB" tutorials in the program's "Documentation Center".

Part I

Modelling

La ciencia jamás podrá descubrir todos los secretos de la naturaleza,
ya que la ciencia la hacen los hombres y éstos son parte de ella.

J. Juan ROSALES García (1967 — ...), pers. comm., Saint Petersburg, 2001,
qtd. in *Ecuaciones diferenciales ordinarias* (2009, auth. Juan Rosales, Manuel
Guía Calderón)

In this part of the lecture notes:

- Chapter 2 presents very handy mathematical tools for the resolution of differential equations, which we will need repeatedly in subsequent chapters.
- Chapter 3 gives examples of mechatronic systems and signals, and the basic notions related thereto.
- Chapter 4 addresses the modelling of mechanical systems.
- Chapter 5 addresses the modelling of electrical systems.
- Chapter 6 addresses the modelling of fluidic systems, a particular type of mechanical systems.
- Chapter 7 addresses the modelling of thermal systems.
- Chapter 8 addresses the modelling of systems combining several of the components studied in chapters 4 to 7, as well as of systems with nonlinear models.

Here is what you need to know beforehand to follow these chapters:

- Differential and integral calculus, up to the usual level of freshman courses on Calculus;
- Kinematics, dynamics, electrical circuits, fluid mechanics, and heat transfer, up to the usual level of freshman courses on Physics, or at least the level of secondary education.

Chapter 2

The Laplace transform

“And, if we transmit through a wormhole, the person is always reconstituted at the other end. We can count on that happening, too.”

There was a pause.

Stern frowned.

“Wait a minute,” he said. “Are you saying that when you transmit, the person is being reconstituted by another universe?”

“In effect, yes. I mean, it has to be. We can’t very well reconstitute them, because we’re not there. We’re in this universe.”

Michael CRICHTON (1942 — †2008), *Timeline* (1999), Black rock

The **Laplace transform** is a very important tool for the resolution of differential equations. In this chapter we will study its definition, its properties, its application to differential equations (which is the reason we are studying this subject), and the related Fourier transform, that we will also need. *Laplace transform*

2.1 Definition

Definition 2.1. Let $t \in \mathbb{R}$ be a real variable, and $f(t) \in \mathbb{R}$ a real-valued function. The Laplace transform of function f , denoted by $\mathcal{L}[f(t)]$ or by $F(s)$, is a complex-valued function $F(s) \in \mathbb{C}$ of complex variable $s \in \mathbb{C}$, given by

$$\mathcal{L}[f(t)] = \int_0^{+\infty} f(t)e^{-st} dt \quad \square \quad (2.1)$$

Remark 2.1. Strictly speaking, operation \mathcal{L} is the Laplace transformation, and the result of applying \mathcal{L} to a function gives us its Laplace transform. But it is common to call the operation itself Laplace transform as well. \square

Remark 2.2. In (2.1), function $f(t)$ only has to be defined for $t \geq 0$. This would not be so if we were using the **bilateral Laplace transform**, which is an alternative definition given by *Bilateral Laplace transform*

$$\mathcal{L}[f(t)] = \int_{-\infty}^{+\infty} f(t)e^{-st} dt \quad (2.2)$$

This bilateral Laplace transform is seldom used; we will use (2.1) instead, as is common, and will need (2.2) only in Chapter 37. The price to pay for being able to work with functions defined in \mathbb{R}^+ only will be addressed below in section 2.4. \square

Remark 2.3. The Laplace transform is part of a group of transforms known as integral transforms, given by

$$\mathcal{T}[f(t)] = \int_0^{+\infty} f(t)K(s, t) dt \quad (2.3)$$

where \mathcal{T} is a generic transform and $K(s, t)$ is a function called kernel. In the case of the Laplace transform, the kernel is $K(s, t) = e^{-st}$. \square

The Laplace transform of function $f(t)$ will only exist if the improper integral in (2.1) converges. This will happen in one of two cases:

Existence of the Laplace transform

- If $f(t)$ is bounded in its domain \mathbb{R}^+ , the integrand $f(t)e^{-st}$ will obviously tend to 0 as $t \rightarrow +\infty$.
- If $f(t)$ tends to infinity as $t \rightarrow +\infty$, but does so slower than e^{-st} tends to 0, the integrand will still tend to 0. More rigorously, $f(t)$ must be of exponential order, i.e. there must be positive real constants $M, c \in \mathbb{R}$ such that

$$|f(t)| \leq M e^{ct}, \quad 0 \leq t \leq \infty. \quad (2.4)$$

Otherwise, the integrand of (2.1) does not tend to 0 and it is obvious that the improper integral will be infinite. For complete rigour we also have to require $f(t)$ to be piecewise continuous for $F(s)$ to exist, otherwise the Riemann integral would not exist.

Remark 2.4. In fact (2.1) may converge only for some values of s , and thus have a region of convergence which is smaller than \mathbb{C} ; but then it can be analytically extended to the rest of the complex plane. This is a question we will not worry about. \square

2.2 Finding Laplace transforms

Heaviside function

Example 2.1. Let $f(t)$ be function

$$H(t) = \begin{cases} 1, & \text{if } t \geq 0 \\ 0, & \text{if } t < 0 \end{cases}, \quad (2.5)$$

$\mathcal{L}[H(t)]$

known as the Heaviside function. Then

$$\mathcal{L}[H(t)] = \int_0^{+\infty} 1 \times e^{-st} dt = \left[\frac{e^{-st}}{-s} \right]_0^{+\infty} = \frac{e^{-\infty}}{-s} - \frac{e^0}{-s} = \frac{1}{s}. \quad \square \quad (2.6)$$

$\mathcal{L}[e^{-at}]$

Example 2.2. Let $f(t)$ be a negative exponential, $f(t) = e^{-at}$. Then

$$\begin{aligned} \mathcal{L}[e^{-at}] &= \int_0^{+\infty} e^{-at} e^{-st} dt \\ &= \left[\frac{e^{-(a+s)t}}{-a-s} \right]_0^{+\infty} = -\frac{e^{-\infty}}{s+a} - \left(-\frac{e^0}{s+a} \right) = \frac{1}{s+a}. \quad \square \end{aligned} \quad (2.7)$$

While Laplace transforms can be found from definition as in the two examples above, in practice they are found from tables, such as the one in Table 2.1.

To use with profit Laplace transform tables, it is necessary to prove first the following result.

\mathcal{L} is linear

Theorem 2.1. The Laplace transform is a linear operator:

$$\mathcal{L}[k f(t)] = k F(s), \quad k \in \mathbb{R} \quad (2.8)$$

$$\mathcal{L}[f(t) + g(t)] = F(s) + G(s) \quad (2.9)$$

Proof. Both (2.8) and (2.9) are proved from the linearity of the integration operator:

$$\mathcal{L}[k f(t)] = \int_0^{+\infty} k f(t) e^{-st} dt = k \int_0^{+\infty} f(t) e^{-st} dt = k F(s) \quad (2.10)$$

$$\begin{aligned} \mathcal{L}[f(t) + g(t)] &= \int_0^{+\infty} (f(t) + g(t)) e^{-st} dt \\ &= \int_0^{+\infty} f(t) e^{-st} dt + \int_0^{+\infty} g(t) e^{-st} dt = F(s) + G(s) \quad \square \end{aligned} \quad (2.11)$$

Example 2.3. The Laplace transform of $f(t) = 5t$ is obtained from line 3 of Table 2.1 together with (2.8):

$$\mathcal{L}[5t] = 5\mathcal{L}[t] = \frac{5}{s^2} \quad \square \quad (2.12)$$

Table 2.1: Table of Laplace transforms

	$x(t)$	$X(s)$
1	$\delta(t)$	1
2	$H(t)$	$\frac{1}{s}$
3	t	$\frac{1}{s^2}$
4	t^2	$\frac{2}{s^3}$
5	e^{-at}	$\frac{1}{s+a}$
6	$1 - e^{-at}$	$\frac{a}{s(s+a)}$
7	te^{-at}	$\frac{1}{(s+a)^2}$
8	$t^n e^{-at}, n \in \mathbb{N}$	$\frac{n!}{(s+a)^{n+1}}$
9	$\sin(\omega t)$	$\frac{\omega}{s^2 + \omega^2}$
10	$\cos(\omega t)$	$\frac{s}{s^2 + \omega^2}$
11	$e^{-at} \sin(\omega t)$	$\frac{\omega}{(s+a)^2 + \omega^2}$
12	$e^{-at} \cos(\omega t)$	$\frac{s+a}{(s+a)^2 + \omega^2}$
13	$\frac{1}{b-a}(e^{-at} - e^{-bt})$	$\frac{1}{(s+a)(s+b)}$
14	$\frac{1}{ab} \left(1 + \frac{1}{a-b}(be^{-at} - ae^{-bt}) \right)$	$\frac{1}{s(s+a)(s+b)}$
15	$\frac{\omega_n}{\Xi} e^{-\xi\omega_n t} \sin(\omega_n \Xi t)$	$\frac{\omega_n^2}{s^2 + 2\xi\omega_n s + \omega_n^2}$
16	$-\frac{1}{\Xi} e^{-\xi\omega_n t} \sin(\omega_n \Xi t - \phi)$	$\frac{s}{s^2 + 2\xi\omega_n s + \omega_n^2}$
17	$1 - \frac{1}{\Xi} e^{-\xi\omega_n t} \sin(\omega_n \Xi t + \phi)$	$\frac{\omega_n^2}{s(s^2 + 2\xi\omega_n s + \omega_n^2)}$

In this table: $\Xi = \sqrt{1 - \xi^2}$; $\phi = \arctan \frac{\Xi}{\xi}$

Example 2.4. The Laplace transform of $f(t) = 1 - (1+t)e^{-3t}$ is obtained from lines 6 and 7 of Table 2.1 together with (2.9):

$$\begin{aligned}\mathcal{L}[f(t)] &= \mathcal{L}[1 - e^{-3t} - te^{-3t}] = \mathcal{L}[1 - e^{-3t}] - \mathcal{L}[te^{-3t}] \\ &= \frac{3}{s(s+3)} - \frac{1}{(s+3)^2} = \frac{3s+3^2-s}{s(s+3)^2} = \frac{2s+9}{s^3+6s^2+9s} \quad \square\end{aligned}\quad (2.13)$$

2.3 Finding inverse Laplace transforms

Inverse Laplace transform Laplace transform tables can also be used to find inverse Laplace transforms, i.e. finding the $f(t)$ corresponding to a given $F(s) = \mathcal{L}[f(t)]$. This operation is denoted by $f(t) = \mathcal{L}^{-1}[F(s)]$.

Example 2.5. The inverse Laplace transform of $F(s) = \frac{10}{s+10}$ is obtained from line 5 of Table 2.1 together with (2.8):

$$\mathcal{L}^{-1}\left[\frac{10}{s+10}\right] = 10\mathcal{L}^{-1}\left[\frac{1}{s+10}\right] = 10e^{-10t} \quad \square\quad (2.14)$$

Partial fraction expansion **Example 2.6.** The inverse Laplace transform of $F(s) = \frac{s+2}{s^2+13s+30}$ is obtained from line 5 of Table 2.1 together with (2.8). But for that it is necessary to develop $F(s)$ in a **partial fraction expansion**. First we find the roots of the polynomial in the denominator, which are -3 and -10 . So $s^2+13s+30 = (s+3)(s+10)$, and we can write

$$\frac{s+2}{s^2+13s+30} = \frac{A}{s+3} + \frac{B}{s+10} \quad (2.15)$$

where A and B still have to be determined:

$$\frac{A}{s+3} + \frac{B}{s+10} = \frac{As+10A+Bs+3B}{(s+3)(s+10)} = \frac{s(A+B) + (10A+3B)}{s^2+13s+30} \quad (2.16)$$

Obviously we want that

$$\begin{cases} A+B=1 \\ 10A+3B=2 \end{cases} \Leftrightarrow \begin{cases} B=1-A \\ 10A+3-3A=2 \end{cases} \Leftrightarrow \begin{cases} B=\frac{8}{7} \\ A=-\frac{1}{7} \end{cases} \quad (2.17)$$

So $\frac{s+2}{s^2+13s+30} = \frac{-\frac{1}{7}}{s+3} + \frac{\frac{8}{7}}{s+10}$, and finally

$$\begin{aligned}\mathcal{L}^{-1}\left[\frac{s+2}{s^2+13s+30}\right] &= \mathcal{L}^{-1}\left[\frac{-\frac{1}{7}}{s+3} + \frac{\frac{8}{7}}{s+10}\right] \\ &= \mathcal{L}^{-1}\left[\frac{-\frac{1}{7}}{s+3}\right] + \mathcal{L}^{-1}\left[\frac{\frac{8}{7}}{s+10}\right] = -\frac{1}{7}e^{-3t} + \frac{8}{7}e^{-10t} \quad \square\end{aligned}\quad (2.18)$$

Remark 2.5. Notice that the result in line 13 of Table 2.1 can be obtained from line 5 also using a partial fraction expansion:

$$\frac{1}{(s+a)(s+b)} = \frac{A}{s+a} + \frac{B}{s+b} = \frac{As+Ab+Bs+aB}{(s+a)(s+b)} = \frac{s(A+B) + (Ab+aB)}{(s+a)(s+b)} \quad (2.19)$$

We want

$$\begin{cases} A+B=0 \\ Ab+aB=1 \end{cases} \Leftrightarrow \begin{cases} B=-A \\ Ab-aA=1 \end{cases} \Leftrightarrow \begin{cases} B=\frac{-1}{b-a} \\ A=\frac{1}{b-a} \end{cases} \quad (2.20)$$

and thus

$$\begin{aligned}\mathcal{L}^{-1}\left[\frac{1}{(s+a)(s+b)}\right] &= \mathcal{L}^{-1}\left[\frac{\frac{1}{b-a}}{s+a}\right] + \mathcal{L}^{-1}\left[\frac{\frac{-1}{b-a}}{s+b}\right] \\ &= \frac{1}{b-a}e^{-at} + \frac{-1}{b-a}e^{-bt} = \frac{1}{b-a}(e^{-at} - e^{-bt}) \quad \square\end{aligned}\quad (2.21)$$

al fraction expansion
complex roots

Example 2.7. The inverse Laplace transform of $F(s) = \frac{4s^2+13s-2}{(s^2+2s+2)(s+4)}$ is obtained from lines 5, 15 and 16 of Table 2.1 together with 2.8 and 2.9. The transforms in lines 14 and 15 are used because the roots of $4s^2 + 13s - 2$ are complex and not real ($-1 \pm j$, to be precise). So we will leave that second order term intact and we make

$$\begin{aligned} \frac{4s^2 + 13s - 2}{(s^2 + 2s + 2)(s + 4)} &= \frac{As + B}{s^2 + 2s + 2} + \frac{C}{s + 4} = \frac{As^2 + 4As + Bs + 4B + Cs^2 + 2Cs + 2C}{(s^2 + 2s + 2)(s + 4)} \\ &= \frac{s^2(A + C) + s(4A + B + 2C) + (4B + 2C)}{(s^2 + 2s + 2)(s + 4)} \end{aligned} \quad (2.22)$$

Hence

$$\begin{cases} A + C = 4 \\ 4A + B + 2C = 13 \\ 4B + 2C = -2 \end{cases} \Leftrightarrow \begin{cases} C = 4 - A \\ 4A + B + 8 - 2A = 13 \\ 4A - 3B = 15 \end{cases} \Leftrightarrow \begin{cases} C = 4 - A \\ 2A + B = 5 \\ 4A - 3B = 15 \end{cases} \Leftrightarrow \begin{cases} C = 1 \\ A = 3 \\ B = -1 \end{cases} \quad (2.23)$$

Finally,

$$\begin{aligned} \mathcal{L}^{-1} \left[\frac{4s^2 + 13s - 2}{(s^2 + 2s + 2)(s + 4)} \right] &= \mathcal{L}^{-1} \left[\frac{3s - 1}{s^2 + 2s + 2} + \frac{1}{s + 4} \right] \\ &= 3\mathcal{L}^{-1} \left[\frac{s}{s^2 + 2s + 2} \right] - \frac{1}{2}\mathcal{L}^{-1} \left[\frac{2}{s^2 + 2s + 2} \right] + \mathcal{L}^{-1} \left[\frac{1}{s + 4} \right] \end{aligned} \quad (2.24)$$

and since for the first two terms we have

$$\omega = \sqrt{2} \quad (2.25)$$

$$\xi\omega = 1 \quad (2.26)$$

$$\xi = \frac{1}{\sqrt{2}} \quad (2.27)$$

$$\Xi = \sqrt{1 - \frac{1}{2}} = \frac{1}{\sqrt{2}} \quad (2.28)$$

$$\omega\Xi = 1 \quad (2.29)$$

$$\varphi = \arctan \frac{\frac{1}{\sqrt{2}}}{\frac{1}{\sqrt{2}}} = \frac{\pi}{4} \quad (2.30)$$

we arrive at

$$\begin{aligned} \mathcal{L}^{-1} \left[\frac{4s^2 + 13s - 2}{(s^2 + 2s + 2)(s + 4)} \right] &= -3\sqrt{2}e^{-t} \sin \left(t - \frac{\pi}{4} \right) - \frac{1}{2}2e^{-t} \sin(t) + e^{-4t} \\ &= e^{-4t} + e^{-t} \left[-3\sqrt{2} \left(\sin t \cos \frac{\pi}{4} - \cos t \sin \frac{\pi}{4} \right) - \sin t \right] \\ &= e^{-4t} + e^{-t} \left[-3\sqrt{2} \left(\sin t \frac{1}{\sqrt{2}} - \cos t \frac{1}{\sqrt{2}} \right) - \sin t \right] \\ &= e^{-4t} + e^{-t} (-4 \sin t + 3 \cos t) \quad \square \end{aligned} \quad (2.31)$$

Remark 2.6. If in the example above we had decided to expand the second order term and use only line 5 of Table 2.1, we would have arrived at the very same result, albeit with more lengthy and tedious calculations involving complex numbers. We would have to separate $\frac{3s-1}{s^2+2s+2}$ in two as follows:

$$\begin{aligned} \frac{3s - 1}{s^2 + 2s + 2} &= \frac{A + Bj}{s + 1 + j} + \frac{C + Dj}{s + 1 - j} \\ &= \frac{As + A - Aj + Bjs + Bj + B + Cs + C + Cj + Djs + Dj - D}{s^2 + s - js + s + 1 - j + js + j + 1} \\ &= \frac{s(A + C) + js(B + D) + (A + B + C - D) + j(-A + B + C + D)}{s^2 + 2s + 2} \end{aligned} \quad (2.32)$$

Then

$$\begin{cases} A + C = 3 \\ B + D = 0 \\ A + B + C - D = -1 \\ -A + B + C + D = 0 \end{cases} \Leftrightarrow \begin{cases} C = 3 - A \\ D = -B \\ A + B + 3 - A + B = -1 \\ -A + B + 3 - A - B = 0 \end{cases} \Leftrightarrow \begin{cases} C = \frac{3}{2} \\ D = 2 \\ B = -2 \\ A = \frac{3}{2} \end{cases} \quad (2.33)$$

Consequently

$$\begin{aligned}
\mathcal{L}^{-1} \left[\frac{4s^2 + 13s - 2}{(s^2 + 2s + 2)(s + 4)} \right] &= \mathcal{L}^{-1} \left[\frac{\frac{3}{2} - 2j}{s + 1 + j} + \frac{\frac{3}{2} + 2j}{s + 1 - j} + \frac{1}{s + 4} \right] \\
&= \left(\frac{3}{2} - 2j \right) \mathcal{L}^{-1} \left[\frac{1}{s + 1 + j} \right] + \left(\frac{3}{2} + 2j \right) \mathcal{L}^{-1} \left[\frac{1}{s + 1 - j} \right] + \mathcal{L}^{-1} \left[\frac{1}{s + 4} \right] \\
&= \left(\frac{3}{2} - 2j \right) e^{-(1+j)t} + \left(\frac{3}{2} + 2j \right) e^{-(1-j)t} + e^{-4t} \\
&= e^{-4t} + \left(\frac{3}{2} - 2j \right) e^{-t} (\cos(-t) + j \sin(-t)) + \left(\frac{3}{2} + 2j \right) e^{-t} (\cos t + j \sin t) \\
&= e^{-4t} + e^{-t} \left(\frac{3}{2} \cos t - \frac{3}{2} j \sin t - 2j \cos t - 2 \sin t + \right. \\
&\quad \left. + \frac{3}{2} \cos t + \frac{3}{2} j \sin t + 2j \cos t - 2 \sin t \right) \\
&= e^{-4t} + e^{-t} (3 \cos t - 4 \sin t)
\end{aligned} \tag{2.34}$$

Notice how all the complex terms appear in complex conjugates, so that the imaginary parts cancel out. This has to be the case, since $f(t)$ is a real-valued function. \square

Partial fraction expansion with multiple roots **Example 2.8.** The inverse Laplace transform of $F(s) = \frac{s^2 + 22s + 119}{(s + 10)^3}$ is obtained from lines 5, 7 and 8 of Table 2.1 together with (2.8) and (2.9):

$$\begin{aligned}
\frac{s^2 + 22s + 119}{(s + 10)^3} &= \frac{A}{s + 10} + \frac{B}{(s + 10)^2} + \frac{C}{(s + 10)^3} \\
&= \frac{As^2 + 20As + 100A + Bs + 10B + C}{(s + 10)^3}
\end{aligned} \tag{2.35}$$

Hence

$$\begin{cases} A = 1 \\ 20A + B = 22 \\ 100A + 10B + C = 119 \end{cases} \Leftrightarrow \begin{cases} A = 1 \\ B = 2 \\ C = -1 \end{cases} \tag{2.36}$$

Finally,

$$\begin{aligned}
\mathcal{L}^{-1} \left[\frac{s^2 + 22s + 119}{(s + 10)^3} \right] &= \mathcal{L}^{-1} \left[\frac{1}{s + 10} + \frac{2}{(s + 10)^2} + \frac{-1}{(s + 10)^3} \right] \\
&= \mathcal{L}^{-1} \left[\frac{1}{s + 10} \right] + 2\mathcal{L}^{-1} \left[\frac{2}{(s + 10)^2} \right] - \frac{1}{2} \mathcal{L}^{-1} \left[\frac{2}{(s + 10)^3} \right] \\
&= e^{-10t} + 2te^{-10t} - \frac{1}{2}t^2 e^{-10t} = e^{-10t} \left(1 + 2t - \frac{1}{2}t^2 \right) \quad \square
\end{aligned} \tag{2.37}$$

Division of polynomials

Example 2.9. The inverse Laplace transform of $F(s) = \frac{2s + 145}{s + 70}$ is obtained from lines 1 and 5 of Table 2.1, but for that it is necessary to begin by dividing the numerator of $F(s)$ by the denominator. Because the denominator is of first order, in this case polynomial division can be carried out with Ruffini's rule (otherwise a long division would be necessary):

$$\begin{array}{r|rr}
 & 2 & 145 \\
-70 & & -140 \\
\hline
 & 2 & 5
\end{array} \tag{2.38}$$

So $\frac{2s + 145}{s + 70} = 2 + \frac{5}{s + 70}$, and finally

$$\begin{aligned}
\mathcal{L}^{-1} \left[\frac{2s + 145}{s + 70} \right] &= 2\mathcal{L}^{-1} [1] + 5\mathcal{L}^{-1} \left[\frac{1}{s + 70} \right] \\
&= 2\delta(t) + e^{-70t} \quad \square
\end{aligned} \tag{2.39}$$

All polynomial operations mentioned in this sections can be performed with MATLAB using the following commands:

- `roots` finds the roots of a polynomial, represented by a vector with its coefficients (in decreasing order of the exponent);
- `conv` multiplies two polynomials, represented by two vectors as above;
- `residue` performs polynomial division and partial fraction expansion, as needed, for a rational function, given the numerator and denominator polynomials represented by two vectors as above.

Example 2.10. The roots of $s^2 + 3s + 2$ are -2 and -1 :

MATLAB's *command*
roots

```
>> roots([1 3 2])
ans =
    -2
    -1
```

□

Example 2.11. The roots of $4s^3 + 3s^2 + 2s + 1$ are -0.6058 , $-0.0721 + 0.6383j$ and $-0.0721 - 0.6383j$:

```
>> roots([4 3 2 1])
ans =
-0.6058 + 0.0000i
-0.0721 + 0.6383i
-0.0721 - 0.6383i
```

□

Example 2.12. The product of $s^2 + 2s + 3$ and $4s^3 + 5s^2 + 6s + 7$ is $4s^5 + 13s^4 + 28s^3 + 34s^2 + 32s + 21$:

MATLAB's *command* *conv*

```
>> conv([1 2 3],[4 5 6 7])
ans =
     4     13     28     34     32     21
```

□

Example 2.13. The partial fraction expansion (2.18) from Example 2.6 is obtained as

MATLAB's *command*
residue

```
>> [r,p,k] = residue([1 2],[1 13 30])
r =
    1.1429
   -0.1429
p =
   -10
    -3
k =
    []
```

Vector **r** contains the **residues** or numerators of the fractions in the partial fraction expansion. Vector **p** contains the **poles** or roots of the denominator of the original expression. Vector **k** contains (the coefficients of the polynomial which is) the integer part of the polynomial division, which in this case is 0 because the order of the denominator is higher than the order of the numerator.

Residues
Poles

The polynomials of the original rational function can be recovered feeding this function back vectors **r**, **p** and **k**:

```
>> [num,den] = residue(r,p,k)
num =
     1     2
den =
     1    13    30
```

□

Example 2.14. The partial fraction expansion (2.34) from Example 2.7 and Remark 2.6 is obtained as

```
>> [r,p,k] = residue([4 13 -2],conv([1 2 2],[1 4]))
r =
    1.0000 + 0.0000i
    1.5000 + 2.0000i
    1.5000 - 2.0000i
p =
   -4.0000 + 0.0000i
   -1.0000 + 1.0000i
   -1.0000 - 1.0000i
k =
     []
```

□

Example 2.15. The partial fraction expansion from Example 2.9 is obtained as

```
>> [r,p,k] = residue([2 145],[1 70])
r =
     5
p =
   -70
k =
     2
```

Notice how this time there is an integer part of the polynomial division, since the order of the numerator is not lower than the order of the denominator. □

Example 2.16. From

```
>> [r,p,k] = residue([1 2 3 4 5 6],[7 8 9 10])
r =
    0.1451 + 0.0000i
   -0.0276 - 0.2064i
   -0.0276 + 0.2064i
p =
   -1.1269 + 0.0000i
   -0.0080 + 1.1259i
   -0.0080 - 1.1259i
k =
    0.1429    0.1224    0.1050
```

we learn that

$$\frac{s^5 + 2s^4 + 3s^3 + 4s^2 + 5s + 6}{7s^3 + 8s^2 + 9s + 10} \quad (2.40)$$

$$= 0.1429s^2 + 0.1224s + 0.1050 + \frac{0.1451}{s + 1.1269} + \frac{-0.0276 - 0.2064j}{s + 0.0080 - 1.1259j} + \frac{-0.0276 + 0.2064j}{s + 0.0080 + 1.1259j} \quad \square$$

2.4 Important properties: derivatives and integrals

Now that we know how to find Laplace transforms, it is time to wonder why we are studying them. To answer this, first we have to establish some very important results.

\mathcal{L} of the derivative

Theorem 2.2. If $\mathcal{L}[f(t)] = F(s)$, then

$$\mathcal{L}[f'(t)] = sF(s) - f(0) \quad (2.41)$$

Proof. Apply integration by parts $\int uv' = uv - \int u'v$ to definition (2.1):

$$\begin{aligned} \mathcal{L}[f'(t)] &= \int_0^{+\infty} \overbrace{f(t)}^u \overbrace{e^{-st}}^{v'} dt \\ &= \left[f(t) \frac{e^{-st}}{-s} \right]_0^{+\infty} - \int_0^{+\infty} f'(t) \frac{e^{-st}}{-s} dt \\ &= \lim_{t \rightarrow +\infty} \left(f(t) \frac{e^{-st}}{-s} \right) - f(0) \frac{e^0}{-s} + \frac{1}{s} \int_0^{+\infty} f'(t) e^{-st} dt \end{aligned} \quad (2.42)$$

The limit has to be 0, otherwise $F(s)$ would not exist. The integral is, by definition, $\mathcal{L}[f'(t)]$. From here (2.41) is obtained rearranging terms. \square

Corollary 2.1. If $\mathcal{L}[f(t)] = F(s)$, then

$$\mathcal{L}[f''(t)] = s^2 F(s) - s f(0) - f'(0) \quad (2.43)$$

Proof. Apply (2.41) to itself:

$$\mathcal{L}[f''(t)] = s \mathcal{L}[f'(t)] - f'(0) = s(s F(s) - f(0)) - f'(0) \quad (2.44)$$

Then rearrange terms. \square

Corollary 2.2. If $\mathcal{L}[f(t)] = F(s)$, then

$$\begin{aligned} \mathcal{L}\left[\frac{d^n}{dt^n} f(t)\right] &= s^n F(s) - s^{n-1} f(0) - s^{n-2} f'(0) - \dots - \frac{d^{n-1} f(t)}{dt^{n-1}} \Big|_{t=0} \\ &= s^n F(s) - \sum_{k=1}^n s^{n-k} \frac{d^{k-1} f(t)}{dt^{k-1}} \Big|_{t=0} \end{aligned} \quad (2.45)$$

Proof. This is proved by mathematical induction. The first case is (2.41). The inductive step is proved applying (2.41) to (2.45) as follows:

$$\begin{aligned} \mathcal{L}\left[\frac{d^{n+1}}{dt^{n+1}} f(t)\right] &= s \mathcal{L}\left[\frac{d^n}{dt^n} f(t)\right] - \frac{d^n f(t)}{dt^n} \Big|_{t=0} \quad (2.46) \\ &= s \left(s^n F(s) - \sum_{k=1}^n s^{n-k} \frac{d^{k-1} f(t)}{dt^{k-1}} \Big|_{t=0} \right) - \frac{d^n f(t)}{dt^n} \Big|_{t=0} \\ &= s^{n+1} F(s) - \left(\sum_{k=1}^n s^{n-k+1} \frac{d^{k-1} f(t)}{dt^{k-1}} \Big|_{t=0} \right) - \frac{d^n f(t)}{dt^n} \Big|_{t=0} \\ &= s^{n+1} F(s) - \left(\sum_{k=1}^n s^{n+1-k} \frac{d^{k-1} f(t)}{dt^{k-1}} \Big|_{t=0} \right) - \sum_{k=n+1}^{\infty} s^{n+1-k} \frac{d^{k-1} f(t)}{dt^{k-1}} \Big|_{t=0} \quad \square \end{aligned}$$

Theorem 2.3. If $\mathcal{L}[f(t)] = F(s)$, then

\mathcal{L} of the integral

$$\mathcal{L}\left[\int_0^t f(t) dt\right] = \frac{1}{s} F(s) \quad (2.47)$$

Proof. In (2.42), make

$$f(t) = \int_0^t g(t) dt, \quad (2.48)$$

whence $f'(t) = g(t)$. Then

$$\mathcal{L}\left[\int_0^t g(t) dt\right] = -\int_0^0 g(t) dt \frac{1}{-s} + \frac{1}{s} \int_0^{+\infty} g(t) e^{-st} dt \quad (2.49)$$

The first integral is 0, the second is $\mathcal{L}[g(t)]$. \square

Remark 2.7. Notice that the Laplace transform of a derivative (2.41) involves $f(0)$, the value of the function itself at $t = 0$. This is because we are using the Laplace transform as defined by (2.1), rather than the bilateral Laplace transform (2.2).

2.5 What do we need this for?

We are now in position of answering the question above: we need Laplace transforms as a very useful tool to solve differential equations.

Use \mathcal{L} to solve differential equations

Example 2.17. Solve the following differential equation, assuming that $y(0) = 0$:

$$y'(t) + y(t) = e^{-t} \quad (2.50)$$

Apply the Laplace transform to obtain

$$\begin{aligned}\mathcal{L}[y'(t) + y(t)] &= \mathcal{L}[e^{-t}] \Leftrightarrow sY(s) + Y(s) = \frac{1}{s+1} \Leftrightarrow Y(s) = \frac{1}{(s+1)^2} \Leftrightarrow \\ &\Leftrightarrow \mathcal{L}^{-1}[Y(s)] = \mathcal{L}^{-1}\left[\frac{1}{(s+1)^2}\right] \Leftrightarrow y(t) = te^{-t}\end{aligned}\quad (2.51)$$

It is easy to verify that this is indeed the solution: $y'(t) = e^{-t} - te^{-t}$, and thus

$$y'(t) + y(t) = e^{-t} \Leftrightarrow e^{-t} - te^{-t} + te^{-t} = e^{-t}, \quad (2.52)$$

as desired. \square

Notice how the Laplace transform turned the differential equation in t into an algebraic equation in s , which is trivial to solve. All that is left is to apply the inverse Laplace transform to turn the solution in s into a solution in t .

Take care of non-null initial conditions

Initial conditions must be taken into account if they are not zero.

Example 2.18. Solve the following differential equation, assuming that $y(0) = \frac{1}{3}$ and $y'(0) = 0$:

$$y''(t) + 4y'(t) + 3y(t) = 4e^t \quad (2.53)$$

Using the Laplace transform, we get

$$\begin{aligned}s^2Y(s) - \frac{1}{3}s - 0 + 4\left(sY(s) - \frac{1}{3}\right) + 3Y(s) &= \frac{4}{s-1} \Leftrightarrow \\ \Leftrightarrow Y(s)(s^2 + 4s + 3) - \frac{s}{3} - \frac{4}{3} &= \frac{4}{s-1}\end{aligned}\quad (2.54)$$

Because $s^2 + 4s + 3 = (s+1)(s+3)$, we get

$$Y(s) = \frac{4}{(s-1)(s+1)(s+3)} + \frac{1}{3} \frac{s+4}{(s+1)(s+3)} \quad (2.55)$$

We now need two partial fraction expansions:

$$\begin{aligned}\frac{4}{(s-1)(s+1)(s+3)} + \frac{1}{3} \frac{s+4}{(s+1)(s+3)} &= \frac{A}{s-1} + \frac{B}{s+1} + \frac{C}{s+3} + \frac{1}{3} \left(\frac{D}{s+1} + \frac{E}{s+3} \right) \\ &= \frac{A(s^2 + 4s + 3) + B(s^2 + 2s - 3) + C(s^2 - 1)}{(s-1)(s+1)(s+3)} + \frac{1}{3} \left(\frac{Ds + 3D + Es + E}{(s+1)(s+3)} \right) \\ &= \frac{s^2(A+B+C) + s(4A+2B) + (3A-3B-C)}{(s-1)(s+1)(s+3)} + \frac{1}{3} \left(\frac{s(D+E) + (3D+E)}{(s+1)(s+3)} \right)\end{aligned}\quad (2.56)$$

whence

$$\begin{cases} A+B+C=0 \\ 4A+B=0 \\ 3A-3B-C=4 \end{cases} \Leftrightarrow \begin{cases} 4A-2B=4 \\ 4A+B=0 \\ C=3A-3B-4 \end{cases} \Leftrightarrow \begin{cases} 8A=4 \\ B=-2A \\ C=3A-3B-4 \end{cases} \Leftrightarrow \begin{cases} A=\frac{1}{2} \\ B=-1 \\ C=\frac{1}{2} \end{cases}\quad (2.57)$$

and

$$\begin{cases} D+E=1 \\ 3D+E=4 \end{cases} \Leftrightarrow \begin{cases} E=1-D \\ 2D=3 \end{cases} \Leftrightarrow \begin{cases} E=-\frac{1}{2} \\ D=\frac{3}{2} \end{cases}\quad (2.58)$$

Thus

$$\begin{aligned}y(t) &= \mathcal{L}^{-1}\left[\frac{\frac{1}{2}}{s-1} - \frac{1}{s+1} + \frac{\frac{1}{2}}{s+3} + \frac{1}{3}\left(\frac{\frac{3}{2}}{s+1} - \frac{\frac{1}{2}}{s+3}\right)\right] \\ &= \frac{1}{2}e^t - e^{-t} + \frac{1}{2}e^{-3t} + \frac{1}{3}\left(\frac{3}{2}e^{-t} - \frac{1}{2}e^{-3t}\right) = \frac{1}{2}e^t - \frac{1}{2}e^{-t} + \frac{1}{3}e^{-3t}\end{aligned}\quad (2.59)$$

It is easy to verify that this is indeed the solution: on the one hand,

$$y'(t) = \frac{1}{2}e^t + \frac{1}{2}e^{-t} - e^{-3t} \quad (2.60)$$

$$y''(t) = \frac{1}{2}e^t - \frac{1}{2}e^{-t} + 3e^{-3t} \quad (2.61)$$

and thus

$$\begin{aligned} y''(t) + 4y'(t) + 3y(t) &= \frac{1}{2}e^t - \frac{1}{2}e^{-t} + 3e^{-3t} + 2e^t + 2e^{-t} - 4e^{-3t} + \frac{3}{2}e^t - \frac{3}{2}e^{-t} + e^{-3t} \\ &= 4e^t \end{aligned} \quad (2.62)$$

as desired; on the other hand,

$$y(0) = \frac{1}{2} - \frac{1}{2} + \frac{1}{3} = \frac{1}{3} \quad (2.63)$$

$$y'(0) = \frac{1}{2} + \frac{1}{2} - 1 = 0 \quad (2.64)$$

as required. \square

Remark 2.8. Notice what would have happened if we had forgot to include initial conditions. It would have been as if initial conditions were null, and we would have got

$$s^2Y(s) + 4sY(s) + 3Y(s) = \frac{4}{s-1} \Leftrightarrow Y(s)(s^2 + 4s + 3) = \frac{4}{s-1} \quad (2.65)$$

and then

$$y(t) = \mathcal{L}^{-1} \left[\frac{\frac{1}{2}}{s-1} - \frac{1}{s+1} + \frac{\frac{1}{2}}{s+3} \right] = \frac{1}{2}e^t - e^{-t} + \frac{1}{2}e^{-3t} \quad (2.66)$$

In this case,

$$y'(t) = \frac{1}{2}e^t + e^{-t} - \frac{3}{2}e^{-3t} \quad (2.67)$$

$$y''(t) = \frac{1}{2}e^t - e^{-t} + \frac{9}{2}e^{-3t} \quad (2.68)$$

and so it remains true that

$$\begin{aligned} y''(t) + 4y'(t) + 3y(t) &= \frac{1}{2}e^t - e^{-t} + \frac{9}{2}e^{-3t} + 2e^t + 4e^{-t} - \frac{12}{2}e^{-3t} + \frac{3}{2}e^t - 3e^{-t} + \frac{3}{2}e^{-3t} \\ &= 4e^t \end{aligned} \quad (2.69)$$

but the initial conditions are indeed

$$y(0) = \frac{1}{2} - 1 + \frac{1}{2} = 0 \quad (2.70)$$

$$y'(0) = \frac{1}{2} + 1 - \frac{3}{2} = 0 \quad (2.71)$$

To conclude: if in fact initial conditions were as in Example 2.18, and if we had written (2.65) instead of (2.54), we would get a wrong result. \square

2.6 More important properties: initial and final values, convolution

Before we are done with Laplace transforms, we must establish some additional important properties that will often be needed.

Theorem 2.4. If $f(t)$ and $f'(t)$ have Laplace transforms,

Final value theorem

$$\lim_{t \rightarrow +\infty} f(t) = \lim_{s \rightarrow 0} sF(s) \quad (2.72)$$

provided that $\lim_{t \rightarrow +\infty} f(t) \in \mathbb{R}$.

Proof. Apply a limit to (2.41) to get

$$\begin{aligned} \lim_{s \rightarrow 0} \mathcal{L}[f'(t)] &= \lim_{s \rightarrow 0} (sF(s) - f(0)) \\ \Leftrightarrow f(0) + \lim_{s \rightarrow 0} \int_0^{+\infty} f'(t)e^{-st} dt &= \lim_{s \rightarrow 0} sF(s) \\ \Leftrightarrow f(0) + \int_0^{+\infty} \lim_{s \rightarrow 0} (f'(t)e^{-st}) dt &= \lim_{s \rightarrow 0} sF(s) \\ \Leftrightarrow f(0) + \int_0^{+\infty} f'(t) dt &= \lim_{s \rightarrow 0} sF(s) \\ \Leftrightarrow f(0) + \lim_{t \rightarrow +\infty} f(t) - f(0) &= \lim_{s \rightarrow 0} sF(s) \quad \square \end{aligned} \quad (2.73)$$

Example 2.19. Let $f(t) = e^{-at}$, $a > 0$. We know that $\lim_{t \rightarrow +\infty} f(t) = 0$. We have $F(s) = \frac{1}{s+a}$. And $\lim_{s \rightarrow 0} s F(s) = \lim_{s \rightarrow 0} \frac{s}{s+a} = 0$.

Notice that, when $a < 0$, it is still true that $F(s) = \frac{1}{s+a}$ and that $\lim_{s \rightarrow 0} s F(s) = \lim_{s \rightarrow 0} \frac{s}{s+a} = 0$. But now $\lim_{t \rightarrow +\infty} f(t) = +\infty$, which is not real. \square

Example 2.20. Let $F(s) = \frac{1}{s(s+a)}$, $a > 0$. We have $\lim_{s \rightarrow 0} s F(s) = \lim_{s \rightarrow 0} \frac{1}{s+a} = \frac{1}{a}$. At the same time, $f(t) = \frac{1}{a}(1 - e^{-at})$, and $\lim_{t \rightarrow +\infty} f(t) = \frac{1}{a}$.

When $a < 0$, we are in a situation similar to that of the former example: we still have $\lim_{s \rightarrow 0} s F(s) = \frac{1}{a}$, but $\lim_{t \rightarrow +\infty} f(t) = +\infty$. \square

Initial value theorem

Theorem 2.5. If $f(t)$ and $f'(t)$ have Laplace transforms,

$$\lim_{t \rightarrow 0^+} f(t) = \lim_{s \rightarrow +\infty} s F(s) \quad (2.74)$$

provided that $\lim_{s \rightarrow +\infty} s F(s) \in \mathbb{R}$.

Proof. Apply a limit to (2.41) to get

$$\begin{aligned} \lim_{s \rightarrow +\infty} \mathcal{L}[f'(t)] &= \lim_{s \rightarrow +\infty} (s F(s) - f(0)) \\ \Leftrightarrow f(0) + \lim_{s \rightarrow +\infty} \int_0^{+\infty} f'(t) e^{-st} dt &= \lim_{s \rightarrow +\infty} s F(s) \end{aligned} \quad (2.75)$$

In the integrand, e^{-st} goes to zero as $s \rightarrow +\infty$. If $f'(t)$ has a Laplace transform, it must be of exponential order, and thus e^{-st} goes to zero faster enough to ensure that $\lim_{s \rightarrow +\infty} \int_0^{+\infty} f'(t) e^{-st} dt = 0$. Because we are assuming the unilateral Laplace transform definition, $f(0)$ is in reality $\lim_{t \rightarrow 0^+} f(t)$, as whatever may happen for $t < 0$ is not taken into account. \square

Example 2.21. Let $f(t) = e^{-at}$. We know that $\lim_{t \rightarrow 0^+} f(t) = 1$. We have $F(s) = \frac{1}{s+a}$. And $\lim_{s \rightarrow +\infty} s F(s) = \lim_{s \rightarrow 0} \frac{s}{s+a} = 1$.

Notice that, unlike what happened when we applied the final value theorem in Example 2.19, there is now no need to restrict $a > 0$.

Example 2.22. Let $F(s) = \frac{1}{s(s+a)}$. We have $\lim_{s \rightarrow +\infty} s F(s) = \lim_{s \rightarrow +\infty} \frac{1}{s+a} = 0$. At the same time, $f(t) = \frac{1}{a}(1 - e^{-at})$, and $\lim_{t \rightarrow +\infty} f(t) = 0$. There is again no need now to make $a > 0$.

Convolution

Definition 2.2. Given two functions $f(t)$ and $g(t)$ defined for $t \in \mathbb{R}^+$, their **convolution**, denoted by $*$, is a function of t given by

$$f(t) * g(t) = \int_0^t f(t-\tau)g(\tau) d\tau \quad (2.76)$$

Theorem 2.6. Convolution is commutative.

Proof. Use the change of variables $\mathbf{t} = t - \tau$, for which $d\tau = -d\mathbf{t}$. With this change of variables, when $\tau = 0$ we have $\mathbf{t} = t$, and when $\tau = t$ we have $\mathbf{t} = 0$. Apply this to (2.76) to get

$$\begin{aligned} f(t) * g(t) &= \int_0^t f(t-\tau)g(\tau) d\tau \\ &= - \int_t^0 f(\mathbf{t})g(t-\mathbf{t}) d\mathbf{t} \\ &= \int_0^t f(\tau)g(t-\tau) d\tau = g(t) * f(t) \quad \square \end{aligned} \quad (2.77)$$

Theorem 2.7. If these Laplace transforms exist,

$$\mathcal{L}[f(t) * g(t)] = F(s) G(s) \quad (2.78)$$

Proof.

$$\mathcal{L}[f(t) * g(t)] = \int_0^{+\infty} \left(\int_0^t f(t-\tau)g(\tau) d\tau \right) e^{-st} dt \quad (2.79)$$

We can change the limits of integration of the inner integral by including a Heaviside function $H(t - \tau)$:

$$\mathcal{L}[f(t) * g(t)] = \int_0^{+\infty} \left(\int_0^{+\infty} f(t - \tau)H(t - \tau)g(\tau) d\tau \right) e^{-st} dt \quad (2.80)$$

$H(t - \tau) = 1$ if $t - \tau \geq 0 \Leftrightarrow \tau \leq t$, which is the range of values in (2.79). But $H(t - \tau) = 0$ if $t - \tau < 0 \Leftrightarrow \tau > t$, the additional range of values added in (2.79), which thus does not change the result. We can now change the order of integration:

$$\begin{aligned} \mathcal{L}[f(t) * g(t)] &= \int_0^{+\infty} \left(\int_0^{+\infty} f(t - \tau)H(t - \tau)g(\tau) d\tau \right) e^{-st} dt \\ &= \int_0^{+\infty} \int_0^{+\infty} f(t - \tau)H(t - \tau)g(\tau)e^{-st} dt d\tau \\ &= \int_0^{+\infty} g(\tau) \int_0^{+\infty} f(t - \tau)H(t - \tau)e^{-st} dt d\tau \end{aligned} \quad (2.81)$$

We now apply to the inner integral the change of variables $\mathfrak{t} = t - \tau$, for which $dt = d\mathfrak{t}$. With this change of variables, when $t = 0$ we have $\mathfrak{t} = -\tau$, and when $t \rightarrow +\infty$ we have $\mathfrak{t} \rightarrow +\infty$ too.

$$\mathcal{L}[f(t) * g(t)] = \int_0^{+\infty} g(\tau) \int_{-\tau}^{+\infty} f(\mathfrak{t})H(\mathfrak{t})e^{-s(\tau+\mathfrak{t})} d\mathfrak{t} d\tau \quad (2.82)$$

We have $H(\mathfrak{t}) = 1$ if $\mathfrak{t} \geq 0$ and $H(\mathfrak{t}) = 0$ if $\mathfrak{t} < 0$, so the integration limits can be changed accordingly:

$$\mathcal{L}[f(t) * g(t)] = \int_0^{+\infty} g(\tau) \int_0^{+\infty} f(\mathfrak{t})e^{-s\tau}e^{-s\mathfrak{t}} d\mathfrak{t} d\tau \quad (2.83)$$

All that is left is taking outside integrals terms that do not depend on the corresponding variables:

$$\begin{aligned} \mathcal{L}[f(t) * g(t)] &= \int_0^{+\infty} g(\tau)e^{-s\tau} \left(\int_0^{+\infty} f(\mathfrak{t})e^{-s\mathfrak{t}} d\mathfrak{t} \right) d\tau \\ &= \int_0^{+\infty} f(\mathfrak{t})e^{-s\mathfrak{t}} d\mathfrak{t} \int_0^{+\infty} g(\tau)e^{-s\tau} d\tau \end{aligned} \quad (2.84)$$

and these integrals are the definitions of $F(s)$ and $G(s)$. \square

Example 2.23. From $\mathcal{L}^{-1} \left[\frac{1}{s} \right] = H(t)$ we get

$$\mathcal{L}^{-1} \left[\frac{1}{s^2} \right] = \mathcal{L}^{-1} \left[\frac{1}{s} \frac{1}{s} \right] = \int_0^t H(t - \tau)H(\tau) d\tau = \int_0^t 1 d\tau = t \quad \square \quad (2.85)$$

Remark 2.9. The function obtained is known as the unit slope ramp:

Unit slope ramp

$$f(t) = \begin{cases} t, & \text{if } t \geq 0 \\ 0, & \text{if } t < 0 \end{cases} \quad \square \quad (2.86)$$

Table 2.2 gives a list of important properties of the Laplace transform.

2.7 The Fourier transform

Definition 2.3. If $F(s)$ is the Laplace transform of $f(t)$, then the **Fourier transform** of $f(t)$, denoted by $\mathcal{F}[f(t)]$, is the restriction of $F(s)$ to purely imaginary values of s , i.e. to the imaginary axis of the complex plane, and

$$\mathcal{F}[f(t)] = \mathcal{L}[f(t)]|_{s=j\omega} = F(j\omega), \quad \omega \in \mathbb{R} \quad \square \quad (2.87)$$

See Figure 2.1.

Remark 2.10. Notice that:

Table 2.2: Laplace transform properties

	$x(t)$	$X(s)$
1	$Ax_1(t) + Bx_2(t)$	$AX_1(s) + BX_2(s)$
2	$ax(at)$	$X\left(\frac{s}{a}\right)$
3	$e^{at}x(t)$	$X(s-a)$
4	$\begin{cases} x(t-a) & t > a \\ 0 & t < a \end{cases}$	$e^{-as}X(s)$
5	$\frac{dx(t)}{dt}$	$sX(s) - x(0)$
6	$\frac{d^2x(t)}{dt^2}$	$s^2X(s) - sx(0) - x'(0)$
7	$\frac{d^n x(t)}{dt^n}$	$s^n X(s) - s^{n-1}x(0) - \dots - x^{(n-1)}(0)$
8	$-tx(t)$	$\frac{dX(s)}{ds}$
9	$t^2x(t)$	$\frac{d^2X(s)}{ds^2}$
10	$(-1)^n t^n x(t)$	$\frac{d^n X(s)}{ds^n}$
11	$\int_0^t x(u) du$	$\frac{1}{s}X(s)$
12	$\int_0^t \dots \int_0^t x(u) du = \int_0^t \frac{(t-u)^{(n-1)}}{(n-1)!} x(u) du$	$\frac{1}{s^n}X(s)$
13	$x_1(t) * x_2(t) = \int_0^t x_1(u) x_2(t-u) du$	$X_1(s) X_2(s)$
14	$\frac{1}{t}x(t)$	$\int_s^\infty X(u) du$
15	$x(t) = x(t+T)$	$\frac{1}{1-e^{-sT}} \int_0^T e^{-su} X(u) du$
16	$x(0)$	$\lim_{s \rightarrow \infty} sX(s)$
17	$x(\infty) = \lim_{t \rightarrow \infty} x(t)$	$\lim_{s \rightarrow 0} sX(s)$

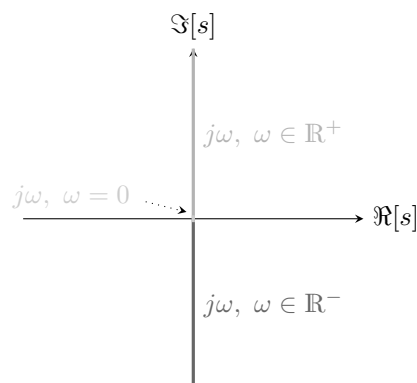


Figure 2.1: The imaginary axis in the complex plane.

- $f(t)$ is a real-valued function that depends on a real variable: $f(t) \in \mathbb{R}$, and $t \in \mathbb{R}$;
- the Laplace transform of $f(t)$, which is $F(s) = \mathcal{L}[f(t)]$, is a complex-valued function that depends on a complex variable: $F(s) \in \mathbb{C}$, and $s \in \mathbb{C}$;
- the Fourier transform of $f(t)$, which is $F(j\omega) = \mathcal{F}[f(t)]$, is a complex-valued function that depends on a real variable, that is the coordinate along the imaginary axis: $F(j\omega) \in \mathbb{C}$, and $\omega \in \mathbb{R}$. \square

Example 2.24. Let $f(t) = e^{-t} - e^{-10t}$. Then

$$\begin{aligned}
 F(s) &= \frac{9}{(s+1)(s+10)} & (2.88) \\
 F(j\omega) &= \frac{9}{(j\omega+1)(j\omega+10)} \\
 &= \frac{9}{(10-\omega^2) + j11\omega} \\
 &= \frac{9((10-\omega^2) - j11\omega)}{((10-\omega^2) + j11\omega)((10-\omega^2) - j11\omega)} \\
 &= \frac{9(10-\omega^2) - j99\omega}{(10-\omega^2)^2 + 121\omega^2} \\
 &= \frac{90-9\omega^2}{\omega^4+101\omega^2+100} + j\frac{-99\omega}{\omega^4+101\omega^2+100} \quad \square & (2.89)
 \end{aligned}$$

Example 2.25. Let $F(j\omega) = \frac{\omega_0}{\omega_0^2 - \omega^2}$, where ω_0 is a real constant. The function *Inverse Fourier transform* $f(t)$ of which $F(j\omega)$ is the Fourier transform is the **inverse Fourier transform** of $F(j\omega)$, and is given by

$$\begin{aligned}
 f(t) &= \mathcal{F}^{-1}[F(j\omega)] = \mathcal{F}^{-1}\left[\frac{\omega_0}{\omega_0^2 - \omega^2}\right] = \mathcal{F}^{-1}\left[\frac{\omega_0}{\omega_0^2 + (j\omega)^2}\right] \\
 &= \mathcal{L}^{-1}\left[\frac{\omega_0}{\omega_0^2 + s^2}\right] = \sin(\omega_0 t) \quad \square & (2.90)
 \end{aligned}$$

While it should now be clear what we need Laplace transforms for, we will only see what we need Fourier transforms for in chapter 10. A more detailed study of this transform is found in Chapter 38.

Glossary

I said it in Hebrew — I said it in Dutch —
 I said it in German and Greek:
 But I wholly forgot (and it vexes me much)
 That English is what you speak!

Lewis CARROLL (1832 — †1898), *The hunting of the Snark* (1876), 4

bilateral Laplace transform transformada de Laplace bilateral

convolution convolução

differential equation equação diferencial

exponential order function função de ordem exponencial

Fourier transform transformada de Fourier

integral transform transformada integral

Laplace transform transformada de Laplace

Laplace transformation transformação de Laplace

partial fraction expansion expansão em frações parciais

pole polo

polynomial division divisão de polinômios

residue resíduo

Exercises

- Find the Laplace transforms of the following functions:
 - $f(t) = 80e^{-0.9t}$
 - $f(t) = 1000 - e^{-6t}$
 - $f(t) = 97.5 \times 10^{-3} \sin(0.2785t) + 546.9 \times 10^{-3} e^{0.9575t} \cos(0.9649t)$
 - $f(t) = \sin\left(5t + \frac{\pi}{6}\right)$ *Hint:* remember that $\sin(a+b) = \sin a \cos b + \cos a \sin b$.
- Find the inverse Laplace transforms of the following functions:
 - $F(s) = \frac{1}{3s^2+15s+18}$
 - $F(s) = \frac{1}{5s^2+6s+5}$
 - $F(s) = \frac{8s^2+34s-2}{s^3+3s^2-4s}$
 - $F(s) = \frac{s^2+2s+8}{2s+4}$
 - $F(s) = \frac{-s^2+5s-2}{s^3-2s^2-4s+8}$
- Consider differential equation (2.50) from Example 2.17, but now with the initial condition $y(0) = 2$.
 - Show that $y(t) = \mathcal{L}^{-1}\left[\frac{2s}{(s+1)^2} + \frac{3}{(s+1)^2}\right]$.
 - Show from (2.41) that $\mathcal{L}[e^{-t}(1-t)] = \frac{s}{(s+1)^2}$.
 - Find $y(t)$ and check that it verifies both the differential equation (2.50) and the new initial condition.
- Solve the following differential equations:
 - $y''(t) + y(t) = te^{-t}$, $y(0) = 0$, $y'(0) = 0$
 - $y''(t) + y(t) = te^{-t}$, $y(0) = \frac{1}{2}$, $y'(0) = -\frac{1}{2}$
 - $y''(t) + y(t) = 10t - 20$, $y(0) = 0$, $y'(0) = 0$
 - $3y''(t) + 7y'(t) + 2y(t) = 0$, $y(0) = -5$, $y'(0) = 10$
- Use the final value and initial value theorems to find the initial and final values of the inverse Laplace transforms of the functions of Exercise 2.
- Find the Fourier transforms of the functions of exercises 1 and 2, putting them in the form $F(j\omega) = \Re[F(j\omega)] + j\Im[F(j\omega)]$.
- Prove the result in line 7 of Table 2.1. *Hint:* use (2.78) together with the result in line 5.
- Prove the result in line 8 of Table 2.1. *Hint:* use mathematical induction.
- Show that:
 - the Fourier transform is given by

$$\mathcal{F}[f(t)] = \int_0^{+\infty} f(t)e^{-j\omega t} dt \quad (2.91)$$

Hint: apply (2.87) to (2.1).
 - the bilateral Fourier transform is given by

$$\mathcal{F}[f(t)] = \int_{-\infty}^{+\infty} f(t)e^{-j\omega t} dt \quad (2.92)$$

Hint: apply (2.87) to (2.2).

Chapter 3

Examples of mechatronic systems and signals

Eia comboios, eia pontes, eia hotéis à hora do jantar,
Eia aparelhos de todas as espécies, férreos, brutos, mínimos,
Instrumentos de precisão, aparelhos de triturar, de cavar,
Engenhos, brocas, máquinas rotativas !
Eia ! eia ! eia !
Eia electricidade, nervos doentes da Matéria !
Eia telegrafia-sem-fios, simpatia metálica do Inconsciente !

Álvaro de CAMPOS, heteronym of Fernando PESSOA (1888 — †1935), *Ode triunfal*, Orpheu, I 1, January–March 1915

In this chapter we discuss different types of mechatronic signals and systems, and present examples of each.

3.1 Systems

In chapter 1 we have already defined system as the part of the Universe we want to study.

A system made up of physical components may be called a **plant**. A system which is a combination of operations may be called a **process**. *Plant*
Process

Example 3.1. WECs, mentioned in Example 1.1, are plants. Figures 3.1, 3.2 and 3.3 show three different WECs; many other such devices exist. □

Example 3.2. If we want to study the wave elevation at a certain onshore location as a function of the weather on the middle of the ocean, we will be studying a process.

The variables describing the characteristics of the system that we are interested in are its **outputs**. The variables on which the outputs depend are the system's **inputs**. *Outputs*
Inputs in the general sense

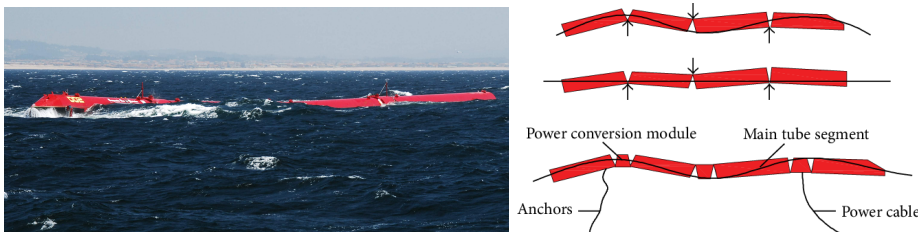


Figure 3.1: The Pelamis, a floating near-shore Wave Energy Converter, at Aguçadoura, Portugal (source: left, Wikimedia; right, DOI 10.1155/2013/186056). Waves cause an angular movement of the several sections of the device. This movement pumps oil in a closed circuit; high pressure oil is then used to run a turbine driving a usual rotational generator.

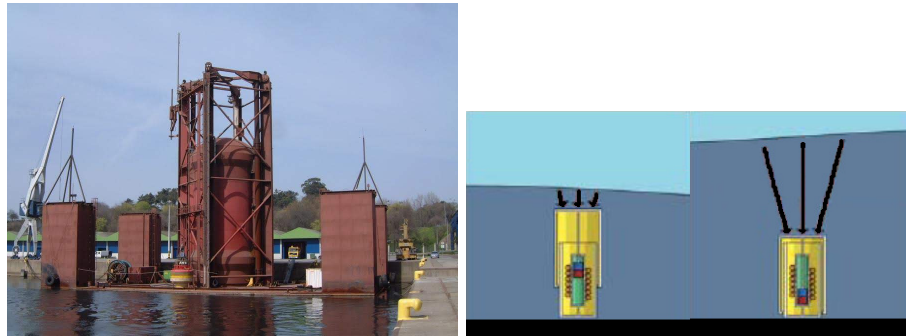


Figure 3.2: The Archimedes Wave Swing, a submerged offshore Wave Energy Converter before submersion, at Viana do Castelo, Portugal. The device is filled with air which is compressed when wave crests pass and expands during wave troughs. The heaving movement of the AWS upper part moves an electrical linear generator.

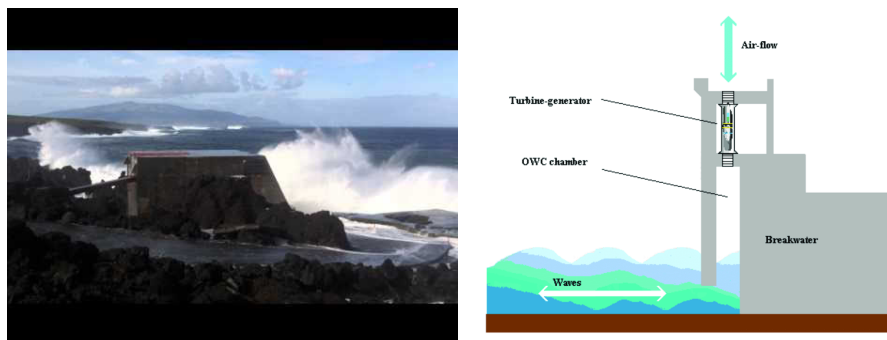


Figure 3.3: The Pico Power Plant (concluded in 1999, decommissioned in 2018), an onshore Wave Energy Converter of the Oscillating Water Column (OWC) type (source: left, WavEC; right, DOI 10.3390/en1112939). In an OWC, the heaving movement of the water inside a chamber compresses and expands the air, which can flow in and out the chamber through a turbine designed to always rotate in the same direction irrespective of the sense of the flow. The turbine drives a usual rotational generator.



Figure 3.4: A Panhard & Levassor Type A motorcar, the first mass produced car in the world, driven by the French priest Jules Gavois (1863 – †1946) in 1891 (source: Wikimedia). This car still does not have a steering wheel (first introduced in 1894), but only a tiller.

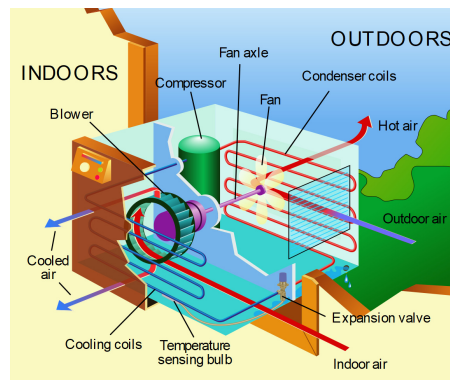


Figure 3.5: A window unit air conditioning system (source: Wikimedia). There are many other types of AC units.

Example 3.3. An internal combustion engine motorcar (see Figure 3.4) is a plant. We are usually interested in its position, velocity, and attitude. We also may want to know the rotation speed of the motor, the temperature of the oil, the fuel consumption, or other such values. All these are outputs. They depend on the position of the steering wheel, the position of the accelerator and brake pedals, the gear selected, the condition and inclination of the road the car is running on, the direction and speed of the wind, the outside temperature, and other such values. These are the inputs. Not all the outputs depend on all the inputs. \square

A **control system** is one devised to make one or more of the system's outputs follow some **reference**.

Control system

Reference

Example 3.4. An air conditioning (AC) unit (see Figure 3.5) is a mechatronic system that heats or cools a room to a temperature set by the user. It is consequently a control system. The value of the temperature selected by the AC user is the reference. The room's temperature is the output of the plant that has to follow this reference. \square

Example 3.5. The wind at the location of a wind turbine is related to the temperature, the solar exposition, and the atmospheric pressure, among other variables. This is a process we cannot control. It is not a control system. \square

For a control system to exist, it must be possible to modify one or more of the inputs, so as to affect the desired outputs and thereby cause them to follow the reference. Such inputs are called **manipulated variables** or **inputs in the strict sense**. The inputs of the system that cannot be modified are called **disturbances**. When studying control systems, it is usual to call simply

Inputs in the strict sense

Disturbances

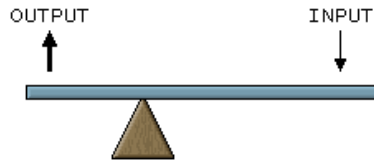


Figure 3.6: A lever, an example of a linear SISO system without dynamics (source: Wikimedia).

inputs to the inputs in the strict sense, and to call outputs only to the variable or variables that have to follow a reference.

Example 3.6. In the case of the OWC from Example 3.1 and Figure 3.3, the sea waves are disturbances, since we cannot control them. The rotation speed of the turbine is an input in the strict sense, since we can manipulate it (e.g. varying the resistance of the electrical generator). If the OWC chamber has a relief valve, the pressure in the chamber will be also an input, since we can change it opening or closing the relief valve. \square

Example 3.7. In the case of the car from Example 3.3, the positions of the steering wheel and of the pedals are inputs. If the car has a manual gear box, the gear selected is an input too; if the gear box is automatic, it is not. The gusts of wind are a disturbance, since we cannot modify them. If we are studying the temperature of the motor of the car, this will depend on the outside temperature, which we cannot control and is therefore a disturbance. \square

A system with only one input and only one output is a Single-Input, Single-Output (**SISO**) system. A system with more than one input and more than one output is a Multi-Input, Multi-Output (**MIMO**) system. It is of course possible to have Single-Input, Multiple-Output (**SIMO**) systems, and Multiple-Input, Single-Output (**MISO**) systems. These are usually considered as particular cases of MIMO systems.

Example 3.8. Both the OWC of Example 3.1 and the car of Example 3.3 are MIMO plants. \square

Example 3.9. The lever in Figure 3.6 is a SISO system: if the extremities are at heights $x(t)$ and $y(t)$, and the first is actuated, then $y(t)$, the output, depends on position $x(t)$, the input, and nothing more. \square

Remark 3.1. We can sometimes model a part of a MIMO system as a separate SISO system. A car is a MIMO system, as we said, but we can model the suspension of one wheel separately, as a SISO system relating the position of the wheel to the position of the vehicle frame, neglecting the influence that all the other inputs have on this particular output. Such a model is an approximation, but for many purposes it is good enough (as we will see in Example 4.2, in Chapter 4 below). \square

SISO system
MIMO system

Model

A system's **model** is the mathematical relation between its outputs, on the one hand, and its inputs in the general sense (inputs in the strict sense and disturbances), on the other.

Linear system
Non-linear system

A system is **linear** if its exact model is linear, and **non-linear** if its exact model is non-linear. Of course, exact non-linear models can be approximated by linear models, and often are, to simplify calculations.

Example 3.10. The lever of Figure 3.6 is a linear plant, since, if its arm lengths are L_x and L_y for the extremities at heights $x(t)$ and $y(t)$ respectively,

$$y(t) = \frac{L_y}{L_x} x(t). \quad \square \quad (3.1)$$

Example 3.11. A Cardan joint (see Figure 3.7) connecting two rotating shafts, with a bent corresponding to angle β , is a non-linear plant, since a rotation of $\theta_1(t)$ in one shaft corresponds to a rotation of the other shaft given by

$$\theta_2(t) = \arctan \frac{\tan \theta_1(t)}{\cos \beta}. \quad (3.2)$$

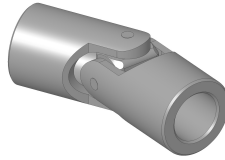


Figure 3.7: A Cardan joint, a non-linear mechanical system without dynamics (source: Wikimedia).

If $\beta \approx 0$, (3.2) can be approximated by

$$\theta_2(t) = \arctan \frac{\tan \theta_1(t)}{1} = \theta_1(t). \quad (3.3)$$

The error incurred in approximating (3.2) by (3.3) depends on how close $\cos \beta$ is to 1. There will be no error at all if the two shafts are perfectly aligned ($\beta = 0$). \square

Example 3.12. A car is also an example of a non-linear plant, as any driver knows. \square

A system is **time-varying** if its exact model changes with time, and **time-invariant** otherwise.

Time-varying system
Time-invariant system

Example 3.13. An airplane consumes enormous amounts of fuel. Thus its mass changes significantly from take-off to landing. Any reasonable model of a plane will have to have a time-varying mass. But it is possible to study a plane, for a short period of time, using an approximation consisting of a time-invariant model, as the mass variation is neglectable in that case. \square

Example 3.14. A drone powered by a battery will not have a similar variation of mass. It is a time-invariant system (unless e.g. its mass changes because it is a parcel-delivering drone). \square

Example 3.15. WECs can have time-varying parameters due to the effects of tides. This is the case of the AWS in Figure 3.2, which is submerged and fixed to the ocean bottom. Consequently, the average height of sea water above it varies from low tide to high time, even if the sea waves remain the same. Other WECs are time-invariant, at least with respect to tides. That is the case of the floating OWC in Figure 3.8, which, precisely because it floats, is not affected by tides. \square

A system has no dynamics if its outputs in a certain time instant do not depend on past values of the inputs or on past values of the disturbances. Otherwise, it is a **dynamic system**. A system without dynamics is called **static system**, which does not mean that it never changes; it means that, if its inputs do not change, neither do the outputs.

Dynamic system
Static system

Example 3.16. Both mechanical systems in Figures 3.6 and 3.7 have no dynamics, since the output $y(t)$ only depends on the current value of the input $u(t)$. Past values of the input are irrelevant. \square

Example 3.17. Consider a pipe with a tap (or a valve) that delivers a flow rate $Q(t)$ given by

$$Q(t) = k_Q f(t) \quad (3.4)$$

where $f(t) \in [0, 1]$ is a variable that tells is if the tap is open ($f(t) = 1$) or closed ($f(t) = 0$). This system is static. But a tap placed far from the point where the flow exits the pipe will deliver a flow given by

$$Q(t) = k_Q f(t - \tau) \quad (3.5)$$

Here, τ is the time the water takes from the tap to the exit of the pipe. This is an example of a dynamic plant, since its output at time instant t depends on a past value of $f(t)$. \square

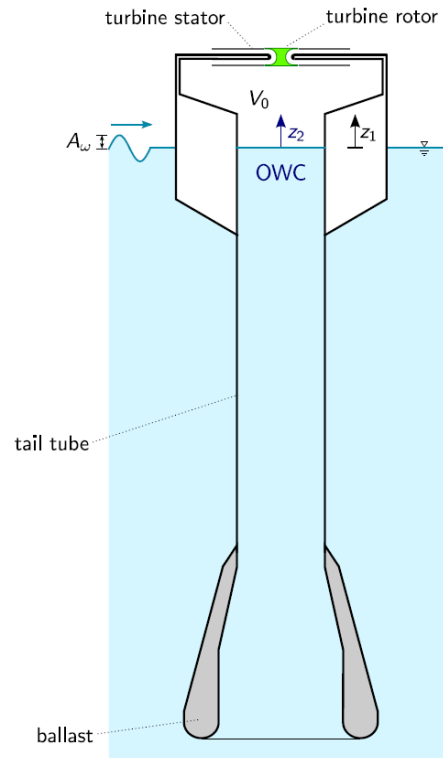


Figure 3.8: A floating OWC (source: DOI 10.1016/j.energy.2016.06.054).

A system is **deterministic** if the same inputs starting from the same initial condition always lead to the same output. A system is **stochastic** if its outputs are not necessarily the same when it is subject to the same inputs beginning with the same initial conditions, or, in other words, if its output is random.

Deterministic system
Stochastic system

Example 3.18. The process from Example 3.5 is stochastic. Even though we may know all those variables, it is impossible to precisely predict the wind speed. The same happens with the process from Example 3.2, and even more so. \square

Example 3.19. Figure 3.9 shows a laboratory setup to test controllers for the lithography industry (which produces microchips with components positioned with precisions of the order of 1 nm). This is a deterministic system. If lithography plants and processes were not deterministic, it would be far more difficult to mass produce microchips. \square

In this course we will only address deterministic, SISO, linear time-invariant (**LTI**) systems.

LTI systems

3.2 Signals

In chapter 1 we have already defined signal as a function of time or space that conveys information about a system. In other words, it is the evolution with time or with space of some variable that conveys information about a system. Most of the signals we will meet depend on time but not on space.

Example 3.20. An image given by a camera is a signal that depends on space, but not on time. A video is a signal that depends on both space and time. \square

Quantised signal
Analogical signal

Some signals can only take values in a discrete set; they are called **quantised signals**. Others can take values in a continuous set; they are called **analogical signals**.

Example 3.21. Consider a turbine, such as the turbine in Figure 3.10, of the Wells type, installed in the Pico Power Plant (shown above in Figure 3.3). Its rotation speed is real valued; it takes values in a continuous set. So the signal consisting in the turbine's rotation speed as a function of time is an analogical signal. \square

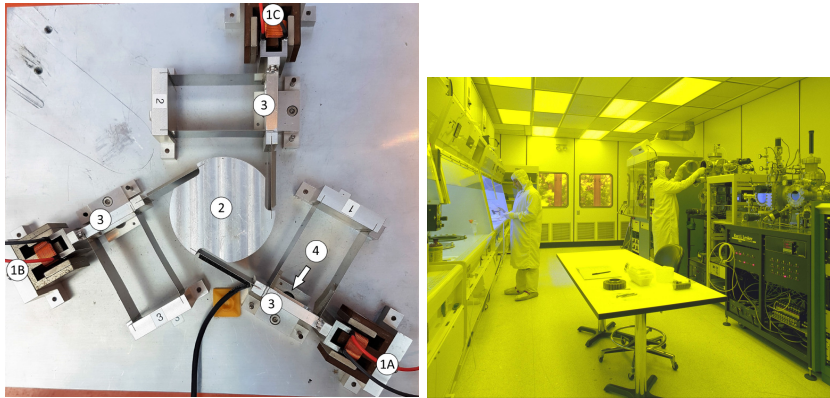


Figure 3.9: Left: precision positioning system used at the Delft University of Technology (source: DOI 10.1007/s11071-019-05130-2). Coil actuators 1 move masses 2, which are connected through flexures to mass 3, the position of which is measured using sensors (encoders) 4. Mass 2 can be positioned with a precision of $1 \mu\text{m}$ or less. Right: NASA clean room for lithography (source: Wikimedia).

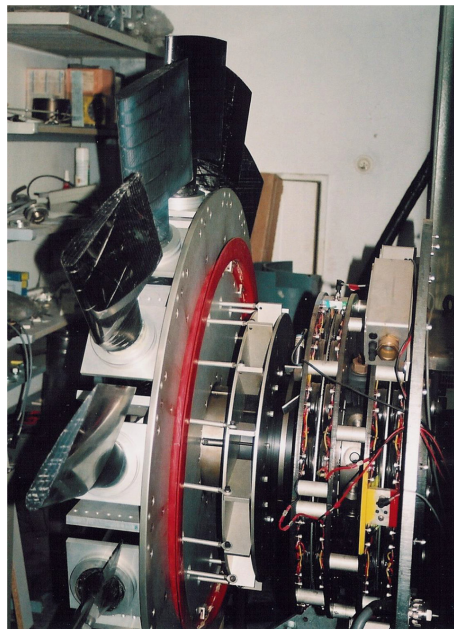


Figure 3.10: The Wells turbine of the Pico Power Plant OWC in Figure 3.3 (source: DOI 10.1016/j.renene.2015.07.086).

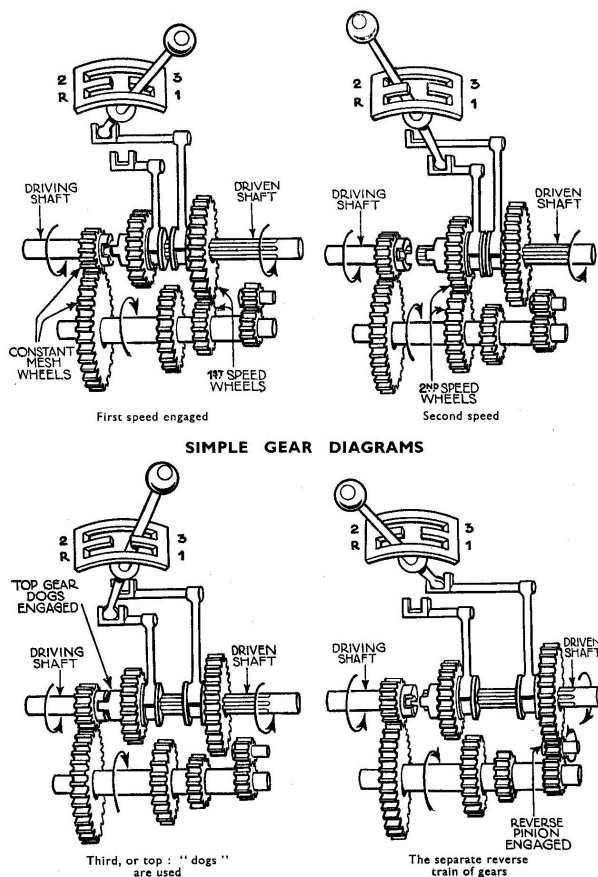


Figure 3.11: Three-speed manual gearbox, typical of cars in the 1930s (source: Wikimedia).

Example 3.22. Consider the gearbox of a car (see Figure 3.11). The signal consisting in the speed engaged as a function of time (neutral, reverse, 1st, 2nd, etc.) takes values in a discrete set. It is a quantised signal. \square

Remark 3.2. It is possible, and sometimes desirable, to approximate a quantised signal by an analogical signal, and vice-versa. \square

Example 3.23. The rotation of a shaft $\theta(t)$ is an analogical signal; it is of course possible to rotate the shaft by an angle as small as desired. But it is often useful to replace it by a discrete signal $\vartheta(t)$ which is the number of revolutions (i.e. the number of 360° rotations of the shaft). This corresponds to an approximation given by $\vartheta(t) = \left\lfloor \frac{\theta(t)}{360^\circ} \right\rfloor$. Figure 3.12 shows a mechanical revolution counter. \square

Example 3.24. A population — be it the number of persons in a country, the number of rabbits in a field, or the number of bacteria on a Petri dish — is a quantised signal. It always increases or decreases in multiples of one, since it is impossible that half a child be born, or that $\frac{3}{4}$ of a rabbit dies. However, if the population is large enough, a variation of one individual is so small that it is possible to assume that it is an analogical signal, and write equations such as

$$\frac{dp(t)}{dt} = b(t)p(t) - d(t)p(t), \quad (3.6)$$

where $p(t)$ is the population, $b(t)$ is the birth rate, and $d(t)$ is the death rate. (Terms for immigration and emigration rates must be included if the population is not isolated.) Such models (and others more complicated, that we will mention in passing below in Chapter 43) are used for instance in Bioengineering and in many other areas. \square

Example 3.25. Strictly speaking, variables such as the fuel admitted to one of the cylinders of an internal combustion engine are also quantised, since the num-



Figure 3.12: Mechanical 19th century revolution counter, from the former Barbadinhos water pumping station (currently the Water Museum), Lisbon. Nowadays mechanical revolution counters are still used, though electronic ones exist.

ber of molecules of fuel admitted is integer. Of course, in practice an analogical value is assumed. \square

Some signals take values for all time instants: they are said to be **continuous**. Others take values only at some time instants: they are said to be discrete in time, or, in short, **discrete**. The time interval between two consecutive values of a discrete signal is the **sampling time**. The sampling time may be variable (if it changes between different samples), or constant. In the later case, which makes mathematical treatment far more simple, the inverse of the sampling time is the **sampling frequency**.

Continuous signal

Discrete signal

Sampling time

Sampling frequency

Example 3.26. The air pressure inside the chamber of an OWC is a continuous signal: it takes a value for every time instant. \square

Example 3.27. The number of students attending the several classes of this course along the semester is a discrete signal: there is a value for each class, and the sampling time is the time between consecutive classes. The sampling time may be constant (if there is e.g. one laboratory class every Monday) or variable (if there are e.g. two lectures per week on Mondays and Wednesdays). \square

Example 3.28. One of the controllers used with the laboratory setup from Example 3.19 in Figure 3.9 provided a discrete control action with sampling frequency 20 kHz. So the sampling time was

$$T_s = \frac{1}{20 \times 10^3} = 50 \times 10^{-6} \text{ s} = 50 \mu\text{s}, \quad (3.7)$$

or, in other words, every 50×10^{-6} s the control action for the coil actuators was updated; or, again, the control action was updated 20×10^3 times per second. The sampling frequency could also be given as

$$\omega_s = \frac{2\pi}{50 \times 10^{-6}} = 2\pi \times 20 \times 10^3 = 125.7 \times 10^3 \text{ rad/s}. \quad \square \quad (3.8)$$

Remark 3.3. Mind the numerical difference between the value of the sampling frequency in Hertz and in radians per second. It is a common source of mistakes in calculations. \square

Remark 3.4. It is possible, and sometimes desirable, to approximate a discrete signal by a continuous signal, and vice-versa. Approximating a continuous signal by a discrete one is an operation called **discretisation**. We will study this issue in more detail below in Chapters 12 and 25. \square

Discretisation

Example 3.29. The control action from Example 3.28 had in fact to be converted into a continuous signal to be applied by the coil actuators. As described, this was done by keeping the control action signal constant between sampling times. The operation corresponds to converting a discrete signal as seen in the left of Figure 3.13 into a continuous signal as seen in the right diagram of that Figure. The conversion of a digital signal into an analogical signal will be addressed in detail in Chapter 25. \square

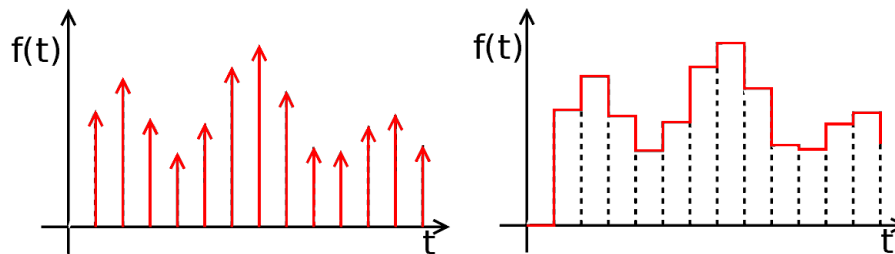


Figure 3.13: Left: discrete signal; right: continuous signal obtained from the discrete signal by keeping the previous value between sampling times (source: Wikimedia, modified).

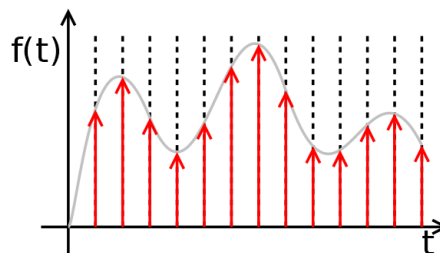


Figure 3.14: Discretising a signal (source: Wikimedia), i.e. approximating a continuous signal (grey) by a discrete one (red).

Example 3.30. Figure 3.14 illustrates the operation of discretisation. □

A signal which is both discrete and quantised is a **digital signal**.

Digital signal

A system, too, is said to be continuous, discrete, or digital, if all its inputs and outputs are respectively continuous, discrete, or digital.

Continuous system

Discrete system

Digital system

Electronic components are nowadays ubiquitous. As a result of sensors, actuators, controllers, etc. being electronic, most signals are digital. Likewise, systems that incorporate such components are digital, inasmuch their inputs and outputs are all digital.

Example 3.31. Consider an industrial oven, seen in Figure 3.15, with a control system to regulate its temperature. The output of this system is the actual temperature inside the oven, and the input is the desired temperature (i.e. the reference of the control system). The oven is heated by gas, and so the gas flow is the manipulated variable that allows controlling the oven. This is a continuous system, since all variables exist in all time instants. But, in all likelihood, a digital sensor will be used for the temperature, and changes in gas flow will also take place at sampling times, after the temperature reading is compared with the reference and processed to find the control action that will better eliminate the error between actual and desired temperatures. So in practice the system will probably be digital. □

Example 3.32. A flush tank for a toilet equipped with a float valve as seen in the top scheme of Figure 3.16 is a control system devoid of any electronic component, and for which all signals are continuous. (See also Figure 3.17.) This is a continuous control system. □

Bounded signal

A signal is **bounded** if it can only assume values in a bounded interval. In engineering, most signals (if not all) are bounded.

Example 3.33. The wave elevation at given coordinates cannot be less than the depth of the sea there. Similarly, the rotation speed of a turbine, or the linear velocity of a shaft, or a voltage in a circuit, are always limited by physical constraints. □

Remark 3.5. Bounded continuous signals can assume infinite values, but bounded quantised signals can only assume a finite number of values. □



Figure 3.15: Industrial oven for aircraft component manufacture (source: Wikimedia).

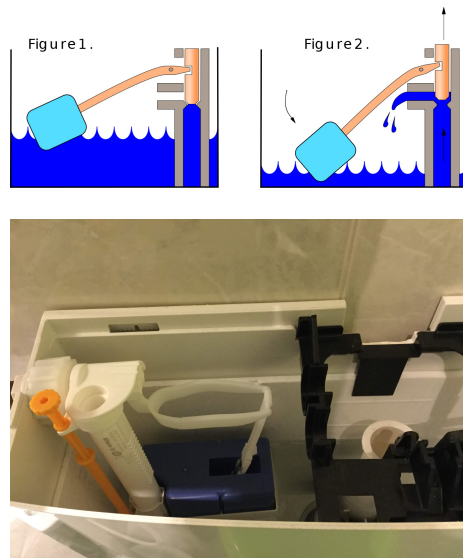


Figure 3.16: Top: float valve mechanism, well known by its use in flush tanks (source: Wikimedia). Bottom: flush tank with a float valve (notice that the lever has two arms, to increase the speed with which the water flow is interrupted as soon as the float raises the level from the lower end of stroke).

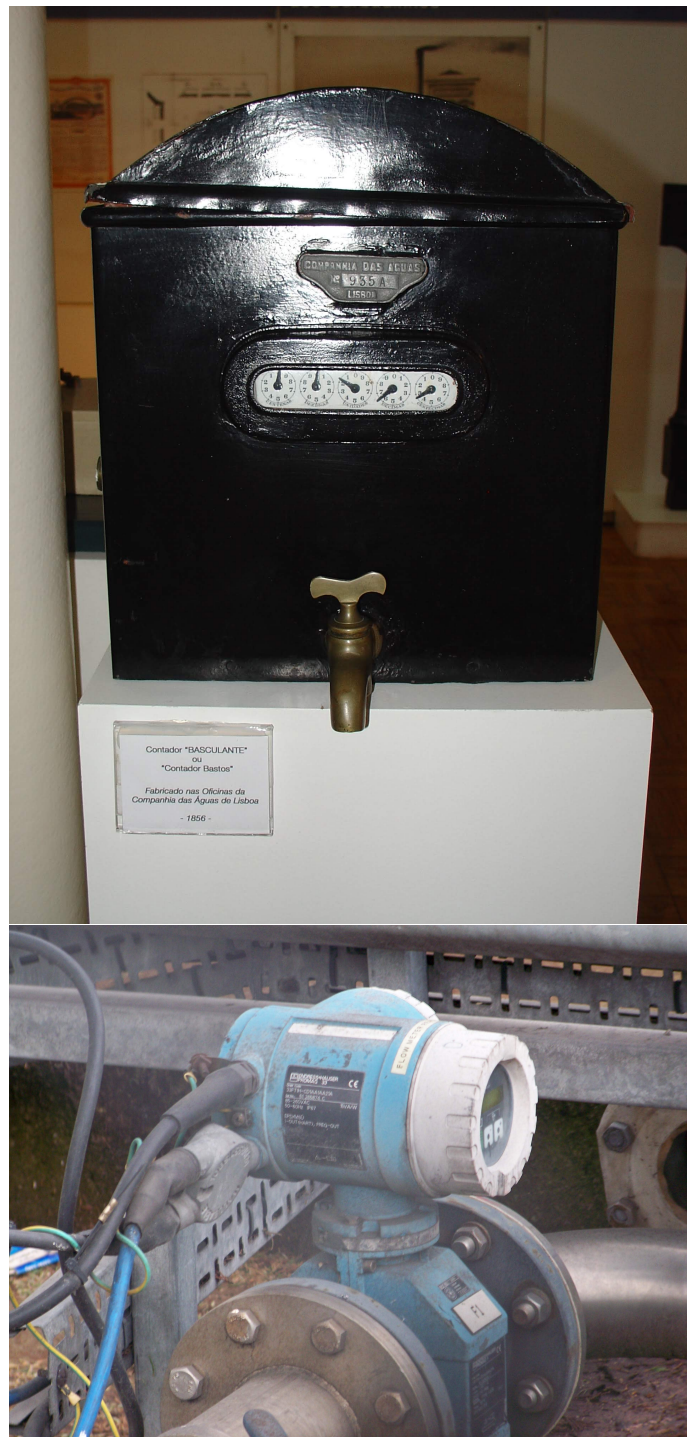


Figure 3.17: Top: a float valve (see Figure 3.16) was also used in water meters devised by António Pinto Bastos in the 1850s, which were used in Lisbon until the 1960s in spite of being obsolescent for a long time by then (source: Wikimedia). These meters were purely mechanical. Bottom: electromagnetic flow meters have no mechanical components; the reading can be sent elsewhere rather than having to be read in the dials *in loco* (source: Wikimedia). We will address sensors for flow measurements in Chapter 13.

3.3 Models

In Section 3.1 we have already defined a system's model as a mathematical relation between its inputs and outputs. There are basically two ways of modelling a system:

1. A model based upon **first principles** is a theoretical construction, resulting from the application of physical laws to the components of the plant. *First principles model*
2. A model based upon **experimental data** results from applying identification methods to data experimentally obtained with the plant. *Experimental model*

It is also possible to combine both these methods.

In this course we will concentrate on models based upon first principles, and you will find abundant examples thereof in Chapters 4 through 8. They can be obtained whenever the way the system works is known. They are the only possibility if the system does not exist yet because it is still being designed and built, or if no experimental data is available. They may be quite hard to obtain if the system comprises many complicated interacting sub-parts. Simplifications can bring down the model to more manageable configurations, but its theoretical origin may mean that results will differ significantly from reality if parameters are wrongly estimated, if too many simplifications are assumed, or if many phenomena are neglected.

When to use first principles models

The models of dynamic continuous LTI systems are given by linear **differential equations**. The models of dynamic digital LTI systems are given by linear **difference equations**. The models of static LTI systems are linear and have neither derivatives nor time differences.

Differential equations

Difference equations

Example 3.34. The static model of the lever (3.1) includes neither differential nor difference equations. It is irrelevant whether $x(t)$ and $y(t)$ are discretised or not. The same happens with the non-linear static model of the Cardan joint (3.2). \square

Example 3.35. Continuous model (3.6) is a differential equation. Suppose that the model is applied to the population of a country, where immigration and emigration are neglectable, and for which population data is available on a yearly basis. Also suppose that birth and death rates are constant and given by $b = 0.03/\text{year}$ and $d = 0.02/\text{year}$. So

$$\frac{dp(t)}{dt} = 0.01p(t) \quad (3.9)$$

Because the sampling time is $T_s = 1$ year, we can perform the following approximation:

$$\left. \frac{dp(t)}{dt} \right|_{t=\text{year } k} \approx \frac{p_k - p_{k-1}}{1 \text{ year}}, \quad (3.10)$$

where p_k is the population in year k , and p_{k-1} is the population in the year before. Notice that this is a first order approximation for the derivative in year k , in which we use the value of the year before, and is consequently called a backward approximation. So we end up with the following difference equation:

$$p_k - p_{k-1} = 0.01p_k \Leftrightarrow p_k = \frac{1}{0.99}p_{k-1}, \quad (3.11)$$

which is an approximation of differential equation (3.9); approximations other than (3.10) could have been used instead. We will address this subject further below in Part V. \square

Example 3.36. Differential equation (2.53) can be approximated by difference equation

$$3y_k = 0.4y_{k-1} + 0.2y_{k-2} + 0.8e^{T_s k} + 1.6e^{T_s (k-1)} + 0.8e^{T_s (k-2)} \quad (3.12)$$

for sampling time T_s . Once more, we will see how to arrive at this result below in Part V. \square

Experimental data should, whenever available, be used to confirm, and if necessary modify, models based upon first principles. This often means that first principles are used to find a structure for a model (the orders of the derivatives in a differential equation, or the number of delays in a difference equation), and then the values of the parameters are found from experimental data: feeding the model the inputs measured, checking the results, and tuning the parameters until they are equal (or at least close) to measured outputs. This can sometimes be done using least squares; sometimes other optimisation methods, such as genetic algorithms, are resorted to. If the outputs of experimental data cannot be made to agree with those of the model, when the inputs are the same, then another model must be obtained; this often happens just because too many simplifications were assumed when deriving the model from first principles. It may be possible to find, from experimental data itself, what modifications to model structure are needed. This area is known as **identification**, and will be addressed below in Parts VI to VIII.

Experimental identification of model parameters

White box model

Models based upon first principles can be called **white box models**, since the reason why the model has a particular structure is known. If experimental data requires changing the structure of the model, a physical interpretation of the new parameters may still be possible. The resulting model is often called a **grey box model**.

Grey box model

There are methods to find a model from experimental data that result in something that has no physical interpretation, neither is it expected to have. Still the resulting mathematical model fits the data available, providing the correct outputs for the inputs used in the experimental plant. Such models are called **black box models**, in the sense that we do not understand how they work. Such models include, among others, neural network (NN) models (see an example in Figure 3.18) and models based upon fuzzy logic, known as fuzzy models (see Figure 3.19). These modelling techniques are increasingly important, but we will not study them in these Lecture Notes.

Black box model

Glossary

“Lascia stare, cerchiamo un libro greco!”

“Questo?” chiedo io mostrandogli un’opera dalle pagine coperte di caratteri astrusi. E Guglielmo: “No, questo è arabo, sciocco! Aveva ragione Bacone che il primo dovere del sapiente è studiare le lingue!”

“Ma l’arabo non lo sapete neppure voi!” ribattevo piccato, al che Guglielmo mi rispondeva: “Ma almeno capisco quando è arabo!”

Umberto ECO (1932 — †2016), *Il nome della rosa* (1980), Quinto giorno, Sesta

black box model modelo de caixa negra
bounded limitado
control system sistema de controlo
continuous contínuo
deterministic determinístico
difference equation equação às diferenças
digital digital
discrete discreto
disturbance perturbação
dynamic dinâmico
first principles primeiros princípios
grey box model modelo de caixa cinzenta
identification identificação
manipulated variable variável manipulada
mechatronic mecatrónico
mechatronics mecatrónica
multiple input entradas múltiplas
multiple output saídas múltiplas
input entrada
model modelo
output saída

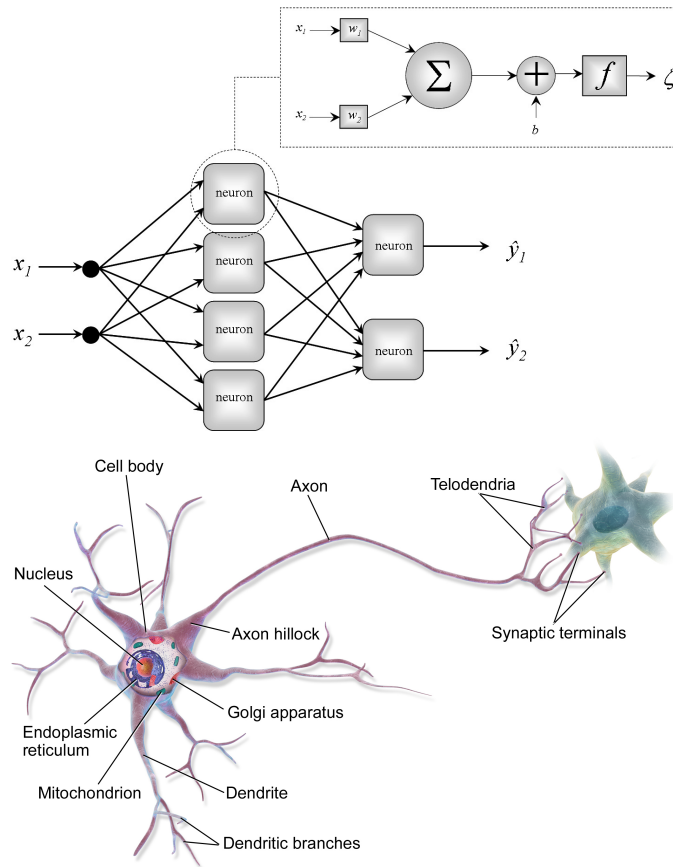


Figure 3.18: Top: scheme of an example of an artificial neural network (source: DOI 10.1016/j.apor.2008.11.002). It is made of several neurons, arranged in layers. These neurons are oversimplified models of biological neurons, seen in the bottom scheme (source: Wikimedia), which are arranged in far more complex patterns. The parameters of an artificial neural network are the configuration of its interconnections and the parameters of each neuron. Neuron parameters can be optimised from experimental data using numerical methods. The NN shown can be used to model a static MIMO system with two inputs and two outputs, or a dynamic system with one input and two outputs if $x_2(t) = x_1(t - Ts)$, in which case it provides a non-linear difference equation model with sampling time T_s . NNs are typically black box models, and parameters are not expected to have any physical meaning at all; recently, however, significant efforts in neural network interpretability have been advanced. We will not study NNs in these Lecture Notes.

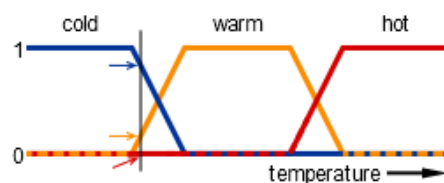


Figure 3.19: In Boolean logic, propositions are either true or false. These two cases correspond respectively to logical values 1 and 0. In fuzzy logic, all intermediate logical values can be used. The plot above shows an example of this (source: Wikimedia). For the temperature shown by the grey line, proposition “temperature is hot” has the logical value 0, proposition “temperature is warm” has the logical value 0.2, and proposition “temperature is cold” has the logical value 0.8. This type of logic can then be used to build models, both static and dynamic. We will not study fuzzy logic or fuzzy models in these Lecture Notes.



Figure 3.20: The Stansted Airport Transit System conveys passengers between Terminals 1 and 2 of Stansted Airport, United Kingdom (source: Wikimedia). Vehicles have no driver. They stop, open doors, close doors, and move between terminals automatically.

plant planta

process processo

reference referência

sampling frequency frequência de amostragem

sampling time tempo de amostragem

signal sinal

single input entrada única

single output saída única

static estático

stochastic estocástico

white box model modelo de caixa branca

Exercises

1. Answer the following questions for each of the mechatronic systems below:

- What are its outputs?
- What are its inputs?
- Is the system SISO or MIMO?
- Which of the inputs can be manipulated, if any?
- Is it a static or a dynamic system?
- Is it time varying or time invariant?
- Is the system continuous or digital?

(a) An automated train system, as seen in Figure 3.20.

(b) A power wheelchair, as seen in Figure 3.21.

(c) A motorboat, as seen in Figure 3.22.

(d) A rigged ship, as seen in Figure 3.22.

(e) A submarine, as seen in Figure 3.23.

(f) A space probe, as seen in Figure 3.24.

(g) A robotic arm, as seen in Figure 3.25.

2. Use the Laplace transform to solve (3.9) for the situation starting at a time when the country's population is 10 million inhabitants. Find the population for $t = 1, 2, 3, \dots$ years. Then use (3.11) to find the evolution of the population starting with year $k = 1$ (corresponding to $t = 0$ years) when the country's population is 10 million inhabitants. Find the population for $k = 2, 3, 4, \dots$ and compare the results with those obtained with (3.9).



Figure 3.21: Physicist Stephen Hawking (1942 — †2018) attending a scientific conference in 2001 (source: Wikimedia).



Figure 3.22: Left: a motorboat with an outboard motor at Zanzibar, Tanzania (source: Wikimedia). Right: Portuguese Navy school ship Sagres (formerly Brazilian school ship Guanabara, formerly German school ship Albert Leo Schlageter; source: Wikimedia).



Figure 3.23: The Portuguese Navy submarine Tridente, of the Tridente class, propelled by a low noise skew back propeller and powered by hydrogen–oxygen fuel cells (source: Wikimedia).



Figure 3.24: Astronomer Carl Sagan (1934 – †1996) with a model of one of the two Viking landers, space probes that descended on Mars in 1976 and worked until 1980 and 1982 (source: Wikimedia). Descent speed was controlled by deploying a parachute and launching three retrorockets (one on each leg) to ensure a soft landing. The descent control system employed an inertial reference unit, four gyroscopes, a radar altimeter, and a landing radar.



Figure 3.25: Two KUKA LWR IV robotic arms extant at the Control, Automation and Robotics Laboratory of Instituto Superior Técnico, Universidade de Lisboa, Portugal. Each robot has seven rotational joints. (Source: Professor Jorge Martins.)

Chapter 4

Modelling mechanical systems

Lex I.

Corpus omne perseverare in statu suo quiescendi vel movendi uniformiter in directum, nisi quatenus a viribus impressis cogitur statum illum mutare. (...)

Lex II.

Mutationem motus proportionalem esse vi motrici impressæ, & fieri secundum lineam rectam qua vis illa imprimitur. (...)

Lex III.

Actioni contrariam semper & æqualem esse reactionem: sive corporum duorum actiones in se mutuo semper esse æquales & in partes contrarias dirigi.

Isaac NEWTON (1643 — †1727), *Philosophiæ Naturalis Principia Mathematica* (1687), Axiomata sive Leges Motus

Ut tensio sic vis; That is, The Power of any Spring is in the same proportion with the Tension thereof: That is, if one power stretch or bend it one space, two will bend it two, and three will bend it three, and so forward.

Robert HOOKE (1635 — †1703), *Lectures de Potentia Restitutiva Or of Spring Explaining the Power of Springing Bodies* (1678)

In this and the following chapters, we will pass in review the basic concepts of system modelling, for different types of components. In this chapter, we concentrate upon mechanical components. Surely you will have already learned, if not all, at least most of these subjects in other courses. However, there are two reasons why a brief review is convenient at this point of your studies:

1. We will systematically resort to the Laplace transform to study dynamic systems, and after seeing too many equations with variable s it is easy to forget that we keep talking about real things — namely, in this course, mechatronic systems that are all around us in our daily life.
2. This is a good time to stress the similarities between apparently very different systems that can be described by the very same equations. We will see that thinking of any system as an energy converter helps to see those parallels.

4.1 Modelling the translation movement

Mechanical systems with movement along a straight line can usually be modelled using three components with the respective three equations:

1. A **mass**. This component stores energy under the form of kinetic energy. *Mass*
To model them, apply Newton's second law (which you can read in Latin *Newton's second law* at the beginning of this chapter):

$$\sum F = \frac{d}{dt} (m(t) \dot{x}(t)) \quad (4.1)$$



Figure 4.1: Usual translation springs. Left: helical or coil spring; centre: volute spring; right: leaf spring. (Source: Wikimedia.)

Here, $\sum F$ is the sum of all forces applied on the mass $m(t)$, which is at position $x(t)$. (Product $m(t)\dot{x}(t)$, as you know, is called **momentum**.) Because we are assuming a movement of translation, we need not bother to use vectors, but the forces must be applied along the direction considered; if not, their projection onto the said direction must be used. And, as we said in Section 3.1, we will only consider LTI systems; since the mass is, in (4.1), a parameter, this restriction means that it will not change with time, and so we are left with

Momentum

$$\sum F = m\ddot{x}(t) \quad (4.2)$$

A mass is usually represented by m or M .

Spring

2. A **spring**. This is a mechanical device that stores energy under the form of elastic potential energy (see Figure 4.1). A translation spring usually follows **Hooke's law** (which you can read in Latin and English at the beginning of this chapter):

Hooke's law

$$F = k(x_1 - x_2) \quad (4.3)$$

Here, F is the force exerted by the spring, x_1 and x_2 are the positions of the extremities of the spring (defined so that $x_1 - x_2$ is the variation in length of the spring measured from the repose length), and k is the spring constant. This constant is usually represented by k or K , and its SI units are N/m. Force F opposes the relative movement of the extremities that is its cause (i.e. force F contracts the spring when it is stretched, and stretches the spring when it is compressed).

Damper

Viscous damping

3. A **damper**. This is a mechanical device that dissipates energy (see Figure 4.2). The most usual model for dampers is **viscous damping**:

$$F = c(\dot{x}_1 - \dot{x}_2) \quad (4.4)$$

Here, F is the force exerted by the damper, \dot{x}_1 and \dot{x}_2 are the velocities of the extremities of the spring (defined so that $\dot{x}_1 - \dot{x}_2$ is the relative velocity of the extremities of the damper), and c is the damping constant. This constant is usually represented by c , C , b or B , and its SI units are N s/m. Force F opposes the relative movement of the extremities that is its cause.

Model (4.4) can also be used to model unintended energy dissipation, such as that due to friction. Notice that since energy dissipation is ubiquitous even a mechanical system consisting only of a mass and a spring will be more exactly modelled by a mass, a spring, and a damper, the latter to account for energy dissipation.

Remark 4.1. Unlike (4.2), Hooke's law (4.3) is often only an approximate model of the phenomenon it addresses. There are three ways in which reality usually deviates from (4.3).

1. The relation between force and variation in length can be non-linear. In any case, as long as the relation is continuous and has a continuous derivative, a linear approximation will be valid in a limited range of length variations (see Figure 4.3).

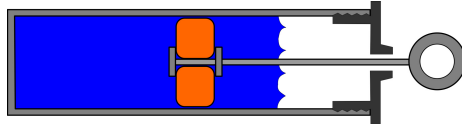


Figure 4.2: Dashpot damper (source: Wikimedia). There are other types of dampers. This one, because it contains a viscous fluid, follows (4.4) rather closely.

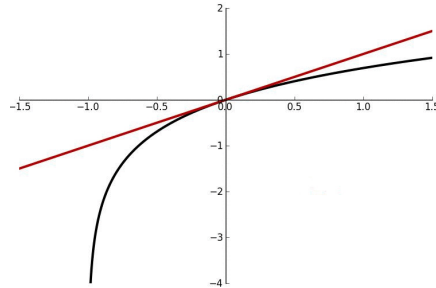


Figure 4.3: A linear approximation of a continuous function with a continuous derivative provides good results in some limited range (source: Wikimedia).

2. Springs that have a different behaviour for positive variations of length ($x > 0$, extension) and negative variations of length ($x < 0$, compression) are not uncommon.
3. In any case, Hooke's law is obviously valid only for a limited range of length variations. \square

Example 4.1. A stainless steel helicoidal spring is 10 cm long. When a traction force of 10 N is applied, its length increases to 12 cm. What force must be applied so that its length increases to 15 cm? What force must be applied so that its length increases to 40×10^3 km?

A length increase of 2×10^{-2} m corresponds to a 10 N force, so $k = \frac{10}{2 \times 10^{-2}} = 500$ N/m. The answer to the first question, when $x = 5$ cm = 5×10^{-2} m, is $F = 500 \times 5 \times 10^{-2} = 25$ N. Or, alternatively, since the length increase is to be $\frac{15-10}{12-10} = 2.5$ times larger than in the original situation, the force should also be 2.5 times larger, i.e. $2.5 \times 10 = 25$ N.

In the second case, it should be obvious that a 10 cm helicoidal spring cannot be stretched to a length which is roughly the perimeter of the Earth. You should not have to calculate the ludicrous result $F = 500 \times 40 \times 10^6 = 2 \times 10^{10}$ N = 20 GN obtained applying the linear relation (4.3) to realise that the spring will surely break well before such a force is applied. You should have by now seen a sufficient number of diagrams such as the one in Figure 4.4 to realise this at once, without even having to look for the yield strength of stainless steel and coming up with an educated guess for the spring's cross-sectional area. \square

Remark 4.2. Our models are approximations of reality. They are valid only for limited ranges of parameters. These important truths cannot be overstated. \square

Remark 4.3. Viscous damping (4.4) is another model of reality that very often is only a rough approximation. Dashpot dampers such as the ones in Figure 4.2 follow this law more closely than other damping phenomena, where damping may be non-linear or, if linear, proportional to another derivative of position x (some damping models even use fractional orders of differentiation). In any case, it is obvious that after a while x will reach its end of stroke, and (4.4) will no longer apply. \square

End of stroke

Combining (4.2)–(4.4) with Newton's third law — which states that when a body exerts a force on another, this latter body exerts an equal force, but opposite in direction, on the first body —, it is possible to find the differential equations that model translation mechanical systems.

Newton's third law

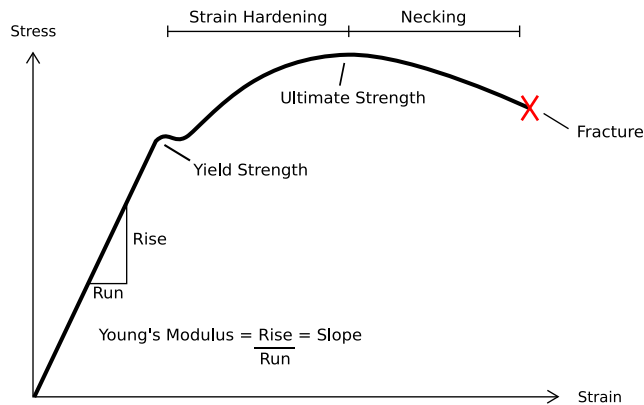


Figure 4.4: Schematic stress-strain curve of steel (source: Wikimedia).

Example 4.2. One of the most simple, but also most useful, mechanical models is the so-called mass–spring–damper system, which can be used to model the behaviour of a mass, on which a force is applied, connected to an inertial referential by a spring and a damper (remember that any real spring also has some damping, so, even in the absence of a dashpot damper or a similar device, energy dissipation must be accounted for). See Figure 4.5. This model can be applied to many systems, among which the vertical behaviour of a car’s suspension (for which of course far more accurate, and complex, models can also be used — see Figure 4.6). *Mass–spring–damper
tem*

Since one of the extremities of the spring is fixed, the force that it exerts on M is

$$F_K(t) = -K x(t) \quad (4.5)$$

or, omitting the dependence on time, $F_K = -K x$. There is a minus sign because, when x increases, the force on M opposes the increase of x . Similarly, the force exerted on M by the damper is

$$F_B(t) = -B \dot{x}(t) \quad (4.6)$$

or $F_B = -B \dot{x}$ to simplify. There is a minus sign because, when x increases, \dot{x} is positive, and the force on M opposes the increase of x . Thus

$$F(t) - K x(t) - B \dot{x}(t) = M \ddot{x}(t) \quad (4.7)$$

We will now assume that initial conditions are zero. Applying the Laplace transform,

$$F(s) - KX(s) - BsX(s) = Ms^2X(s) \quad (4.8)$$

which we can rearrange as

$$\frac{X(s)}{F(s)} = \frac{1}{Ms^2 + Bs + K} \quad \square \quad (4.9)$$

Transfer function

The form (4.9) in which the model of the mass–spring–damper system was put is called **transfer function**. It is very practical for the resolution of dynamic models.

Definition 4.1. Given a SISO system modelled by a differential equation, its transfer function is the ratio of the Laplace transform of the output (in the numerator) and the Laplace transform of the input (in the denominator), assuming that all initial conditions are zero. \square

Remark 4.4. When you see a transfer function, never forget that it is nothing but a differential equation under disguise. The transfer function is a rational function in s , which conceals a dynamic relation in time (or a relation in space, if the differential equation has derivatives in space rather than in time). \square

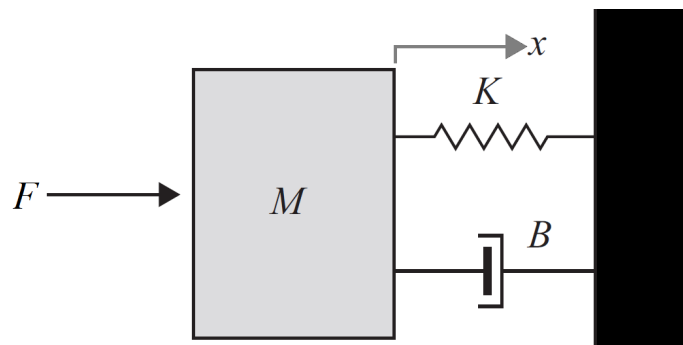


Figure 4.5: A mass–spring–damper system, with mass M , spring constant K , and damping coefficient B .



Figure 4.6: Independent suspension of a car's wheel (source: Wikimedia). The spring is clearly visible. The damper can be seen inside the coils of the spring. Even when the suspension does not consist of one spring and one damper, a mass–spring–damper model can be used, as nearly all suspensions have a spring-like restitution force and some sort of energy dissipation.

Remark 4.5. Notice that it is necessary to assume zero initial conditions to obtain a transfer function. Otherwise, additional terms would appear, and it would be impossible to isolate on one side of the equation the ratio of the Laplace transforms of the output and the input. We will further study this subject in Chapter 8. \square

Example 4.3. Let us find the transfer functions corresponding to (4.2)–(4.4) — i.e. to a mass, to a spring, and to a damper. For the spring and the damper, consider that extremity x_2 is fixed. Assume that the output is position $X(s)$, and force $F(s)$ is the input. Then

$$\frac{X(s)}{F(s)} = \frac{1}{ms^2} \quad (4.10)$$

$$\frac{X(s)}{F(s)} = \frac{1}{k} \quad (4.11)$$

$$\frac{X(s)}{F(s)} = \frac{1}{cs} \quad \square \quad (4.12)$$

Since transfer functions are functions of s , they are usually represented by one capital letter, such as F or G ; when $F(s)$ is used to represent a transfer function, care must be taken not to use the same letter to represent the Laplace transform $F(s)$ of a force $f(t)$.

Example 4.4. Suppose that force $F(t) = \sin(t)$ is applied to the system in Figure 4.5, in which $M = 1$ kg, $B = 3.5$ Ns/m, $K = 1.5$ N/m. What is the output $x(t)$?

The system's transfer function is

$$G(s) = \frac{X(s)}{F(s)} = \frac{1}{s^2 + 3.5s + 1.5} = \frac{1}{(s + 3)(s + 0.5)} \quad (4.13)$$

We have

$$F(s) = \frac{1}{s^2 + 1} \quad (4.14)$$

and thus

$$\begin{aligned} X(s) &= \overbrace{\frac{1}{s^2 + 1}}^{\mathcal{L} \text{ of the input}} \overbrace{\frac{1}{(s + 3)(s + 0.5)}}^{\text{transfer function}} = \frac{as + b}{s^2 + 1} + \frac{c}{s + 3} + \frac{d}{s + 0.5} \\ &= \frac{(as + b)(s^2 + 3.5s + 1.5) + c(s^2 + 1)(s + 0.5) + d(s^2 + 1)(s + 3)}{(s^2 + 1)(s + 3)(s + 0.5)} \\ &= \frac{s^3(a + c + d) + s^2(3.5a + b + 0.5c + 3d) + s(1.5a + 3.5b + c + d) + (1.5b + 0.5c + d)}{(s^2 + 1)(s + 3)(s + 0.5)} \end{aligned} \quad (4.15)$$

whence

$$\begin{aligned} \begin{cases} a + c + d = 0 \\ 3.5a + b + 0.5c + 3d = 0 \\ 1.5a + 3.5b + c + d = 0 \\ 1.5b + 0.5c + 3d = 1 \end{cases} &\Leftrightarrow \begin{cases} c + d = -a \\ 3.5a - 0.5b = -1 \\ 0.5a + 3.5b = 0 \\ c + 6d = 2 - 3b \end{cases} \Leftrightarrow \begin{cases} c + d = -a \\ 50b = 2 \\ a = -7b \\ c + 6d = 2 - 3b \end{cases} \\ \Leftrightarrow \begin{cases} c + d = \frac{7}{25} \\ b = \frac{1}{25} \\ a = -\frac{7}{25} \\ c + 6d = \frac{47}{25} \end{cases} &\Leftrightarrow \begin{cases} c = \frac{7}{25} - d \\ - \\ - \\ 5d = \frac{40}{25} \end{cases} \Leftrightarrow \begin{cases} c = -\frac{1}{25} \\ - \\ - \\ d = \frac{8}{25} \end{cases} \end{aligned} \quad (4.16)$$

Finally,

$$\begin{aligned} x(t) &= \mathcal{L}^{-1} \left[\frac{-\frac{7}{25}s}{s^2 + 1} + \frac{\frac{1}{25}}{s^2 + 1} + \frac{-\frac{1}{25}}{s + 3} + \frac{\frac{8}{25}}{s + 0.5} \right] \\ &= -\frac{7}{25} \cos(t) + \frac{1}{25} \sin(t) - \frac{1}{25} e^{-3t} + \frac{8}{25} e^{-0.5t} \quad \square \end{aligned} \quad (4.17)$$

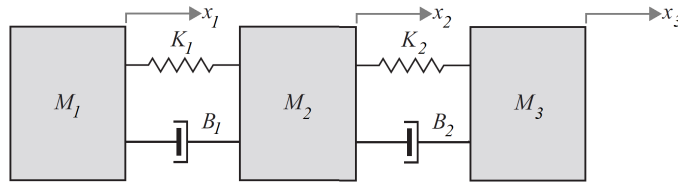


Figure 4.7: The system from Example 4.5, modelled by (4.20).

Example 4.5. Generalise the mass–spring–damper system of Example 4.2 to include three masses connected by springs and dampers as seen in Figure 4.7. The forces exerted by these components will be

$$\begin{cases} f_{K_1} = K_1(x_2 - x_1) \\ f_{K_2} = K_2(x_3 - x_2) \\ f_{B_1} = B_1(\dot{x}_2 - \dot{x}_1) \\ f_{B_2} = B_2(\dot{x}_3 - \dot{x}_2) \end{cases} \quad (4.18)$$

Applying Newton’s law to the masses, we get

$$\begin{cases} K_1(x_2 - x_1) + B_1(\dot{x}_2 - \dot{x}_1) = M_1\ddot{x}_1 \\ K_2(x_3 - x_2) + B_2(\dot{x}_3 - \dot{x}_2) - K_1(x_2 - x_1) - B_1(\dot{x}_2 - \dot{x}_1) = M_2\ddot{x}_2 \\ -K_2(x_3 - x_2) - B_2(\dot{x}_3 - \dot{x}_2) = M_3\ddot{x}_3 \end{cases} \quad (4.19)$$

Finally, the mathematical model of the system is

$$\begin{cases} M_1\ddot{x}_1 + K_1(x_1 - x_2) + B_1(\dot{x}_1 - \dot{x}_2) = 0 \\ M_2\ddot{x}_2 + K_1(x_2 - x_1) + B_1(\dot{x}_2 - \dot{x}_1) + K_2(x_2 - x_3) + B_2(\dot{x}_2 - \dot{x}_3) = 0 \\ M_3\ddot{x}_3 + K_2(x_3 - x_2) + B_2(\dot{x}_3 - \dot{x}_2) = 0 \end{cases} \quad (4.20)$$

To find a transfer function from the equations above, we would have to know which of the three positions x_1 , x_2 and x_3 is the input and which is the output. \square

Remark 4.6. Remember that it is somewhat irrelevant if positive displacements are assumed to be in one direction or the other. In the example above, positive displacements were arbitrarily assigned to the direction from the left to the right; the opposite could have been assumed, and signs would then be changed in such a way that the resulting model would still be correct. \square

4.2 Simulating transfer functions in MATLAB

There are two ways of creating a transfer function with MATLAB:

- `tf` creates a transfer function, represented by two vectors with the coefficients of the polynomials in the numerator and in the denominator (in decreasing order of the exponent);
- `s = tf('s')` creates the Laplace transform variable s , which can then be manipulated using algebraic operators.

Example 4.6. Transfer function (4.13) from Example 4.4 can be created as `MATLAB’s command tf`

```
>> G = tf(1,[1 3.5 1.5])
```

```
G =
```

```

      1
-----
s^2 + 3.5 s + 1.5
```

Continuous-time transfer function.

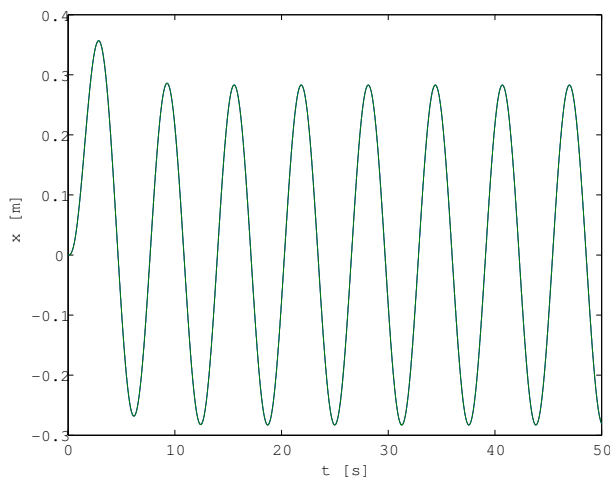


Figure 4.8: Results of Example 4.4.

or else as

```
>> s = tf('s')
```

```
s =
```

```
s
```

Continuous-time transfer function.

```
>> G = 1/(s^2+3.5*s+1.5)
```

```
G =
```

```

      1
-----
s^2 + 3.5 s + 1.5
```

Continuous-time transfer function.

□

Simulation

Command `lsim` (linear simulation) uses numerical methods to solve the differential equation represented by a transfer function for a given input. In other words, it **simulates** the LTI represented by the transfer function.

MATLAB's command `lsim` **Example 4.7.** The output found in Example 4.4 can be obtained and displayed as follows, if transfer function $G(s)$ has been created as above:

```
>> t = 0 : 0.01 : 50;
>> f = sin(t);
>> x = lsim(G, f, t);
>> figure, plot(t, x, t, -7/25*cos(t)+1/25*sin(t)-1/25*exp(-3*t)+8/25*exp(-0.5*t))
>> xlabel('t [s]'), ylabel('x [m]')
```

See Figure 4.8. Notice that we plotted two curves: the first was created with `lsim`, the second is (4.17). As expected, they coincide (there is a small numerical difference, too small to show up in the plot), and only one curve can be seen. □

Remark 4.7. The result (4.17) from Example 4.4 is exact. So the second curve in Figure 4.8 only has those numerical errors resulting from the implementation of the functions. The first curve has the errors resulting from the numerical method with which the differential equation was solved. Of course, both curves are based upon the same transfer function, and thus will suffer from any errors that there may be in that model (e.g. imprecise values of parameters M , B , and K , or neglected non-linearities in the spring or the damper). Do you still remember Remark 4.2? □

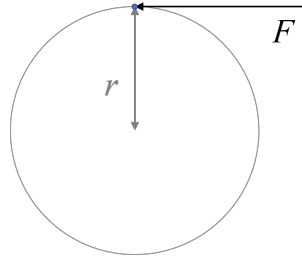


Figure 4.9: Tangential force F for a rotation of radius r applied on a point-like mass.

4.3 Modelling the rotational movement

Mechanical systems with movement of rotation can usually be modelled using three components with the respective three equations:

1. A **moment of inertia**. For these components, apply Newton's second law for rotation:

*Moment of inertia
Newton's second law for
rotation*

Theorem 4.1.

$$\sum \tau = \frac{d}{dt} (J(t) \dot{\omega}(t)) \quad (4.21)$$

Here, $\sum \tau$ is the sum of all torques applied on the moment of inertia $J(t)$, which is at angular position $\omega(t)$.

Proof. Let r be the radius of rotation for which are applied the tangential forces $\sum F$ that cause the torque (see Figure 4.9). Because $x = r\omega$, then $\dot{x} = r\dot{\omega}$, and Newton's second law (4.1) becomes

$$\sum F = \frac{d}{dt} (m(t) r \dot{\omega}(t)) \Leftrightarrow r \sum F = \frac{d}{dt} (m(t) r^2 \dot{\omega}(t)) \quad (4.22)$$

Because the torque τ of a force F is rF , and the moment of inertia J of mass m is mr^2 , (4.21) follows for a point-like mass. In the case of a distributed mass, integrating (4.22) over the volume occupied will then yield the desired result. \square

Corollary 4.1. Considering an LTI system, J will not change with time, and so we are left with

$$\sum \tau = J \ddot{\omega}(t) \quad (4.23)$$

A moment of inertia is usually represented by J or I (the latter letter is avoided when it can be confounded with an electrical current); its SI units are kg m^2 . A torque is usually represented by τ or T ; its SI units are N m .

2. A **torsion spring**. This is a mechanical device that stores energy (see Figure 4.10) and usually follows the **angular form of Hooke's law**:

*Torsion spring
Angular form of Hooke's
law*

$$\tau = \kappa (\omega_1 - \omega_2) \quad (4.24)$$

Here, τ is the torque exerted by the spring, ω_1 and ω_2 are the angular positions of the extremities of the spring, and κ is the spring constant. This constant is usually represented by the Greek character κ or K (to avoid confusion with a translation spring, for which a Latin character is used), and its SI units are N/rad .

3. A **rotary damper**, or **torsional damper**. The most usual model for this mechanical device that dissipates energy is **viscous damping**:

*Rotary (or torsional)
damper*

$$\tau = c (\dot{\omega}_1 - \dot{\omega}_2) \quad (4.25)$$

Here, τ is the torque exerted by the damper, $\dot{\omega}_1$ and $\dot{\omega}_2$ are the angular velocities of the extremities of the damper, and c is the damping constant.



Figure 4.10: A torsion spring mounted on a mousetrap (source: Wikimedia). Notice that Hooke's law for torsion springs will only apply in the range of angles comprised within the ends of stroke (say, from 0 rad to π rad).

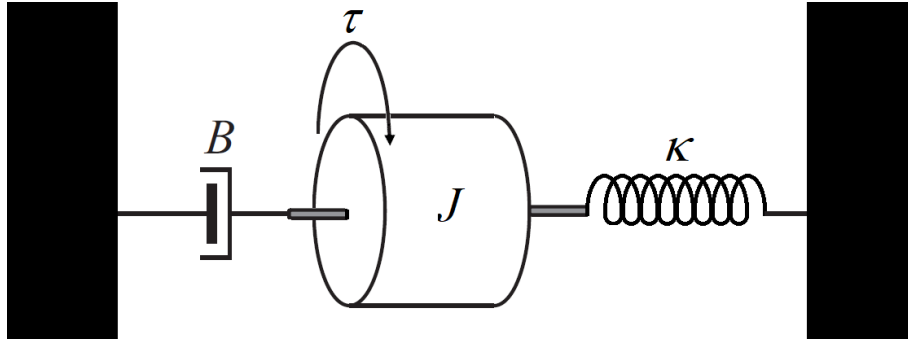


Figure 4.11: A mechanical system comprising a moment of inertia J , a torsion spring with constant κ , and a rotational damper with constant B .

This constant is usually represented by c , C , b or B , just like for the translation case, but its SI units are Ns/rad.

Also like (4.4), model (4.25) can be used to model unintended energy dissipation, such as that due to friction.

Remark 4.8. When dealing with rotation, take care with angular units. Confusion about values in degrees, radians, and rotations is a common source of error. This is true also for angular speed, angular velocity, angular spring constants, etc.. \square

Example 4.8. Consider the system in Figure 4.11. The torque exerted on J by the spring is

$$\tau_{\kappa}(t) = -\kappa \omega(t) \quad (4.26)$$

where ω is the rotation of J in the sense of rotation in which the applied torque τ is positive. The torque exerted by the damper is

$$\tau_B(t) = -B \dot{\omega}(t) \quad (4.27)$$

Thus

$$\tau(t) - \kappa \omega(t) - B \dot{\omega}(t) = J \ddot{\omega}(t) \quad (4.28)$$

Applying the Laplace transform (and assuming, once more, that all initial conditions are zero),

$$T(s) - \kappa \Omega(s) - Bs \Omega(s) = Js^2 \Omega(s) \Leftrightarrow \frac{\Omega(s)}{T(s)} = \frac{1}{Js^2 + Bs + \kappa} \quad \square \quad (4.29)$$

4.4 Energy, effort and flow

A comparison of transfer functions (4.9) and (4.29) shows us that different systems can have similar models. Actually, if the numerical values of M and J ,

and of B (in translation) and B (in rotation), and of K and κ , are the same, then the model will be the same.

As you surely know by now, this happens not only with mechanical systems, but also with systems of other, different types, as we shall see in the following chapters. One of the best ways of studying this parallelism is to see systems as energy converters, and energy E as the integral of the product of two variables, called **effort variable** e and **flow variable** f :

$$E(t) = \int_0^t e(t) f(t) dt \quad (4.30)$$

In other words, the product $e(t) \times f(t)$ is the instantaneous **power** $\dot{E}(t)$.

In the case of a translation movement,

$$\dot{E}(t) = F(t) \dot{x}(t) \Leftrightarrow E(t) = \int_0^t F(t) \dot{x}(t) dt \quad (4.31)$$

In the case of a rotation movement,

$$\dot{E}(t) = \tau(t) \dot{\omega}(t) \Leftrightarrow E(t) = \int_0^t \tau(t) \dot{\omega}(t) dt \quad (4.32)$$

We will consider force F and torque τ as the flow variable, and velocity \dot{x} and angular velocity $\dot{\omega}$ as the effort variable. But notice that it would make no difference if it were the other way round. In any case, their product will be the power. Both choices can be found in published literature.

The components of a system are the **effort accumulator**, the **flow accumulator**, and the **energy dissipator**, as seen in Table 4.1. For both accumulators, energy is the integral of accumulated flow or accumulated effort: elastic potential energy in the case of effort, and kinetic energy in the case of flow. The dissipator dissipates energy and it makes no difference whether it is kinetic or potential energy that it dissipates. Table 4.1 also includes the relations between these quantities.

Definition 4.2. A transfer function of a system that has the flux as input and the effort as output is called **impedance** of that system. A transfer function of a system that has the effort as input and the flux as output is called **admittance** of that system. Consequently, the admittance is the inverse of the impedance. \square

Transfer functions (4.10)–(4.12) can be rewritten so as to give the mechanical impedance of a mass, a spring, and a damper:

$$\frac{sX(s)}{F(s)} = \frac{1}{m.s} \quad (4.33)$$

$$\frac{sX(s)}{F(s)} = \frac{s}{k} \quad (4.34)$$

$$\frac{sX(s)}{F(s)} = \frac{1}{c} \quad (4.35)$$

4.5 Other components

Among the several other components that may be found in mechanical systems, the following ones, because of their general use and of their linearity, deserve a passing mention:

- **Cogwheels**, or gears. These wheels convert rotation movement into another rotation movement, but the wheels need not be in the same plane. For two external cogwheels, the sense of rotation is inverted, whereas for one external and one internal cogwheel it is not (see Figure 4.12).

Theorem 4.2. Let ω_1 and ω_2 be the angular positions of the two cogwheels, and r_1 and r_2 the respective radius. Then

$$\frac{\omega_1}{\omega_2} = \pm \frac{r_2}{r_1} \quad (4.36)$$

Table 4.1: Effort, flow, accumulators and dissipators in mechanical systems

	Translation mechanical system	SI	Rotation mechanical system	SI
effort e flow f	velocity \dot{x} force F	m s^{-1} N	angular velocity $\dot{\omega}$ torque τ	rad s^{-1} N m
effort accumulator accumulated effort $e_a = \int e dt$ relation between accumulated effort and flow $e_a = \varphi(f)$ accumulated energy $E_e = \int e_a df$	spring, with spring constant K position $x = \int \dot{x} dt$ position $x = \frac{1}{K}F$ elastic potential energy $E_e = \frac{1}{2K}F^2$	N/m m J	angular spring, with spring constant κ angular position $\omega = \int \dot{\omega} dt$ angular position $\omega = \frac{1}{\kappa}\tau$ elastic potential energy $E_e = \frac{1}{2\kappa}\tau^2$	N/rad rad J
flow accumulator accumulated flow $f_a = \int f dt$ relation between accumulated flow and effort $f_a = \varphi(e)$ accumulated energy $E_f = \int f_a de$	mass M momentum $p = \int F dt$ momentum $p = M\dot{x}$ kinetic energy $E_f = \frac{1}{2}M\dot{x}^2$	kg kg m s^{-1} J	moment of inertia J angular momentum $h = \int \tau dt$ angular momentum $h = J\dot{\omega}$ kinetic energy $E_f = \frac{1}{2}J\dot{\omega}^2$	kg m^2 $\text{kg m}^2 \text{s}^{-1}$ J
dissipator relation between effort and flow $e = \varphi(f)$ dissipated energy $E_d = \int f de$	damper, with damping constant b $\dot{x} = \frac{1}{b}F$ $E_d = \frac{1}{2}b\dot{x}^2$	N s/m J	rotary damper, with damping constant b $\dot{\omega} = \frac{1}{b}\tau$ $E_d = \frac{1}{2}b\dot{\omega}^2$	N s/rad J

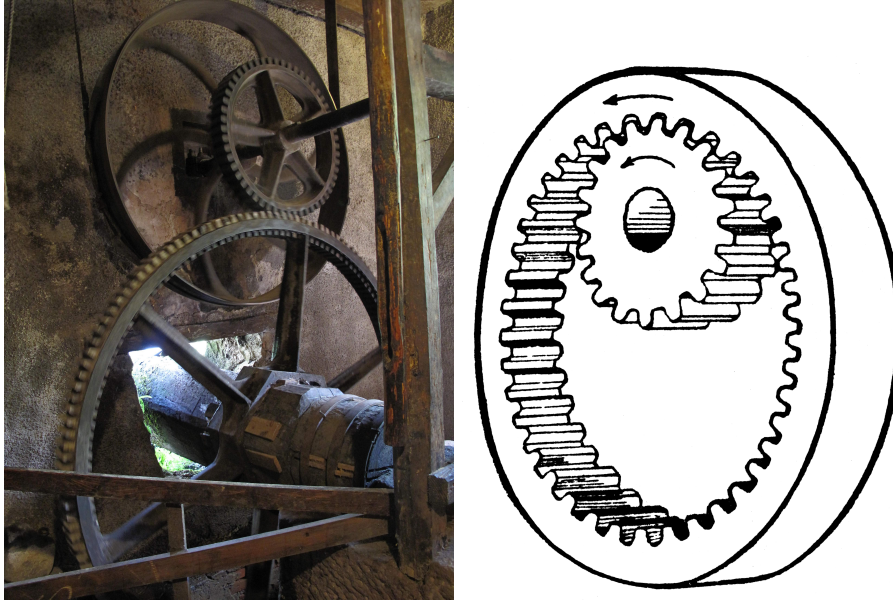


Figure 4.12: Left: two external cogwheels (source: Wikimedia). Right: two cogwheels, the outside one being an internal cogwheel (because the cogs are on the inside), and the inside one being an external cogwheel (because the cogs are on the outside; source: <https://etc.usf.edu/clipart>).

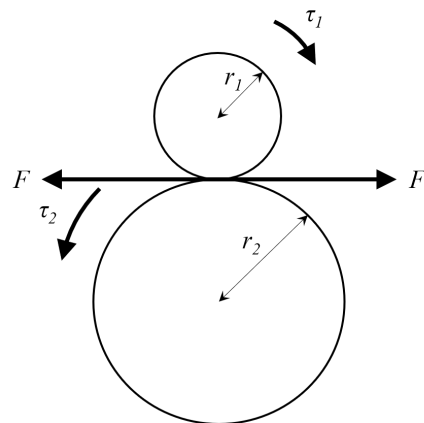


Figure 4.13: Proof of Theorem 4.3.

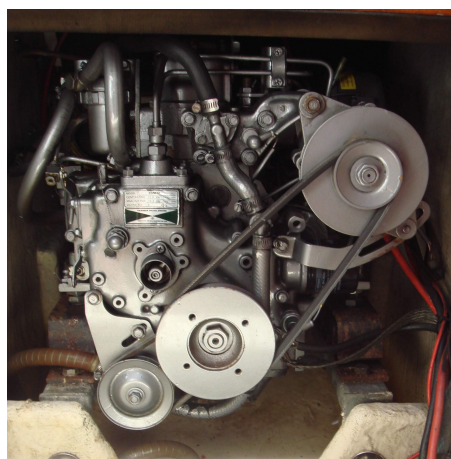


Figure 4.14: Transmission belts in a Diesel engine (source: Wikimedia).

Proof. This is a straightforward consequence of the linear movement of the point of contact between the two cogwheels being the same:

$$\omega_1 r_1 = \omega_2 r_2 \quad \square \quad (4.37)$$

Corollary 4.2. Let n_1 and n_2 be the numbers of cogs, or teeth, in the cogwheels. Then

$$\frac{\omega_1}{\omega_2} = \frac{\dot{\omega}_1}{\dot{\omega}_2} = \frac{\ddot{\omega}_1}{\ddot{\omega}_2} = \pm \frac{r_2}{r_1} = \pm \frac{n_2}{n_1} \quad (4.38)$$

Proof. The spacing of the teeth in both cogwheels has to be the same, otherwise they would not match. Thus the perimeters of the two wheels, $2\pi r_1$ and $2\pi r_2$, are directly proportional to n_1 and n_2 .

As to the angular velocity and acceleration, it suffices to differentiate (4.36). \square

Theorem 4.3. Let τ_1 and τ_2 be the torques of the two cogwheels. Then

$$\frac{\tau_1}{\tau_2} = \frac{r_1}{r_2} = \frac{n_1}{n_2} \quad (4.39)$$

Proof. See Figure 4.13. The forces that each wheel exerts on the other are equal. Thus

$$\tau_1 = F r_1 \Rightarrow F = \frac{\tau_1}{r_1} \quad (4.40)$$

$$\tau_2 = F r_2 \Rightarrow F = \frac{\tau_2}{r_2} \quad (4.41)$$

and the result follows. \square

Belt

- **A transmission belt.** It also converts rotation movement into another rotation movement:

$$\frac{\omega_1}{\omega_2} = \frac{\dot{\omega}_1}{\dot{\omega}_2} = \frac{\ddot{\omega}_1}{\ddot{\omega}_2} = \frac{r_2}{r_1} \quad (4.42)$$

Here ω_1 and ω_2 are the angular positions of the two wheels connected by the belt, and r_1 and r_2 are the respective radius. See Figure 4.14.

Rack and pinion

- **A rack and pinion.** It converts rotation movement into translation movement and vice-versa:

$$x = \omega r \Leftrightarrow \dot{x} = \dot{\omega} r \Leftrightarrow \ddot{x} = \ddot{\omega} r \quad (4.43)$$

Here x is the distance of the translation movement, r the radius of the wheel, and ω the angle of the rotation movement. See Figure 4.15.

Harmonic drive

- **A harmonic drive.** It converts rotation movement into another rotation movement, using an outside internal circular gear, inside which there is an external elliptical gear, to which an elliptical shaft is connected through a rolling bearing:

$$\frac{\omega_1}{\omega_2} = \frac{\dot{\omega}_1}{\dot{\omega}_2} = \frac{\ddot{\omega}_1}{\ddot{\omega}_2} = -\frac{n_1 - n_0}{n_0} \quad (4.44)$$

Here ω_2 is the angular position of the elliptical shaft inside the elliptical gear, ω_1 is the angular position of the shaft connected to the elliptical gear, n_1 is the number of teeth of the said gear, and n_0 is the number of teeth of the outside internal gear (which is fixed). See Figure 4.15.

Remark 4.9. Linear models (4.36)–(4.44) omit non-linear effects that may appear, such as backlash due to gaps between cogs (see Figure 4.16). These effects may sometimes be important but are not our concern here. \square

Remark 4.10. Models (4.36)–(4.44) are not only linear but also static (in the sense that outputs do not depend on past inputs, not in the sense that these components never move, of course). Non-linearities such as backlash make the components in fact dynamic. \square

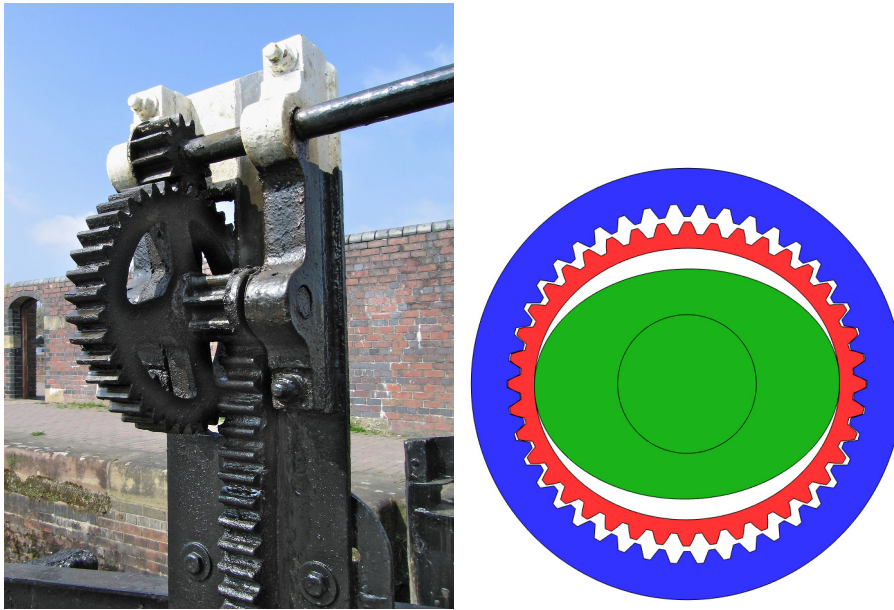


Figure 4.15: Left: rack and pinion in a canal gate; the rack is the linear element (in this case, vertical). Right: schematic of a harmonic drive. (Source: Wikimedia.)

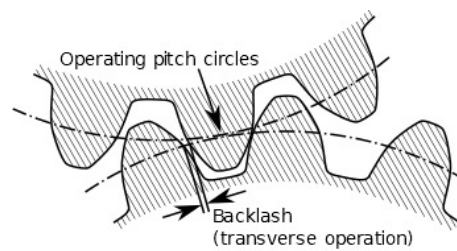


Figure 4.16: Backlash (source: Wikimedia).

Glossary

Als Storm weer bijkomt is het eerste wat hij ziet het vriendelijke gezicht van de jager.

„Wat wilt u dat ik voor u doe, god?”

„Ik... kan je verstaan, maar... maar spreek ik nu jouw taal of jij de mijne? En wat bedoel je met god? Ik ben geen god!”

„Maar natuurlijk bent u dat! Alles wijst erop. U sprak wartaal en begreep mij niet. Nou, het is duidelijk dat de goden de mensen niet begrijpen, anders hadden ze ze allang uitgeroeid! En nu u van de parel der kennis hebt gegeten, begrijpt u er nog niets van. Dat bewijst dat u gek bent! En de goden moeten gek zijn anders hadden ze de wereld nooit zo gemaakt als hij is.”

Don LAWRENCE (1928 — †2003), Martin LODEWIJK (1939 — ...), *Storm*, De kronieken van Pandarve 10, De piraten van Pandarve (1983)

accumulated effort potencial acumulado
accumulated flow fluxo acumulado
admittance admitância
cogwheel roda dentada
coil spring mola helicoidal
compression compressão
damper amortecedor
damping amortecimento
dashpot amortecedor viscoso
dissipator dissipador
effort accumulator acumulador de potencial
effort variable variável de potencial
end of stroke fim de curso
energy energia
extension extensão
flow accumulator acumulador de fluxo
flow variable variável de fluxo
gear roda dentada
harmonic drive redutor harmônico
helical spring mola helicoidal
impedance impedância
leaf spring mola de folhas, mola de lâminas
mass massa
mechanical impedance impedância mecânica
moment of inertia momento de inércia
momentum quantidade de movimento, momento linear
point-like mass massa pontual
power potência
rack and pinion pinhão e cremalheira
rotary damper amortecedor rotativo, amortecedor de torção
simulation simulação
spring mola
spring constant constante de mola
torque binário, torque (bras.)
torsion spring mola de torção
torsional damper amortecedor rotativo, amortecedor de torção
transfer function função de transferência
transmission belt correia de transmissão
volute spring mola de volutas, mola voluta

Exercises

1. Consider the system in Figure 4.17.
 - (a) Find the differential equations that model the system.

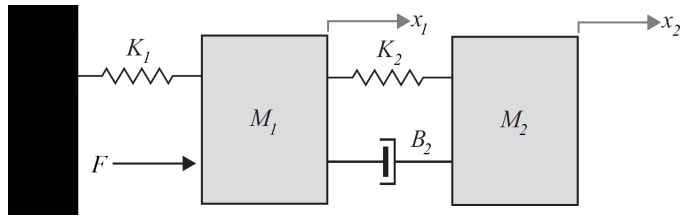


Figure 4.17: System of Exercise 1.

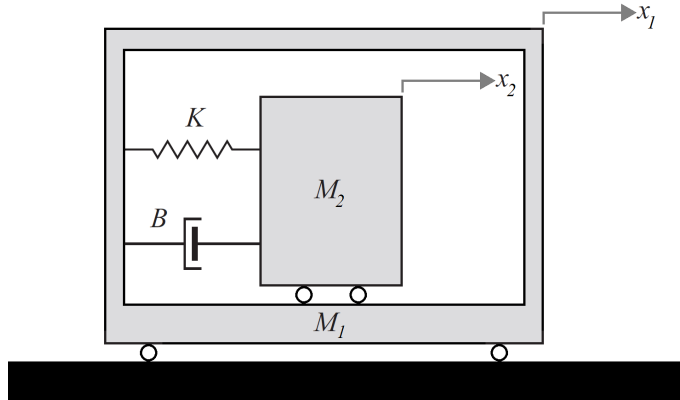


Figure 4.18: System of Exercise 2.

- (b) From the result above, knowing that $M_1 = 1$ kg, $M_2 = 0.5$ kg, $K_1 = 10$ N/m, $K_2 = 2$ N/m, and $B_2 = 4$ N s/m, find transfer function $\frac{X_2(s)}{F(s)}$.
2. Consider the system in Figure 4.18. The wheels have neglectable mass and inertia; they are included to show that the masses move without friction.
- (a) Find the differential equations that model the system.
- (b) From the result above, knowing that $M_1 = 100$ kg, $M_2 = 10$ kg, $K = 50$ N/m, and $B = 25$ N s/m, find transfer function $\frac{X_2(s)}{X_1(s)}$.
3. Consider the system in Figure 4.19. The wheels have neglectable mass and inertia; they are included to show that the masses move without friction.
- (a) Find the differential equations that model the system.
- (b) From the result above, knowing that $M_1 = M_2 = 1$ kg, $K = 5$ N/m, and $B = 10$ N s/m, find transfer function $\frac{X_1(s)}{F(s)}$.
- (c) For the same constants, find transfer function $\frac{X_2(s)}{F(s)}$.
4. Consider the system in Figure 4.20.
- (a) Find the differential equations that model the system.
- (b) From the result above, knowing that $M_1 = 1$ kg, $M_2 = 2$ kg, $M_3 = 3$ kg, $K_1 = 10$ N/m, $K_2 = 20$ N/m, $K_3 = 30$ N/m, $B_1 = 5$ N s/m, $B_2 = 10$ N s/m, and $B_3 = 15$ N s/m, find transfer function $\frac{X_1(s)}{F(s)}$.

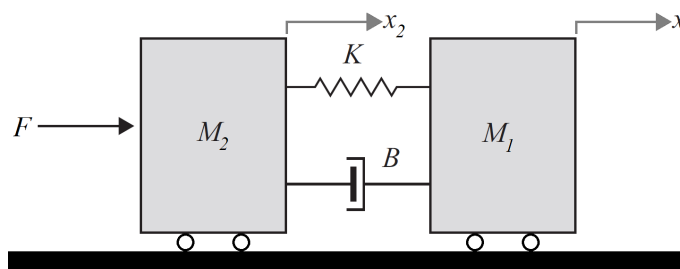


Figure 4.19: System of Exercise 3.

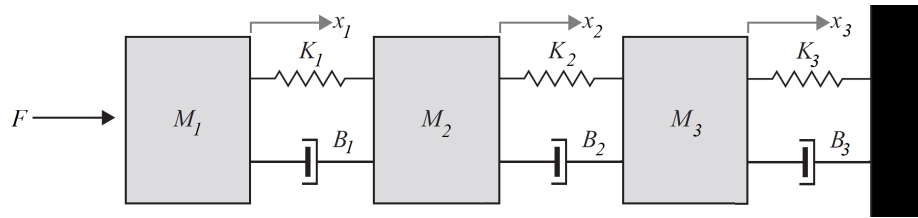


Figure 4.20: System of Exercise 4.

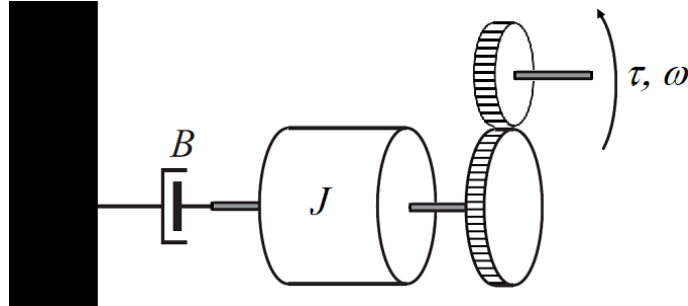


Figure 4.21: System of Exercise 5.

- (c) For the same constants, find transfer function $\frac{X_2(s)}{F(s)}$.
- (d) For the same constants, find transfer function $\frac{X_3(s)}{F(s)}$.
5. Consider the system in Figure 4.21. The cogwheels have neglectable moments of inertia, when compared to J . Let N_u be the number of cogs in the upper cogwheel, and N_l the number of cogs in the lower cogwheel.
- (a) Find the differential equations that model the system.
- (b) From the result above, knowing that $J = 50 \text{ kg m}^2$, $N_u = 20$, $N_l = 30$, and $B = 40 \text{ N s/m}$, find transfer function $\frac{\Omega(s)}{T(s)}$.
6. Consider the system in Figure 4.22. The pinion's centre is fixed, and its moment of inertia I includes the lever actuated by force F .
- (a) Find the differential equations that model the system.
- (b) From the result above, knowing that $r = 0.2 \text{ m}$, $I = 0.8 \text{ kg m}^2$, $m = 20 \text{ kg}$, $k = 1000 \text{ N/m}$, and $b = 480 \text{ N s/m}$, find transfer function $\frac{X(s)}{F(s)}$.
7. Consider the system in Figure 4.23. Force F is applied through a bar of neglectable mass, connected by the spring and the damper to mass m , affected by friction force F_a that follows the law of viscous damping with constant b_a . The bar has velocity v_F ; mass m has velocity F .
- (a) Find the differential equations that model the system.
- (b) From the result above, find transfer function $\frac{V_F(s)}{F(s)}$.
8. Consider the system in Figure 4.24. The position of mass M is $x(t)$.
- (a) Find the differential equations that model the system.
- (b) From the result above, find transfer function $\frac{X(s)}{T(s)}$.
9. For all the systems in the exercises above, find:
- (a) the effort variables;
- (b) the effort accumulators;
- (c) the flow variables;
- (d) the flow accumulators;

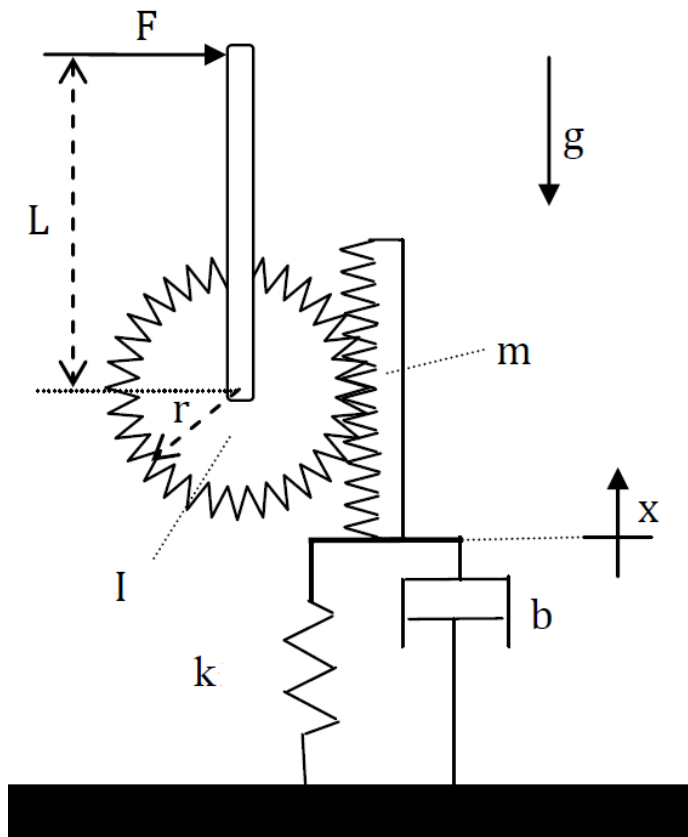


Figure 4.22: System of Exercise 6.

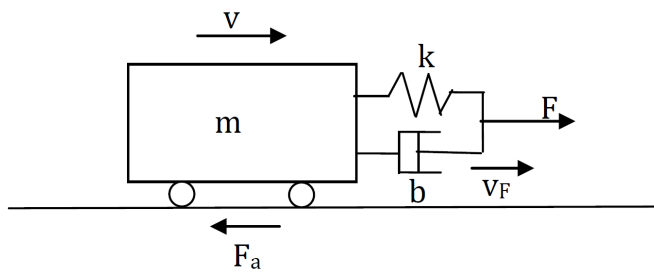


Figure 4.23: System of Exercise 7.

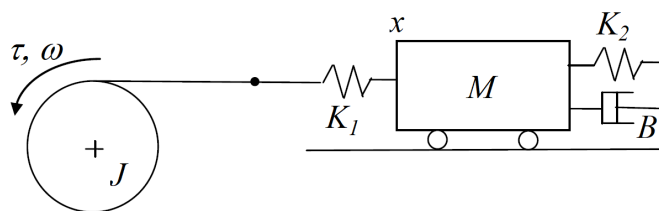


Figure 4.24: System of Exercise 8.

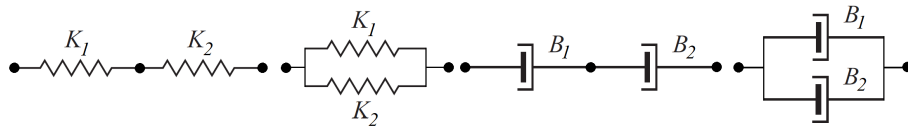


Figure 4.25: From left to right: two springs in series; two springs in parallel; two dampers in series; two dampers in parallel.

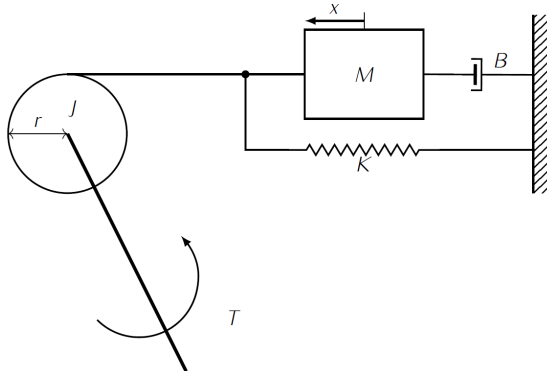


Figure 4.26: Mechanical system from Exercise 11.

- (e) the dissipators.
10. Find the $\frac{X(s)}{F(s)}$ transfer functions, as in (4.10)–(4.12), of the following systems (see Figure 4.25):
- two springs, with constants K_1 and K_2 , in series;
 - two springs, with constants K_1 and K_2 , in parallel;
 - two dampers, with constants B_1 and B_2 , in series;
 - two dampers, with constants B_1 and B_2 , in parallel.
11. Consider the mechanical system in Figure 4.26 comprising a mass M , a damper with viscous friction coefficient B and a spring with stiffness K . These elements are connected to a fixed pulley of radius r with inertia J . The system is driven by a torque T . Let θ be the rotation angle of the pulley.
- Model the system, writing the equations that describe its dynamics.
 - Obtain the transfer function considering the torque T as input and the mass position x as output.

Chapter 5

Modelling electrical systems

By that time Mordor was deservedly being called “smithy of the nations,” and it could trade its manufactured goods for any amounts of food from Khand and Umbar. Trading caravans went back and forth through the Ithilien Crossroads day and night, and more and more voices in Barad-dûr were saying that the country has had enough tinkering with agriculture, which was nothing but a net loss anyway, and the way to go was to develop what nobody else had — namely, metallurgy and chemistry. Indeed, the industrial revolution was well underway; steam engines toiled away in mines and factories, while the early aeronautic successes and experiments with electricity were the talk of the educated classes.

Kirill YESKOV (1956 — . . .), *The Last Ringbearer* (1999), I 3 (transl. Yisroel Markov, 2011)

This chapter addresses the modelling of electrical systems.

5.1 Passive components

The three simplest elements in an electrical circuit are:

1. A **resistor**. This component (see Figure 5.1) dissipates energy according to Ohm’s law: *Resistor*
Ohm’s law

$$R(t) = \frac{U(t)}{I(t)} \quad (5.1)$$

Here R is the resistance, U is the voltage (or tension, or electric potential difference), and I is the current. Notice that $U = U_1 - U_2$, where U_1 and U_2 are the voltages at the resistor’s terminals.

The resistance R of a uniform conductor is directly proportional to its length L and inversely proportional to its cross-section A :

$$R = \rho \frac{L}{A} \quad (5.2)$$

Proportionality constant ρ is called *resistivity*. This variation with length is used to build variable resistances, shown in Figure 5.2.

2. A **capacitor**. This component (see Figure 5.4) stores energy and its most usual model is *Capacitor*

$$U(t) = \frac{1}{C} Q(t) \quad (5.3)$$

where $Q(t)$ is the electrical charge stored, and C is the capacity. Since $I(t) = \frac{dQ(t)}{dt}$,

$$U(t) = \frac{1}{C} \int_0^t I(t) dt \quad (5.4)$$

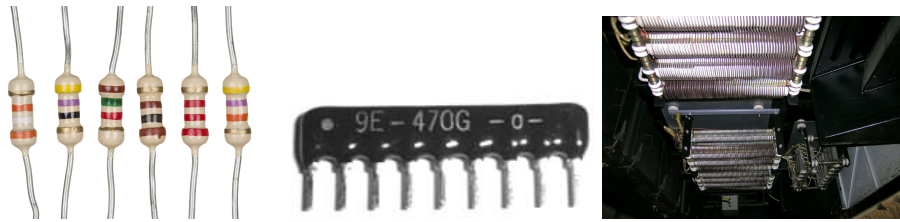


Figure 5.1: Different types of resistors. Left: individual resistors for use in electronics; centre: many resistors in one encasing; right: wirewound resistors for high tensions and currents in a train. (Source: Wikimedia.) There are still other types of resistors.



Figure 5.2: Potentiometers (or variable resistors, or rheostats) have a slider or a screw to move the position of a terminal, and thus the length of the resistor which is actually employed; resistance is proportional to this length (source: Wikimedia). See Figure 5.3.

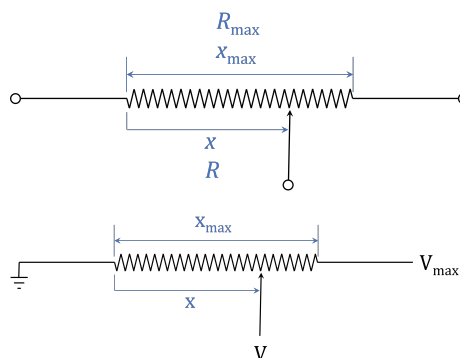


Figure 5.3: Top: a linear potentiometer with length x_{\max} and resistance R_{\max} , connected on the left and on the movable terminal at position x , has a resistance R proportional to the position: $\frac{x}{x_{\max}} = \frac{R}{R_{\max}} \Leftrightarrow R = R_{\max} \frac{x}{x_{\max}}$. Bottom: this potentiometer is used as a tension divider, grounded on the left and at V_{\max} on the right; the voltage at the terminal is given by $\frac{V}{V_{\max}} = \frac{R}{R_{\max}} = \frac{x}{x_{\max}} \Rightarrow V = V_{\max} \frac{x}{x_{\max}} \Leftrightarrow x = x_{\max} \frac{V}{V_{\max}}$.

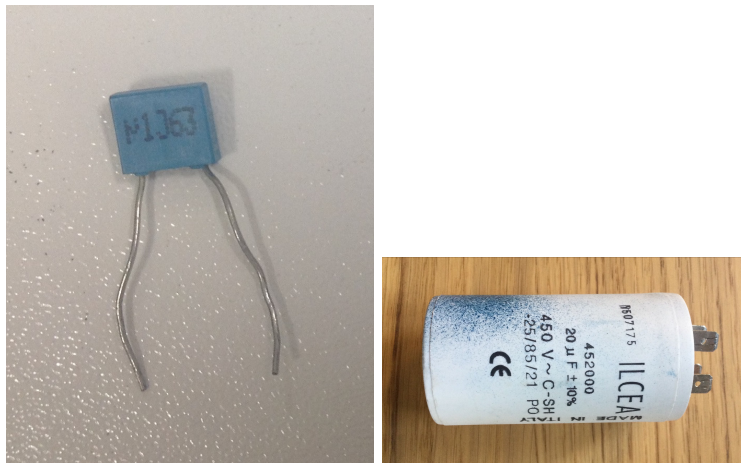


Figure 5.4: Left: 100 nF capacitor; right: 20 μF capacitor.



Figure 5.5: Inductor for electronic circuits.

Differentiating, we get

$$I(t) = C \frac{dU(t)}{dt} \quad (5.5)$$

3. An **inductor**. This component (see Figure 5.5) also stores energy and its most usual model is *Inductor*

$$I(t) = \frac{1}{L} \lambda(t) \quad (5.6)$$

where $\lambda(t) = \int_0^t U(t) dt$ is the flux linkage, and L is the inductance. Differentiating, we get

$$U(t) = L \frac{dI(t)}{dt} \quad (5.7)$$

The transfer functions of the resistor, the capacitor, and the inductor, corresponding to (5.1), (5.5), and (5.7), considering always tension $U(s)$ as the output and current $I(s)$ as the input, are

$$\frac{U(s)}{I(s)} = R \quad (5.8)$$

$$\frac{U(s)}{I(s)} = \frac{1}{Cs} \quad (5.9)$$

$$\frac{U(s)}{I(s)} = Ls \quad (5.10)$$

Remark 5.1. Notice that Ohm's law (5.1) or (5.8) corresponds to a static system. \square

Remark 5.2. It cannot be overstated that relations (5.8)–(5.10) are not followed by many components:

Table 5.1: Effort, flow, accumulators and dissipators in electrical systems

	Electrical system	SI
effort e	voltage U	V
flow f	current I	A
effort accumulator	inductor with induction L	H
accumulated effort $e_a = \int e dt$	flux linkage $\lambda = \int U dt$	Wb
relation between accumulated effort and flow $e_a = \varphi(f)$	flux linkage $\lambda = LI$	
accumulated energy $E_e = \int e_a df$	inductive energy $E_e = \frac{1}{2}LI^2$	J
flow accumulator	capacitor with capacity C	F
accumulated flow $f_a = \int f dt$	charge $Q = \int I dt$	C
relation between accumulated flow and effort $f_a = \varphi(e)$	charge $Q = CU$	
accumulated energy $E_f = \int f_a de$	capacitative energy $E_f = \frac{1}{2}CV^2$	J
dissipator	resistance R	Ω
relation between effort and flow $e = \varphi(f)$	$U = RI$	
dissipated energy $E_d = \int f de$	$E_d = \frac{1}{2}RI^2$	J

- Many resistances do not follow a linear relation between U and I such as (5.1), and are thus called non-ohmic resistors. Still, Ohm's law can be a good approximation in a limited range of values (see Figure 4.3 again). *Non-ohmic resistors*
- Many capacitors have variable capacity C , depending on the voltage applied. Others follow differential equations of fractional order (which we will study in Chapter 35).
- Inductances always have some resistance, which is often not neglectable. So their transfer function would more accurately be $R + Ls$.
- Even when (5.8)–(5.10) are accurately followed, this only happens for a limited range of values. Increase U or I too much, and any electrical component will cease to function (burn, melt...). What is too much depends on the particular component: there are components that cannot stand 1 V while others work at 10^4 V and more.

5.2 Energy, effort and flow

Because $\dot{E}(t) = U(t)I(t)$ and $E(t) = \int_0^t U(t)I(t) dt$, effort and flow variables are U and I . While either can once more play each of the roles, by universal convention,

- U is the effort variable,
- I is the flow variable, and thus
- the inductor is the effort accumulator,
- the capacitor is the flux accumulator,
- the resistor is the dissipator.

Table 5.1 sums up the passing information and relations.

Remark 5.3. Transfer functions (5.8)–(5.10) have the flux as input and the effort and output. They consequently give the impedance of the corresponding components. \square

Electrical impedance

Kirchoff's laws

Kirchoff's current law

- The current law states that the sum of the currents at a circuit's node is zero.

Kirchoff's voltage law

- The voltage law states that the sum of the voltages around a circuit's closed loop is zero.

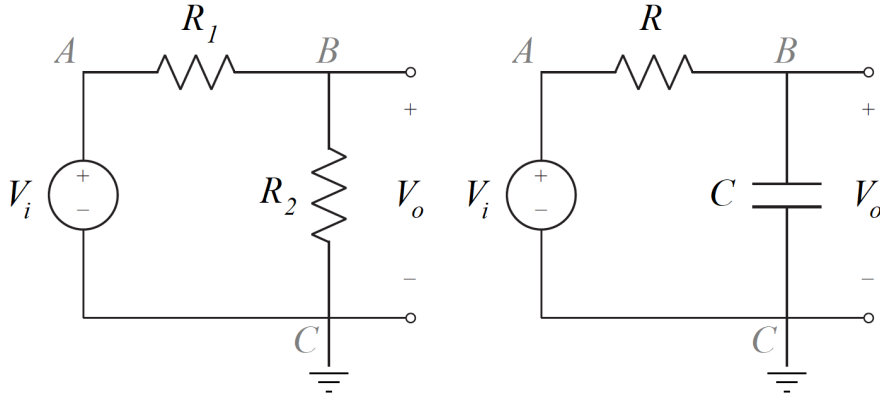


Figure 5.6: Left: voltage divider. Right: RC circuit.

voltage divider

Example 5.1. Consider the system in Figure 5.6 known as voltage divider. The input is $V_i(t)$ and the output is $V_o(t)$. Applying the current law at point B , we see that the current flowing from A to B must be the same that flows from B to C . Applying Ohm's law (5.1) to the two resistances, we see that

$$R_1 = \frac{V_B(t) - V_A}{I(t)} = \frac{V_o(t) - V_i(t)}{I(t)} \Rightarrow I = \frac{V_o - V_i}{R_1} \quad (5.11)$$

$$R_2 = \frac{V_C(t) - V_B(t)}{I(t)} = \frac{0 - V_o(t)}{I(t)} \Rightarrow I = \frac{-V_o}{R_2} \quad (5.12)$$

In the last equalities above, we dropped the dependence on t to alleviate the notation. Consequently,

$$(V_o - V_i)R_2 = -V_oR_1 \Leftrightarrow V_o(R_1 + R_2) = V_iR_2 \Leftrightarrow \frac{V_o}{V_i} = \frac{R_2}{R_1 + R_2} \quad (5.13)$$

Notice that this system is static, and from $\frac{V_o(t)}{V_i(t)} = \frac{R_2}{R_1 + R_2}$ we get $\frac{V_o(s)}{V_i(s)} = \frac{R_2}{R_1 + R_2}$. \square

Remark 5.4. Remember that, similarly to what happens with the positive direction of displacements in mechanical systems, it is irrelevant if a higher tension is presumed to exist to the left or to the right of a component. Current is always assumed to flow from higher to lower tensions; as long as equations are coherently written, if in end current turns out to be negative, this only means that it will flow the other way round. \square

Example 5.2. The transfer function of the system in Figure 5.6 known as RC circuit can be found in almost the same manner, thanks to impedances:

RC circuit

$$R = \frac{V_B(s) - V_A(s)}{I(s)} = \frac{V_o(s) - V_i(s)}{I(s)} \Rightarrow I = \frac{V_o - V_i}{R} \quad (5.14)$$

$$\frac{1}{Cs} = \frac{V_C(s) - V_B(s)}{I(s)} = \frac{0 - V_o(s)}{I(s)} \Rightarrow I = -V_oCs \quad (5.15)$$

In the last equalities above, we dropped the dependence on s to alleviate the notation. Consequently,

$$V_o - V_i = -V_oRCs \Leftrightarrow V_o(1 + RCs) = V_i \Leftrightarrow \frac{V_o}{V_i} = \frac{1}{1 + RCs} \quad (5.16)$$

Notice that this system is dynamic, and from $V_o(s)(1 + RCs) = V_i(s)$ we get $V_o(t) + RC \frac{dV_o(t)}{dt} = V_i(t)$. \square

Example 5.3. Both systems above are particular cases of the generic system *Two generic impedances* in Figure 5.7 with two impedances:

$$Z_1(s) = \frac{V_B(s) - V_A(s)}{I(s)} = \frac{V_o(s) - V_i(s)}{I(s)} \Rightarrow I = \frac{V_o - V_i}{Z_1} \quad (5.17)$$

$$Z_2(s) = \frac{V_C(s) - V_B(s)}{I(s)} = \frac{0 - V_o(s)}{I(s)} \Rightarrow I = \frac{-V_o}{Z_2} \quad (5.18)$$

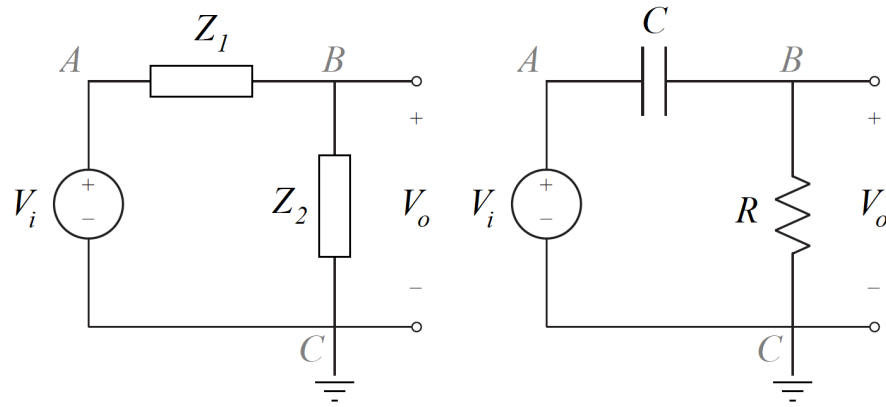


Figure 5.7: Left: generic electrical system with two impedances, of which the voltage divider (Figure 5.6), the RC circuit (Figure 5.6) and the CR circuit (to the right) are particular cases. Right: CR circuit.

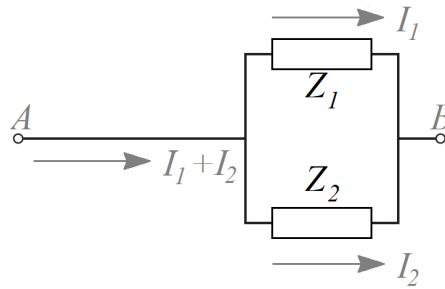


Figure 5.8: Two impedances in parallel.

Consequently,

$$(V_o - V_i)Z_2 = -V_o Z_1 \Leftrightarrow \frac{V_o}{V_i} = \frac{Z_2}{Z_1 + Z_2} \quad (5.19)$$

Replacing $Z_1(s) = R_1$ and $Z_2(s) = R_2$ in (5.19), we obtain (5.13).

Replacing $Z_1(s) = R$ and $Z_2(s) = \frac{1}{Cs}$ in (5.19), we obtain (5.16).

We can also obtain the transfer function of the case where the resistor and the capacitor are switched as also shown in Figure 5.7: when $Z_1(s) = \frac{1}{Cs}$ and $Z_2(s) = R$, we have $\frac{V_o(s)}{V_i(s)} = \frac{R}{R + \frac{1}{Cs}} = \frac{RCs}{1 + RCs}$. This is known as a CR circuit. \square

CR circuit

Impedances in series

Remark 5.5. We have incidentally shown that two impedances Z_1 and Z_2 in series correspond to a single impedance $Z = Z_1 + Z_2$, i.e. two impedances in series are summed, just as two resistances in series are. You know that two resistances R_1 and R_2 in parallel are equivalent to resistance $R = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2}}$, and something similar happens to two impedances in parallel (see Figure 5.8):

Impedances in parallel

$$Z_1 = \frac{V_B - V_A}{I_1} \Leftrightarrow I_1 = \frac{V_B - V_A}{Z_1} \quad (5.20)$$

$$Z_2 = \frac{V_B - V_A}{I_2} \Leftrightarrow I_2 = \frac{V_B - V_A}{Z_2} \quad (5.21)$$

$$Z = \frac{V_B - V_A}{I_1 + I_2} = \frac{V_B - V_A}{\frac{V_B - V_A}{Z_1} + \frac{V_B - V_A}{Z_2}} = \frac{1}{\frac{1}{Z_1} + \frac{1}{Z_2}} \quad (5.22)$$

Because of the parallelism between systems of different types, this is true for mechanical impedances as well (see Exercise 10 from Chapter 2), and for impedances of other types of systems we will study in the next chapters.

RLC circuit

Example 5.4. Consider the system in Figure 5.9 known as RLC circuit. The input is $V_i(t)$ and the output is $V_o(t)$. Applying the current law, we see that the current flowing from A to B must be the same that flows from B to C and

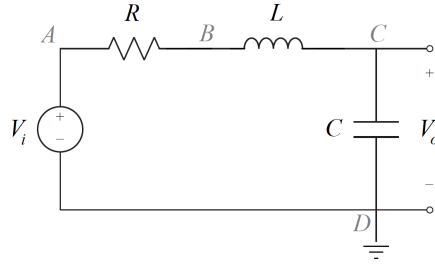


Figure 5.9: RLC circuit.

the same that flows from C to D . Then

$$\begin{cases} R = \frac{V_B(t) - V_A}{I(t)} = \frac{V_B(t) - V_i(t)}{I(t)} \Rightarrow V_B = RI + V_i \\ Ls = \frac{V_C(t) - V_B(t)}{I(t)} = \frac{V_o(t) - V_B(s)}{I(t)} \Rightarrow ILs = V_o - V_B \\ \frac{1}{Cs} = \frac{V_D(s) - V_C(s)}{I(s)} = \frac{0 - V_o(s)}{I(s)} \Rightarrow I = -V_oCs \end{cases} \quad (5.23)$$

We now replace the first equation in the second, and use it together with the third to get

$$\begin{aligned} ILs = V_o - RI - V_i &\Leftrightarrow I(R + Ls) = V_o - V_i & (5.24) \\ \Rightarrow \frac{V_o - V_i}{R + Ls} = -V_oCs &\Leftrightarrow V_o + V_oCRs + V_oCLs^2 = V_i \Leftrightarrow \frac{V_o}{V_i} = \frac{1}{CLs^2 + CRs + 1} \end{aligned}$$

From $V_o(s) + V_o(s)CRs + V_o(s)CLs^2 = V_i(s)$ we get

$$V_o(t) + CR \frac{dV_o(t)}{dt} + CL \frac{d^2V_o(t)}{dt^2} = V_i(t) \quad \square \quad (5.25)$$

Remark 5.6. We could have established (5.25) first, without using impedances:

$$\begin{cases} R = \frac{V_B(t) - V_A(t)}{I(t)} \Rightarrow I(t) = \frac{1}{R}V_B(t) - \frac{1}{R}V_i(t) \\ V_C(t) - V_B(t) = L \frac{dI(t)}{dt} \Rightarrow V_o(t) - V_B(t) = \frac{L}{R} \frac{dV_B(t)}{dt} - \frac{L}{R} \frac{dV_i(t)}{dt} \\ V_D(t) - V_C(t) = \frac{1}{C} \int I(t) dt \Rightarrow -\frac{dV_o(t)}{dt} = \frac{1}{C} \int I(t) dt \end{cases} \quad (5.26)$$

Replacing the first equation in the third, and then the result in the second,

$$-\frac{dV_o(t)}{dt} = \frac{1}{RC}V_B(t) - \frac{1}{RC}V_i(t) \Rightarrow V_B(t) = V_i(t) - RC \frac{dV_o(t)}{dt} \quad (5.27)$$

$$V_o(t) - V_i(t) + RC \frac{dV_o(t)}{dt} = \frac{L}{R} \left(\frac{dV_i(t)}{dt} - RC \frac{d^2V_o(t)}{dt^2} \right) - \frac{L}{R} \frac{dV_i(t)}{dt} \quad (5.28)$$

Rearranging terms in the last equality gives (5.25). Applying the Laplace transform, we then obtained transfer function (5.24). The results are of course the same. Notice that in both cases zero initial conditions were implicitly assumed (i.e. integrals were assumed to be zero at $t = 0$; in the case of the Laplace transform, this means that there is no $f(0)$ term in (2.41)). We will address this further in Chapter 9. \square

Remark 5.7. Transfer function (5.24) is similar to transfer functions (4.9) and (4.11). As you know, this is one case of a so-called electrical equivalent of a mechanical system, or of a mechanical equivalent of an electrical system. The notions of effort and flux make clear why this parallel between models exists: both consist of an effort accumulator, a flux accumulator, and a dissipator. But notice that the parallel is not complete: (4.9) has a flux as input and an accumulated effort as output; both the input and the output of (5.24) are efforts. \square

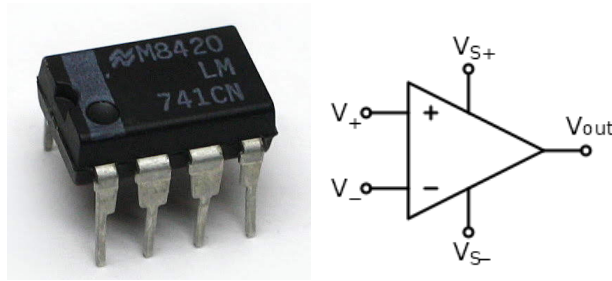


Figure 5.10: Left: an integrated circuit with a 741 OpAmp, one of the most usual types of OpAmps (source: Wikimedia). Other OpAmp types are manufactured in integrated circuits that have several OpAmps each, sharing the same power supply. Right: the symbol of the OpAmp (source: Wikimedia). Power supply tensions are often omitted in diagrams for simplicity, but never forget that an active component without power supply does not work.

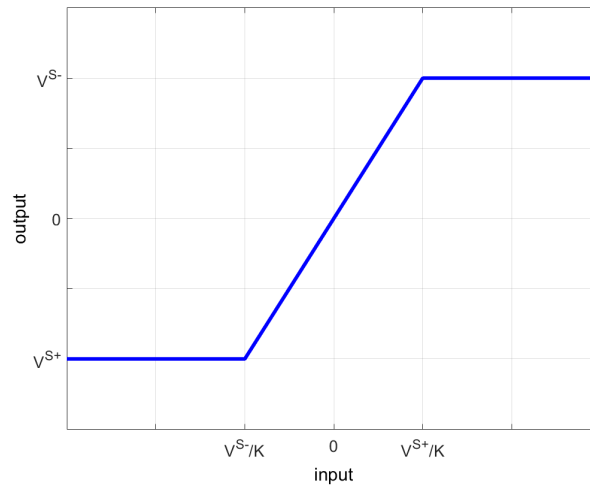


Figure 5.11: The output of an OpAmp.

5.3 The operational amplifier (OpAmp), an active component

Active components

The resistor, the capacitor and the inductor are called passive components because they do not need a source of energy to function. Components that need a source of energy to function are called active components. Among them are diodes and transistors, together with sensors that we will study in Chapter 13. A component we will study right away because of its importance is the **operational amplifier**, or in short the OpAmp.

Passive components

An OpAmp is an electronic component that presents itself as an integrated circuit (see Figure 5.10) and amplifies the difference between its two inputs V^- and V^+ :

$$V^{\text{out}} = K (V^+ - V^-) \quad (5.29)$$

The output V^{out} is limited to the power supply tensions:

$$V^{S-} \leq V^{\text{out}} \leq V^{S+} \quad (5.30)$$

No output if no power supply

As can be expected from the fact that the OpAmp is an active component, if no power is supplied, i.e. if the corresponding pins of the integrated circuit are disconnected and thus $V^{S+} = V^{S-} = 0$ V, then $V^{\text{out}} = 0$ V, i.e. there is no output. The gain of the OpAmp K should ideally be infinite; in practice it is very large, e.g. 10^5 or 10^6 . See Figure 5.11.

The other important characteristic of the OpAmp is that the impedance between its two inputs V^- and V^+ is very large. Ideally it should be infinite; in practice it is 2 M Ω or more.

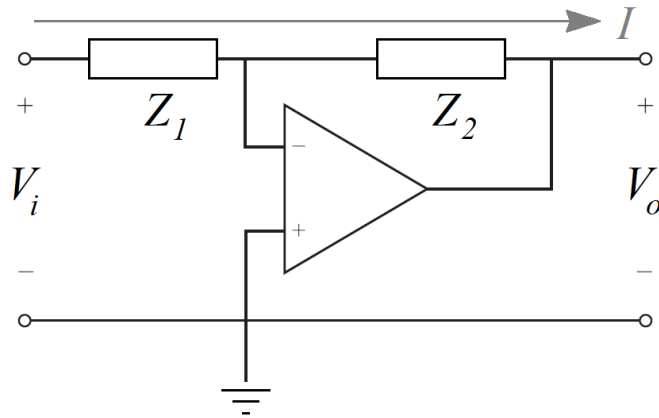


Figure 5.12: Inverting OpAmp with two generic impedances.

Example 5.5. The OpAmp can be used to compare two tensions, connected to the two inputs V^+ and V^- . Because K is very large, if $V^+ > V^-$, even if only by a very small margin, the output will saturate at tension V^{S+} . Likewise, if $V^+ < V^-$, even if only by a very small margin, the output will saturate at tension V^{S-} .

Only if V^+ and V^- are equal, and equal to a great precision, will the output be 0 V. Consider the case of a 741 OpAmp, typically supplied with $V^{S\pm} = \pm 15$ V. Suppose that $K = 10^5$. Then the output V^{out} will not saturate at either +15 V or -15 V only if $|V^+ - V^-| < 15 \times 10^{-5}$ V. \square

Example 5.6. OpAmps are very usually employed in the configuration shown in Figure 5.12, known as **inverting OpAmp** or **inverter**. In this case, because the OpAmp's input impedance is very large, the current I will flow from input V_i to output V_o , as shown in Figure 5.12. Consequently,

Inverting OpAmp or inverter

$$\begin{cases} V_o = K(V^+ - V^-) \Leftrightarrow V^- = -\frac{V_o}{K} \\ Z_2 = \frac{V_o - V^-}{I} \Leftrightarrow I = \frac{V_o - V^-}{Z_2} \\ Z_1 = \frac{V^- - V_i}{I} \Leftrightarrow I = \frac{V^- - V_i}{Z_1} \end{cases} \quad (5.31)$$

Eliminating I and V^- , we get

$$\frac{V_o + \frac{V_o}{K}}{Z_2} = \frac{-\frac{V_o}{K} - V_i}{Z_1} \Leftrightarrow V_o Z_1 + V_o \frac{Z_1}{K} + V_o \frac{Z_1}{K} = -V_i Z_2 \Leftrightarrow \frac{V_o}{V_i} = -\frac{Z_2 K}{Z_1 K + Z_1 + Z_2} \quad (5.32)$$

Because K is large, (5.32) reduces to

$$\frac{V_o}{V_i} = -\frac{Z_2}{Z_1} \quad \square \quad (5.33)$$

Remark 5.8. Notice how (5.33) shows that we can assume

$$V^+ = V^- \quad (5.34)$$

(and so in this case $V^- = 0$). This is because of the high input impedance. \square

Example 5.7. If in Figure 5.12 we make $Z_1 = R_1$ and $Z_2 = R_2$, we obtain the circuit in Figure 5.13, known as **inverting amplifier**, for which

Inverting amplifier

$$\frac{V_o}{V_i} = -\frac{R_2}{R_1} \quad (5.35)$$

Notice that, because $R_1, R_2 > 0$ (there are no negative resistances!), in this circuit the signs of V_i and V_o are always opposite. In spite of the circuit's name, it can

- amplify the input (i.e. $|V_o| > |V_i|$, if $R_2 > R_1$), or

Amplification

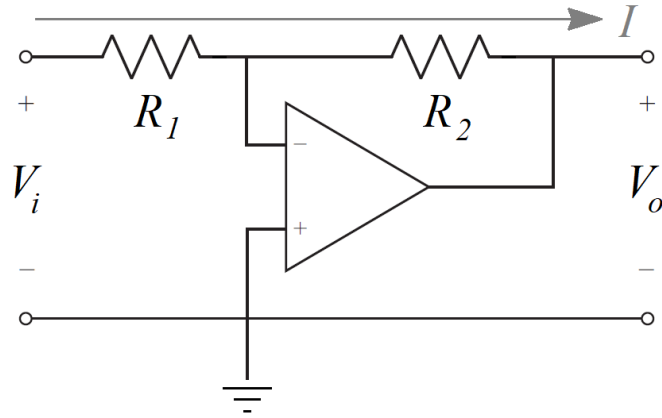


Figure 5.13: Inverter amplifier.

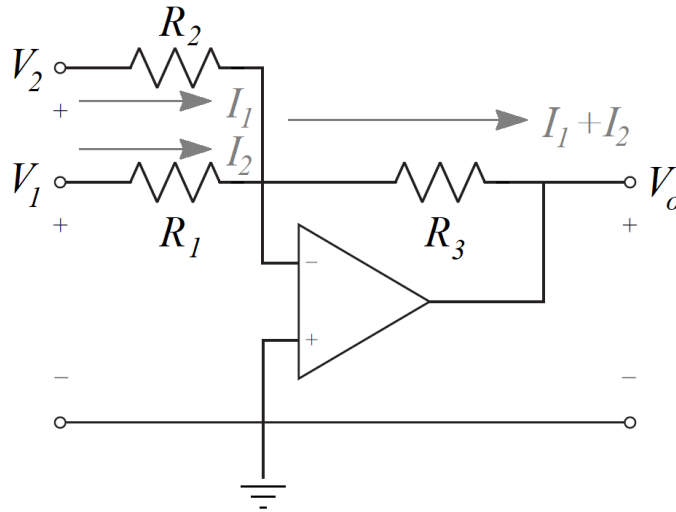


Figure 5.14: Inverting summer or summing circuit.

- attenuate the input (i.e. $|V_o| < |V_i|$, if $R_1 > R_2$).

□ *Attenuation*

Example 5.8. Consider the circuit in Figure 5.14, which is another variation of the negative feedback OpAmp. From (5.34) and the Kirchoff law of nodes, implicit in the currents shown in Figure 5.14, we get

$$\begin{cases} Z_1 = \frac{V_1 - 0}{I_1} \Leftrightarrow I_1 = \frac{V_1}{Z_1} \\ Z_2 = \frac{V_2 - 0}{I_2} \Leftrightarrow I_2 = \frac{V_2}{Z_2} \\ Z_3 = \frac{0 - V_o}{I_1 + I_2} \Leftrightarrow \frac{V_1}{Z_1} + \frac{V_2}{Z_2} = \frac{-V_o}{Z_3} \Leftrightarrow V_o = -\frac{Z_3}{Z_1} V_1 - \frac{Z_3}{Z_2} V_2 \end{cases} \quad (5.36)$$

Consider what happens when all the impedances are resistors:

Inverting summer or inverting summing circuit

- If $Z_1 = Z_2 = Z_3 = R$ this circuit is called **inverting summer** or **inverting summing circuit**. The output V_o is the sum of the two inputs V_1 and V_2 , but with the sign inverted.

Inverting amplifying summer or inverting summing amplifier

- If $Z_1 = Z_2 = R$ and $Z_3 = R_3$ this will be an **inverting amplifying summer** or **inverting summing amplifier**. The amplifying ratio is $-\frac{R_3}{R}$ (and can correspond to amplification or attenuation).

Inverting weighted summer

- If all the resistances are different, we will have an **inverting weighted summer**. If $\frac{R_3}{R_1} + \frac{R_3}{R_2} = 1$ there is no amplification or attenuation; otherwise there is. □

Remark 5.9. Notice that the circuit in Figure 5.14 is a MISO system. □

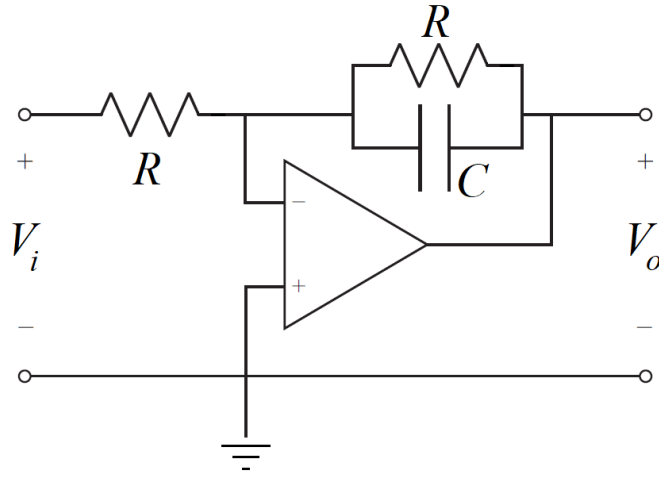


Figure 5.15: Inverting RC circuit with an OpAmp.

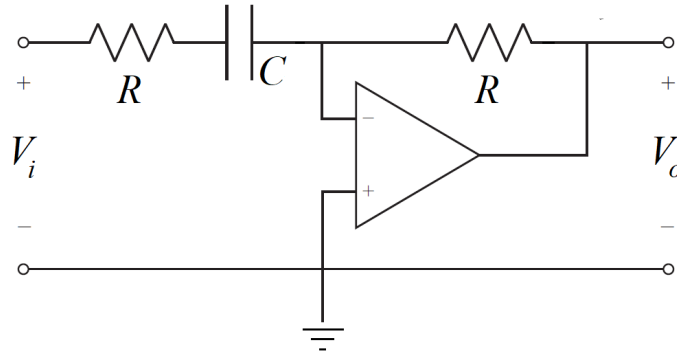


Figure 5.16: Inverting CR circuit with an OpAmp.

Example 5.9. If in Figure 5.12 we have $Z_1 = R$ and Z_2 consists in a resistor R and a capacitor C in parallel, we obtain the circuit in Figure 5.15, with

$$Z_2 = \frac{1}{\frac{1}{R} + \frac{1}{Cs}} = \frac{R}{1 + RCs} \quad (5.37)$$

$$\frac{V_o}{V_i} = -\frac{1}{1 + RCs} \quad (5.38)$$

and similar to the RC circuit from Example 5.2 with transfer function (5.16).

If in Figure 5.12 we have $Z_2 = R$ and Z_1 consists in a resistor R and a capacitor C in series, we obtain the circuit in Figure 5.16, with

$$Z_1 = R + \frac{1}{Cs} = \frac{1 + RCs}{Cs} \quad (5.39)$$

$$\frac{V_o}{V_i} = -\frac{RCs}{1 + RCs} \quad (5.40)$$

and similar to the CR circuit from Example 5.3. \square

Example 5.10. Other than the inverter configuration in Figure 5.12, the most usual configuration with which OpAmps are used is the one in Figure 5.17, known as the **non-inverting OpAmp** or **non-inverter**. Because of the very large

Non-inverting OpAmp or non-inverter

$$\begin{cases} V_o = K(V_i - V^-) \Leftrightarrow V^- = V_i - \frac{V_o}{K} \\ Z_2 = \frac{V_o - V^-}{I} \Leftrightarrow I = \frac{V_o - V^-}{Z_2} \\ Z_1 = \frac{V^- - 0}{I} \Leftrightarrow I = \frac{V^-}{Z_1} \end{cases} \quad (5.41)$$

Eliminating I and V^- , we get

$$\frac{V_o - V_i + \frac{V_o}{K}}{Z_2} = \frac{V_i - \frac{V_o}{K}}{Z_1} \Leftrightarrow V_o Z_1 - V_i Z_1 + V_o \frac{Z_1}{K} = V_i Z_2 - V_o \frac{Z_2}{K} \Leftrightarrow \frac{V_o}{V_i} = \frac{Z_1 + Z_2}{Z_1 + \frac{Z_1}{K} + \frac{Z_2}{K}} \quad (5.42)$$

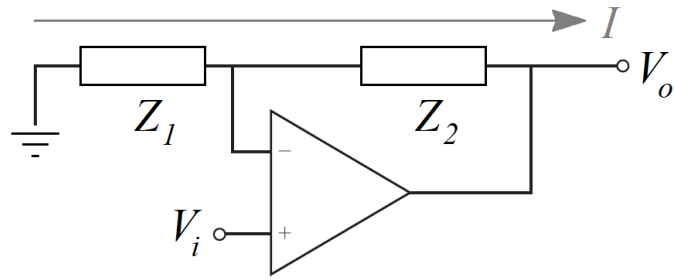


Figure 5.17: Non-inverting OpAmp with two generic impedances.

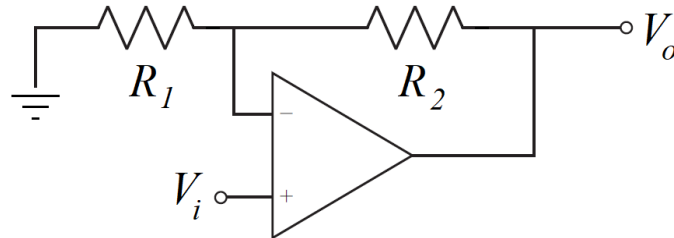


Figure 5.18: Non-inverting amplifier.

Because K is large, (5.42) reduces to

$$\frac{V_o}{V_i} = \frac{Z_1 + Z_2}{Z_1} \quad \square \quad (5.43)$$

Remark 5.10. (5.43) shows once again that we can assume (5.34). We would have arrived sooner at the same result. \square

Non-inverting amplifier

Example 5.11. If in Figure 5.17 we make $Z_1 = R_1$ and $Z_2 = R_2$, we obtain the circuit in Figure 5.18, known as **non-inverting amplifier**, for which

$$\frac{V_o}{V_i} = \frac{R_1 + R_2}{R_1} \quad (5.44)$$

Notice that, because $R_1, R_2 > 0$, not only in this circuit the signs of V_i and V_o are always the same, as the input is always amplified (i.e. $|V_o| > |V_i|$): it is impossible to attenuate the input. \square

Example 5.12. Suppose that we want to amplify a tension 4 times. We can use the non-inverting amplifier of Figure 5.18 with $R_2 = 3R_1$. As an alternative, we can use two inverting amplifiers as in Figure 5.19. \square

Remark 5.11. When we want to attenuate a tension without inverting its signal, a non-inverting amplifier cannot be used, since it must be always true that $\frac{V_o}{V_i} > 1$; two inverting amplifiers in series must be used instead, as in the previous example. \square

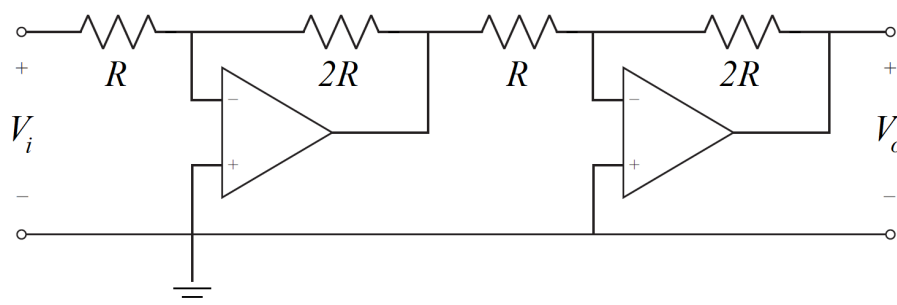


Figure 5.19: Two inverting amplifiers, that amplify the input 4 times.

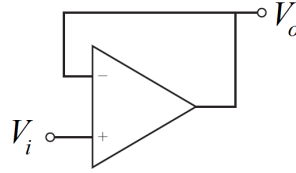


Figure 5.20: Voltage buffer.

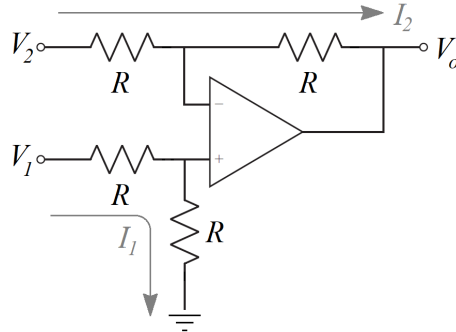


Figure 5.21: Subtractor.

Example 5.13. Consider the circuit in Figure 5.20. Because of (5.34), we have $V^+ = V^- = V_i$, and $V_o = V^-$; hence

$$V_o = V_i \quad (5.45)$$

While at first sight this may seem a good candidate for the prize of the most useless circuit, it is in reality a most useful one. We can be sure that $V_o = V_i$ and that whatever components are connected to V_o will not affect V_i , because there is no current flowing between V_i and V_o . (The source of energy is the OpAmp's power supply.) If it were not for the OpAmp, anything connected to V_o would modify the value of V_i . This circuit is known as **voltage buffer** or **tension buffer**. □ *Voltage buffer*

Example 5.14. The MISO system in Figure 5.21 is known as **subtractor**, because *Subtractor*

$$\begin{cases} R = \frac{V_2 - V^\pm}{I_2} \\ R = \frac{V^\pm - V_o}{I_2} \\ R = \frac{V_1 - V^\pm}{I_1} \\ R = \frac{V^\pm - 0}{I_1} \end{cases} \quad (5.46)$$

From the last two equations, we get $2V^\pm = V_1$. From the first two equations, and replacing this last result,

$$V_2 = 2V^\pm - V_o \Leftrightarrow V_o = V_1 - V_2 \quad \square \quad (5.47)$$

5.4 Other components

Among the several other components that may be found in mechanical systems, we will study the model of the **transformer**, shown in Figure 5.22: *Transformer*

$$\frac{V_P}{V_S} = \frac{N_P}{N_S} \quad (5.48)$$

Here V_P and V_S are the tensions in the two windings, and N_P and N_S are the corresponding numbers of turns in each winding. This is an ideal model; in practice, there are losses, but we will not need to use a more accurate expression.

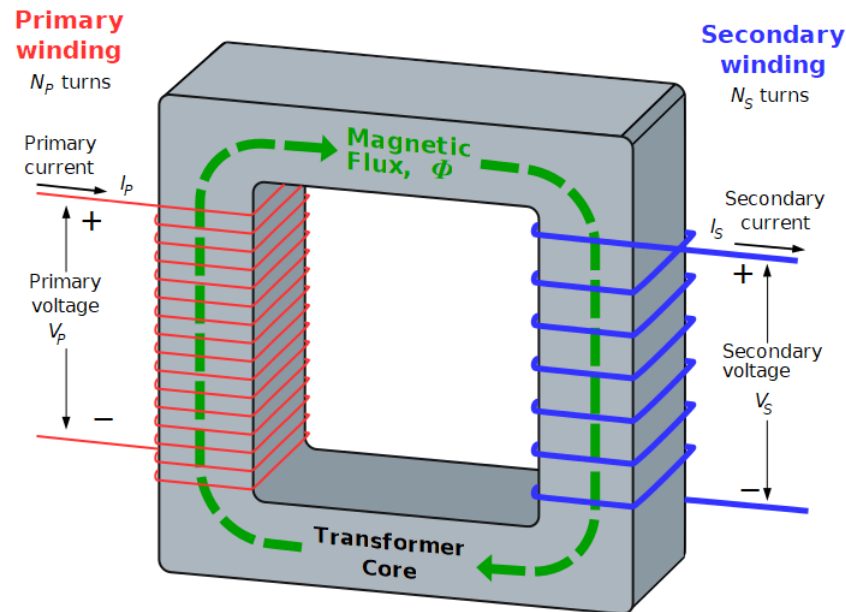


Figure 5.22: Transformer (source: Wikimedia commons).

Glossary

Et le professeur Lidenbrock devait bien s'y connaître, car il passait pour être un véritable polyglotte. Non pas qu'il parlât couramment les deux mille langues et les quatre mille idiomes employés à la surface du globe, mais enfin il en savait sa bonne part.

Jules VERNE (1828 — †1905), *Voyage au centre de la Terre* (1864), 2

active component componente ativo
amplification amplificação
attenuation atenuação
capacitance capacidade elétrica, capacitância (bras.)
capacitor condensador, capacitor (bras.)
current corrente, intensidade (de corrente elétrica)
electric potential difference voltagem, tensão, diferença de potencial elétrico
flux linkage fluxo magnético total
impedance impedância
inductance indutância
inductor bobina, indutor
inverting amplifier amplificador inversor
inverter inversor
inverting amplifying summer somador amplificador inversor
inverting OpAmp AmpOp inversor
inverting summer somador inversor
inverting summing amplifier amplificador somador inversor
inverting summing circuit circuito somador inversor
inverting weighted summer somador inversor ponderado
non-inverter não-inversor
non-inverting OpAmp AmpOp não-inversor
OpAmp AmpOp
operational amplifier amplificador operacional
passive component componente passivo
potentiometer potenciômetro, resistência variável, reóstato
resistance resistência
resistor resistência, resistor (bras.)
rheostat potenciômetro, resistência variável, reóstato
subtractor subtrator
tension voltagem, tensão, diferença de potencial elétrico

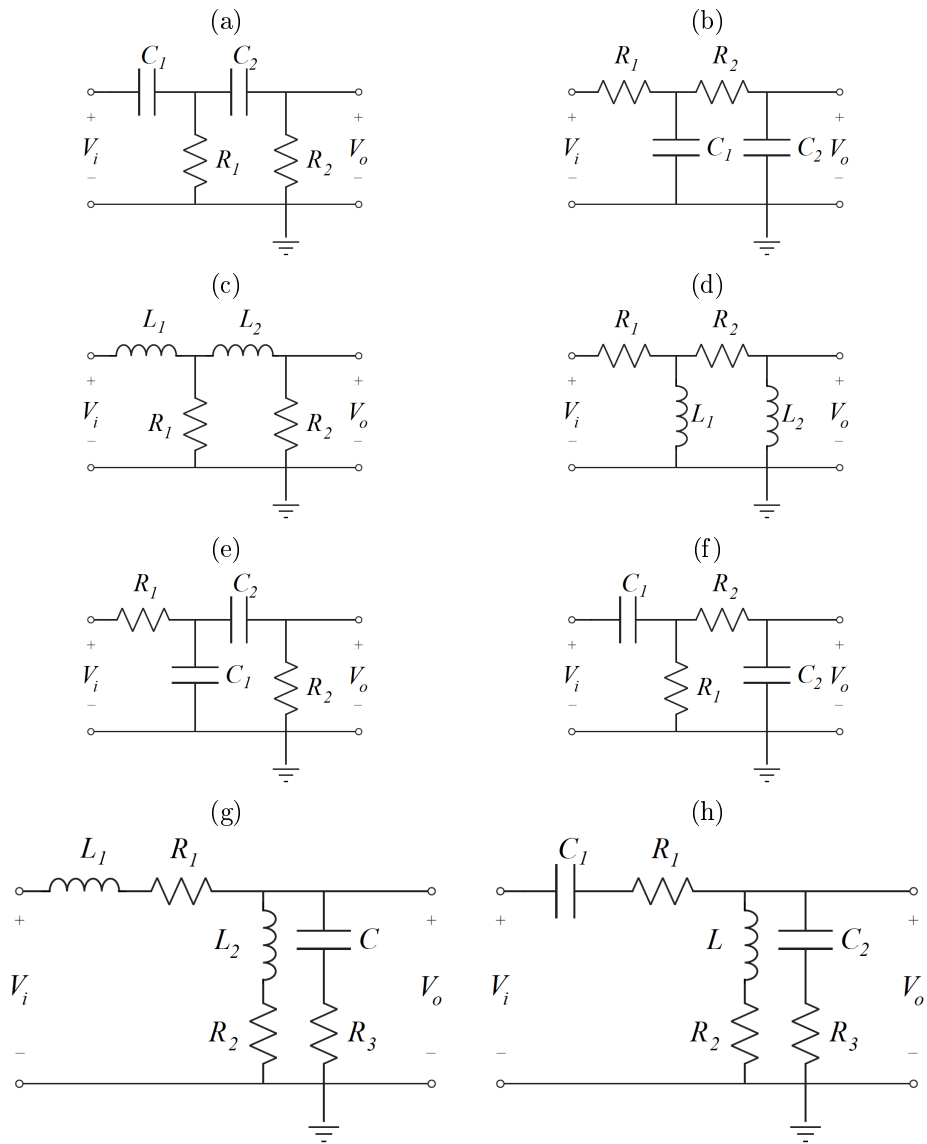


Figure 5.23: Systems of Exercises 1 and 2.

tension buffer AmpOp seguidor de tensão, *buffer* de tensão
variable resistor potenciômetro, resistência variável, reóstato
voltage voltagem, tensão, diferença de potencial elétrico
voltage buffer AmpOp seguidor de tensão, *buffer* de tensão

Exercises

1. Find the equations describing the dynamics of the systems in Figure 5.23, and apply the Laplace transform to the equations to find the corresponding transfer function.
2. Again for the systems in Figure 5.23, find the transfer function directly from the impedances of the components, and apply the inverse Laplace transform to the transfer functions to find the corresponding equations.
3. Show that the differential equations modelling the circuit in Figure 5.24 are similar to those of the mechanical system of Exercise 1 in Chapter 4. Explain why, using the concepts of effort and flow.
4. Find the mechanical systems equivalent to the circuits in Figure 5.25.
5. Find the transfer function of the circuit in Figure 5.12 from the impedance of the components for the following cases:

- (a) Impedance Z_1 is a resistor, impedance Z_2 is a capacitor.

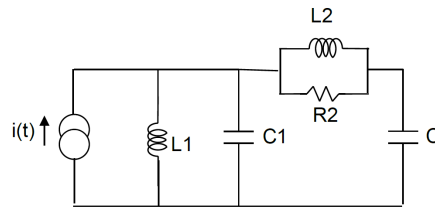


Figure 5.24: Circuit of Exercise 3.

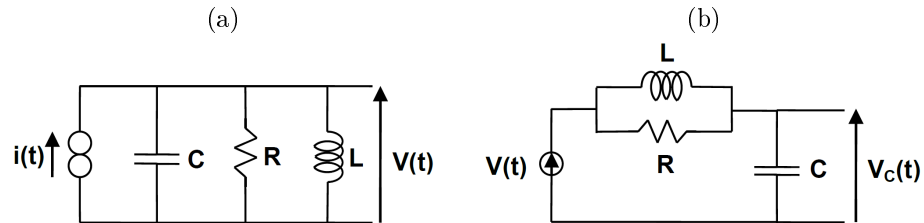


Figure 5.25: Systems of Exercise 4.

- (b) Impedance Z_1 is a capacitor, impedance Z_2 is a resistor.
- (c) Impedance Z_1 is a resistor, impedance Z_2 is an inductor.
- (d) Impedance Z_1 is an inductor, impedance Z_2 is a resistor.
- (e) Both impedances Z_1 and Z_2 are capacitors.
- (f) Both impedances Z_1 and Z_2 are inductors.
- (g) Impedance Z_1 consists in a resistor and a capacitor in series, impedance Z_2 is a resistor.
- (h) Impedance Z_1 consists in a resistor and a capacitor in parallel, impedance Z_2 is a resistor.
- (i) Impedance Z_1 is a resistor, impedance Z_2 consists in a resistor and a capacitor in series.
- (j) Impedance Z_1 is a resistor, impedance Z_2 consists in a resistor and a capacitor in parallel.
- (k) Both impedances Z_1 and Z_2 consist in a resistor and a capacitor in series.
- (l) Both impedances Z_1 and Z_2 consist in a resistor and a capacitor in parallel.
- (m) Impedance Z_1 consists in a resistor and a capacitor in series, impedance Z_2 consists in a resistor and a capacitor in parallel.
- (n) Impedance Z_1 consists in a resistor and a capacitor in parallel, impedance Z_2 consists in a resistor and a capacitor in series.
- (o) Impedance Z_1 consists in a resistor and an inductor in series, impedance Z_2 is a resistor.
- (p) Impedance Z_1 consists in a resistor and an inductor in parallel, impedance Z_2 is a resistor.
- (q) Impedance Z_1 is a resistor, impedance Z_2 consists in a resistor and an inductor in series.
- (r) Impedance Z_1 is a resistor, impedance Z_2 consists in a resistor and an inductor in parallel.
- (s) Both impedances Z_1 and Z_2 consist in a resistor and an inductor in series.
- (t) Both impedances Z_1 and Z_2 consist in a resistor and an inductor in parallel.
- (u) Impedance Z_1 consists in a resistor and an inductor in series, impedance Z_2 consists in a resistor and an inductor in parallel.
- (v) Impedance Z_1 consists in a resistor and an inductor in parallel, impedance Z_2 consists in a resistor and an inductor in series.

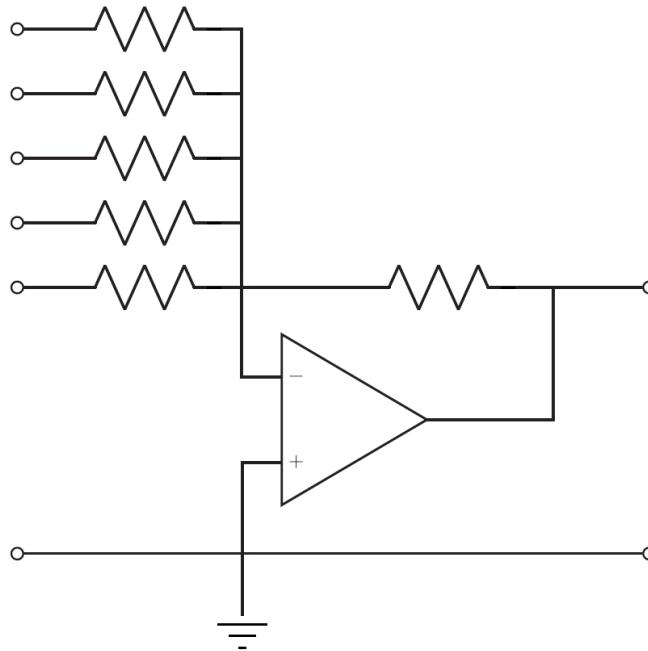


Figure 5.26: Circuit of Exercise 6.

6. Find the transfer function of the circuit in Figure 5.26. Assume that all resistors are equal.
7. How could you perform operation $V_o = V_1 - V_2$ using two OpAmps and without resorting to the subtractor in Figure 5.21?
8. Design a circuit to perform the operation $V_o = V_1 + V_2 - V_3 + 2V_4 - 3V_5$. Use only one OpAmp.
9. Design a circuit to perform the operation $V_o = 10(V_1 + V_2 + \frac{1}{2}V_3)$. Use two OpAmps.
10. Modify the subtractor in Figure 5.21 so as to give:
 - (a) $V_o = V_1 - \frac{1}{3}V_2$
 - (b) $V_o = 5(V_1 - V_2)$
11. Suppose you are using an OpAmp with power supply $V^{S\pm} = \pm 20$ V as comparator. Use MATLAB to plot the expected output V_o for $0 \leq t \leq 10$ s and the following inputs:
 - (a) $V^+ = \sin(t\pi)$ V and $V^- = 5$ V
 - (b) $V^+ = 5$ V and $V^- = \sin(t\pi)$ V
 - (c) $V^+ = 10 \sin(t\pi)$ V and $V^- = 5$ V
 - (d) $V^+ = 5$ V and $V^- = 10 \sin(t\pi)$ V

Chapter 6

Modelling fluidic systems

Suppose a solid held above the surface of a liquid and partially immersed: a portion of the liquid is displaced, and the level of the liquid rises. But, by this rise of level, a little bit more of the solid is of course immersed, and so there is a new displacement of a second portion of the liquid, and a consequent rise of level. Again, this second rise of level causes a yet further immersion, and by consequence another displacement of liquid and another rise. It is self-evident that this process must continue till the entire solid is immersed, and that the liquid will then begin to immerse whatever holds the solid, which, being connected with it, must for the time be considered a part of it. If you hold a stick, six feet long, with its end in a tumbler of water, and wait long enough, you must eventually be immersed. The question as to the source from which the water is supplied—which belongs to a high branch of mathematics, and is therefore beyond our present scope—does not apply to the sea. Let us therefore take the familiar instance of a man standing at the edge of the sea, at ebb-tide, with a solid in his hand, which he partially immerses: he remains steadfast and unmoved, and we all know that he must be drowned.

Lewis CARROLL (1832 — †1898), *A tangled tale* (1885), Knot IX

In this chapter we are concerned with fluid flow in pipes (not with fluid flow with a free surface). Fluidic systems can be accurately modelled using the Navier-Stokes equations, which you learn in a different course. Fortunately, in many cases of fluid flow in pipes it is possible to use simplified equations as follows. *Pipe flow*

6.1 Energy, effort and flow

Energy E is given by the integral over distance x of the force F exerted by the fluid:

$$E = \int_0^x F dx \quad (6.1)$$

The force is equal to the product of the pressure p and the cross-sectional area A , so

$$F = p A \Rightarrow E = \int_0^x p A dx \quad (6.2)$$

The **volume flow rate** (or volumetric flow rate) Q is the derivative of the volume $V = Ax$, given by *Volume flow rate*

$$Q = A \frac{dx}{dt} \quad (6.3)$$

where we assume a constant A . With some abuse of notation, we can write $A = \frac{Q dt}{dx}$ and replace this in (6.2) to rewrite the integral in (6.1) as

$$E = \int_0^t pQ dt \quad (6.4)$$

Table 6.1: Effort, flow, accumulators and dissipators in fluidic systems

	Fluidic system	SI
effort e	pressure P	Pa
flow f	volume flow rate Q	m ³ /s
effort accumulator	fluidic inductance with inertance L	kg m ⁻⁴
accumulated effort $e_a = \int e dt$	fluidic moment $\Gamma = \int p dt$	Pa s
relation between accumulated effort and flow $e_a = \varphi(f)$	fluidic moment $\Gamma = LQ$	
accumulated energy $E_e = \int e_a df$	kinetic energy of the flow $E_e = \frac{1}{2}LQ^2$	J
flow accumulator	reservoir with capacitance C	F
accumulated flow $f_a = \int f dt$	volume $V = \int Q dt$	m ³
relation between accumulated flow and effort $f_a = \varphi(e)$	volume $V = Cp$	
accumulated energy $E_f = \int f_a de$	potential energy of the flow $E_f = \frac{1}{2}Cp^2$	J
dissipator	fluidic resistance R	kg s ⁻¹ m ⁻⁴
relation between effort and flow $e = \varphi(f)$	$p = RQ$	
dissipated energy $E_d = \int f de$	$E_d = \frac{1}{2}RQ^2$	J

So **pressure** p and **volume flow rate** Q can be used as effort and flow. By an understandable universal convention, Q is always considered as the flow, and p as the effort.

Table 6.1 sums up the passing information and relations. The next section presents the basic components mentioned in that Table.

6.2 Basic components of a fluidic system

The basic components of a fluidic system are the following:

*Reservoir
Tank*

1. A **reservoir** or **tank**, which may either have a free surface (see Figure 6.1) or not. Tanks of the first case are often used with liquids; closed tanks are the only option when the fluid is a gas, since the gas might otherwise escape, even if its density is higher than that of the air.

Open tank

In the case of a tank with a free surface, there must be a pipe at the bottom (otherwise the tank could not be emptied). The pressure p at that point, as you know, is

$$p = \rho gh \quad (6.5)$$

where ρ is the fluid density, g is the acceleration of gravity, and h is the height of the fluid in the tank. But the volume of fluid in the tank is given by $V = Ah$, and so, replacing $h = \frac{V}{A}$ in (6.5) and solving in order to V ,

Capacitance

$$V = p \underbrace{\frac{A}{\rho g}}_{\text{capacitance } C} \quad (6.6)$$

Pressurised tank

In the case of reservoirs without a free surface, it can also be shown that $V = pC$, where the value of capacitance C will depend on whether the fluid is a liquid, a gas undergoing an isothermal compression or expansion, or a gas undergoing an adiabatic compression or expansion. We need not worry with that, as long as the value of C is known.

Fluidic inductance

2. A **fluidic inductance**. This is in fact one of the two phenomena that take place in a pipe. Its model is an application of Newton's second law (4.1) to the fluid contained in a length ℓ of pipe (see Figure 6.2):

$$\underbrace{Ap}_{\text{force}} = \underbrace{\rho A \ell}_{\text{mass}} \frac{d^2x}{dt^2} \quad (6.7)$$

The force is the product of the cross-sectional area and the pressure (or rather the difference of pressures at the two extremities of the fluid separated by length ℓ). Integrating both sides, and introducing the **fluidic moment** $\Gamma = \int p dt$,

$$\Gamma = \rho \ell \frac{dx}{dt} \quad (6.8)$$



Figure 6.1: A water reservoir at the Évora train station. (Source: Wikimedia.)

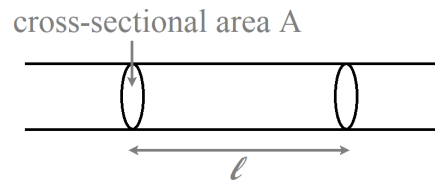


Figure 6.2: A pipe with cross-sectional area A .

From (6.3) we know that $\frac{dx}{dt} = \frac{Q}{A}$, so

$$\Gamma = \underbrace{\frac{\rho \ell}{A}}_{\text{inertance } L} Q \quad (6.9)$$

3. A **pressure drop**. This is the other phenomena taking place in any pipe, due to the resistance of viscous forces both between the fluid and the wall of the pipe and between fluid particles themselves. In another course you learn about the difference between **laminar flow** (i.e. a situation in which fluid particles move essentially in the direction of the flow only) and **turbulent flow** (i.e. a situation in which fluid particles move in a far more chaotic manner).

Pressure drop

Laminar flow

Turbulent flow

Here it suffices to notice that in laminar flow the **Hagen-Poiseuille equation** applies:

Hagen-Poiseuille equation for laminar flow

$$p = \underbrace{\frac{8\mu\ell}{\pi r^4}}_{\text{fluidic resistance } R} Q \quad (6.10)$$

Here p is the pressure drop over length ℓ of the pipe, μ is the dynamic viscosity, and r is the pipe radius. (If the cross-section of the pipe is not circular, then $r = \sqrt{\frac{A}{\pi}}$.) This expression was first determined experimentally, and then proved from the Navier-Stokes equations; all that we need to worry about is the value of the fluidic resistance.

The pressure drop is always higher for turbulent flow than for laminar flow, and the relation between p and Q is no longer linear. However, it may be linearised around a convenient point, so as to find an approximate value of resistance $R = \frac{p}{Q}$ valid in some range of values of these variables. (See Figure 4.3 again.)

The pressure also drops when the flow crosses valves, bends (or shouldered fittings), orifices, diameter reducers, etc.. Model $p = RQ$ is usually a good fit for those situations as well.

Remark 6.1. Pipes have both inertance and resistance. Of course, it may be that one of the two is neglectable when compared to the other; but in reality both are present. \square

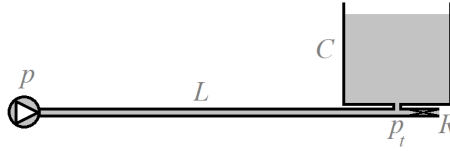


Figure 6.3: System of Example 6.1.

Fluidic impedances

Remark 6.2. The impedances of these components are as follows:

$$\frac{P(s)}{Q(s)} = \frac{1}{Cs} \quad (6.11)$$

$$\frac{P(s)}{Q(s)} = Ls \quad (6.12)$$

$$\frac{P(s)}{Q(s)} = R \quad \square \quad (6.13)$$

Pipe flow can be modelled putting together the equations describing these components with the conservation of mass.

Example 6.1. Consider the system in Figure 6.3, supplied with water by a pump providing a pressure p , that flows through a long pipe with inductance L and neglectable resistance, and either fills a tank with capacitance C or leaves the system through a valve with resistance R . We want to know the pressure below the tank p_t .

Let $q(t)$ be the volume flow rate through the long pipe, that is then divided into the flow feeding the tank $q_t(t)$ and the flow through the valve $q_v(t)$. Using the impedances, we get

$$\begin{cases} \frac{P(s) - P_t(s)}{Q(s)} = Ls \\ Q(s) = Q_t(s) + Q_v(s) \\ \frac{P_t(s)}{Q_t(s)} = \frac{1}{Cs} \\ \frac{P_t(s)}{Q_v(s)} = R \end{cases} \Rightarrow \begin{cases} P(s) - P_t(s) = LsP_t(s) \left(\frac{1}{R} + Cs \right) \\ Q(s) = P_t(s) \left(\frac{1}{R} + Cs \right) \\ Q_t(s) = P_t(s)Cs \\ Q_v(s) = \frac{1}{R}P_t(s) \end{cases} \quad (6.14)$$

The first equation then gives the desired answer:

$$P(s) = P_t(s) \left(1 + \frac{L}{R}s + LCs^2 \right) \Leftrightarrow \frac{P_t(s)}{P(s)} = \frac{1}{1 + \frac{L}{R}s + LCs^2} \quad \square \quad (6.15)$$

Remark 6.3. Notice that (6.15) is similar to the model of a mass–spring–damper system (4.9) or the model of an RLC system (5.24). \square

Remark 6.4. Liquids can be presumed to be incompressible, so ρ is constant and independent from p . Thus (6.5) shows that there is a one-to-one relation between p and h , where h is the **hydraulic head**. So, when the fluid is a liquid, p is often replaced by ρgh . To do this, of course, the density of the liquid used in the system must be fixed in advance; the most usual cases are water, brine, and crude oil.

Hydraulic head

Using the head instead of the pressure

Model (6.9) of a fluidic inductance tells us that $\int p dt = \Gamma = LQ$; applying the Laplace transform, this becomes

Fluidic inductance

$$\frac{p}{s} = LQ \Leftrightarrow \frac{\rho gh}{s} = LQ \Leftrightarrow h = \frac{L}{\rho g} sQ \quad (6.16)$$

Here, $L^* = \frac{L}{\rho g}$ is the inductance relating hydraulic head and flow; L is the inductance relating pressure and flow. Notice that the SI units of L are kg m^{-4} ; those of L^* are $\text{s}^2 \text{m}^{-2}$.

Fluidic resistance

Likewise, model (6.10) of fluidic resistance tells us that $p = RQ$; so

$$\rho gh = RQ \Leftrightarrow h = \frac{R}{\rho g} Q \quad (6.17)$$

Here, $R^* = \frac{R}{\rho g}$ is the resistance relating hydraulic head and flow; R is the resistance relating pressure and flow. Notice that the SI units of R are $\text{kg s}^{-1} \text{m}^{-4}$; those of R^* are s m^{-2} . \square

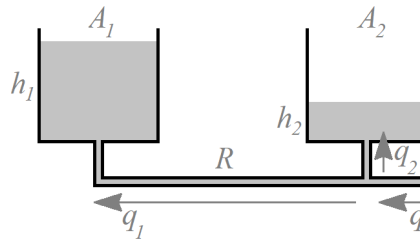


Figure 6.4: System of Example 6.2.

Example 6.2. Consider the system in Figure 6.4, with two water reservoirs fed by a pump that delivers a flow q , and connected by a pipe with neglectable inductance and resistance R . We have

$$\begin{cases} q(t) = q_1(t) + q_2(t) \\ q_1(t) = A_1 \dot{h}_1(t) \\ q_2(t) = A_2 \dot{h}_2(t) \\ \frac{h_2(t) - h_1(t)}{R} = q_1(t) \end{cases} \Rightarrow \begin{cases} Q(s) = A_1 H_1(s)s + A_2 H_2(s)s \\ Q_1(s) = A_1 H_1(s) \\ Q_2(s) = A_2 H_2(s)s \\ H_2(s) - H_1(s) = RA_1 H_1(s)s \end{cases} \quad (6.18)$$

Thus

$$\begin{cases} Q(s) = A_1 H_1(s)s + A_2 H_1(s)(RA_1 s + 1)s \\ H_2(s) = H_1(s)(RA_1 s + 1) \end{cases} \Leftrightarrow \begin{cases} Q(s) = H_1(s)(RA_1 A_2 s^2 + A_1 s + A_2 s) \\ H_2(s) = H_1(s)(RA_1 s + 1) \end{cases} \Leftrightarrow \begin{cases} \frac{H_1(s)}{Q(s)} = \frac{1}{RA_1 A_2 s^2 + (A_1 + A_2)s} \\ \frac{H_2(s)}{Q(s)} = \frac{RA_1 s + 1}{RA_1 A_2 s^2 + (A_1 + A_2)s} \end{cases} \quad \square \quad (6.19)$$

Remark 6.5. When modelling pipe flows, pay special attention to non-linearities. All pipes have some maximum value that the flow can take. So in a pipe like that in Figure 6.4 we have $-Q_{\max} \leq Q \leq Q_{\max}$. But in some systems pipes feed tanks from above: see for instance Figure 6.8 below, corresponding to Exercise 2. In that case, $0 \leq Q_3 \leq Q_{3\max}$. In other words, the flow can fill up the tank below, but it cannot empty it.

Neglecting non-linearities leads to absurd models and ludicrous conclusions, such as those mocked by mathematician Charles Dogdson in the quote at the beginning of this chapter. \square

6.3 Other components

Among the several other components that may be found in pipe flow systems, the hydraulic press deserves a passing mention. Its principle is shown in Figure 6.5. Because of (6.2), a similar pressure on both sides means that

$$\frac{F_1}{A_1} = \frac{F_2}{A_2} \Leftrightarrow F_2 = F_1 \frac{A_2}{A_1} \quad (6.20)$$

where F_1 and F_2 are the forces exerted on the two pistons, with areas A_1 and A_2 . This principle is used in presses such as that in Figure 6.6.

Glossary

Había en el puerto gran multitud de buques de todas clases y tamaños, resplandeciendo entre ellos, llamando la atención y hasta excitando la admiración y la envidia de los españoles, un enorme y hermosísimo navío, construido con tal perfección, lujo y elegancia, que era una maravilla.

Los españoles, naturalmente, tuvieron la curiosidad de saber quién era el dueño del navío y encargaron al secretario que, sirviendo de intérprete, se lo preguntase a algunos alemanes que habían venido a

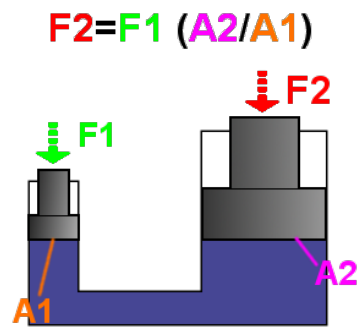


Figure 6.5: Principle of the hydraulic press (source: Wikimedia).



Figure 6.6: Hydraulic press (source: Wikimedia).

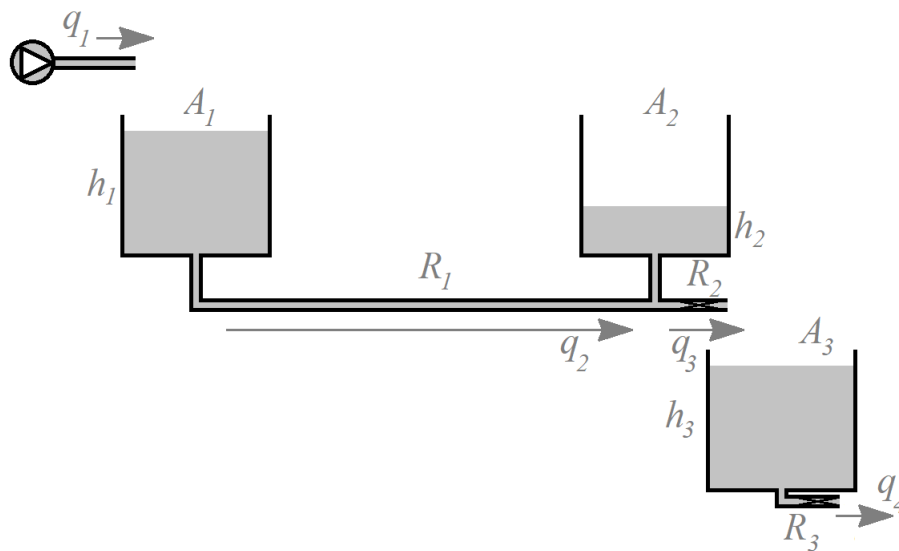


Figure 6.7: System of Exercise 2.

bordo.

Lo preguntó el secretario y dijo luego a sus paisanos y camaradas:

— El buque es propiedad de un poderoso comerciante y naviero de

esta ciudad en que estamos, el cual se llama el señor Nichtverstehen.

Juan VALERA (1824 — †1905), *Cuentos y chascarrillos andaluces* (1896), El señor Nichtverstehen

brine água salgada (lit. salmoura)

fluidic inductance indutância fluídica

fluidic moment momento fluídico

fluidic resistance resistêcia fluídica

hydraulic head altura de coluna de fluido (de água, de água salgada, de crude)

hydraulic press prensa hidráulica

inertance indutância fluídica

pressure pressão

pressure drop perda de carga

reservoir reservatório, tanque

tank reservatório, tanque

volume flow rate caudal volumétrico

volumetric flow rate caudal volumétrico

Exercises

1. Consider the system from Example 6.1, shown in Figure 6.3. Find its mechanical equivalent.
2. Consider the system in Figure 6.7, fed by a pump delivering volume flow rate $q_1(t)$. Tanks 1 and 2 are connected by a pipe with neglectable inertance and fluidic resistance R_1 ; tanks 2 and 3 are emptied through valves with resistances R_2 and R_3 respectively. Find transfer functions $\frac{H_1(s)}{Q_1(s)}$, $\frac{H_2(s)}{Q_1(s)}$ and $\frac{H_3(s)}{Q_1(s)}$.
3. Consider the system in Figure 6.8, fed by a pump delivering volume flow rate $q_1(t)$. Find transfer functions $\frac{H_1(s)}{Q_1(s)}$, $\frac{H_2(s)}{Q_1(s)}$, $\frac{H_3(s)}{Q_1(s)}$ and $\frac{Q_5(s)}{Q_1(s)}$.

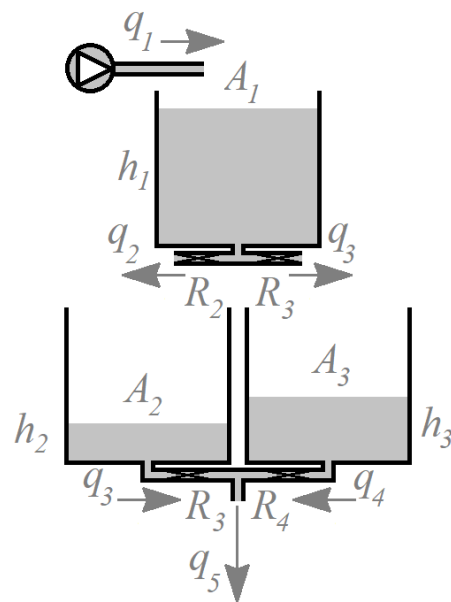


Figure 6.8: System of Exercise 3.

Chapter 7

Modelling thermal systems

— Mais qu’entends-tu par le vide ? demanda Michel, est-ce le vide absolu ?

— C’est le vide absolument privé d’air.

— Et dans lequel l’air n’est remplacé par rien ?

— Si. Par l’éther, répondit Barbicane.

— Ah ! Et qu’est-ce que l’éther ?

— L’éther, mon ami, c’est une agglomération d’atomes impondérables, qui, relativement à leurs dimensions, disent les ouvrages de physique moléculaire, sont aussi éloignés les uns des autres que les corps célestes le sont dans l’espace. Leur distance, cependant, est inférieure à un trois-millionièmes de millimètre. Ce sont ces atomes qui, par leur mouvement vibratoire, produisent la lumière et la chaleur, en faisant par seconde quatre cent trente trillions d’ondulations, n’ayant que quatre à six dix-millièmes de millimètre d’amplitude.

Jules VERNE (1828 — †1905), *Autour de la Lune* (1869), 5

This chapter concerns thermal systems. You will study heat transfer in depth in another course, but simple cases can be modelled with a formulation similar to that employed to the systems in the previous chapters.

7.1 Energy, effort and flow

There is, however, an important difference. By convention:

- **temperature** T is the effort variable;
- **heat flow rate** q is the flow variable.

Temperature

Heat flow rate

However, heat is energy; it is in fact kinetic energy of molecules and atoms. (As you may know, an omnipresent æther was postulated for some centuries to explain the propagation of heat and especially of light, but this hypothesis, of which you may find a popular explanation at the beginning of this chapter, has been abandoned for about one century, having been by then contested for decades because of experimental results with which it could not reconciled.) Consequently, it is not true in thermal systems that energy is the integral of the product of effort and flow — see (4.30) —, as the variable used for flow is an energy rate itself. And, as a result, the parallels with the other types of systems we have been studying are not perfect. *Heat is energy*

7.2 Basic components of a thermal system

There is only one type of accumulator, that of heat, i.e. of flow:

Heat accumulator

$$H(t) = mC_p(T(t) - T(0)) \quad (7.1)$$

Here $H(t) = \int_0^t q(t) dt$ is the accumulated heat in mass m , which has a **specific heat** C_p and is heated from temperature $T(0)$ to temperature $T(t)$.

Specific heat

Dissipation can take place in three different ways:

Conduction

- In **conduction** there is no macroscopic movement of the solid or fluid that undergoes the process. Heat is transmitted, or rather diffused, at the molecular and atomic levels. The heat flow, which we assume to be positive in the sense of increasing values of position x , is

$$q(t) = -kA \frac{\partial T(x,t)}{\partial x} \quad (7.2)$$

where A is the cross-sectional area and k is the thermal conductivity. Assuming that the temperature distribution over x is linear over a distance L , (7.2) becomes

$$q(t) = \underbrace{\frac{kA}{L}}_{\text{conduction heat transfer coefficient } h_c} (T(0,t) - T(L,t)) \quad (7.3)$$

Notice that the minus sign is gone because $\frac{\partial T(x,t)}{\partial x} = \lim_{L \rightarrow 0} \frac{T(L,t) - T(0,t)}{L}$.

Convection

- In **convection** there is macroscopic movement of the fluid where heat transfer is taking place. In solids matter cannot move and this way and consequently there can be no convection. If the fluid movement is due solely to the temperature gradients, this is called **free convection**; if fluid movement is due at least in part to some other reason (like fluid flow in pipes, or a blower), this is called **forced convection**. In any case, the heat flow, again assumed positive in the sense of increasing values of position x , is

Free convection

Forced convection

$$q(t) = \underbrace{hA}_{\text{convection heat transfer coefficient } h_h} (T(0,t) - T(L,t)) \quad (7.4)$$

Here h is the convection heat transfer coefficient, and A is the cross-sectional area over which heat transfer takes place.

Radiation

- In **radiation** heat is propagated by the emission of photons. It is the only heat transmission that can take place in a vacuum. It is also the only one corresponding to a non-linear law:

$$q(t) = C_r A (T_1^4 - T_2^4) \quad (7.5)$$

Here C_r is a proportionality constant that we need not delve into. This law, of course, can be linearised around some point, and the result will be approximately valid in a vicinity thereof; for instance:

$$q(t) = \underbrace{C_r A (T_1 + T_2) (T_1^2 + T_2^2)}_{\text{radiation heat transfer coefficient } h_r} (T_1 - T_2) \quad (7.6)$$

Notice that all cases — conduction, convection, and linearised radiation — can be reduced to the following form:

$$\Delta T = Rq \quad (7.7)$$

Thermal resistance

Here R is thermal resistance and ΔT is the temperature difference.

The relations are summed up in Table 7.1. The corresponding impedances are

$$\frac{\Delta T(s)}{Q(s)} = \frac{1}{mC_p s} \quad (7.8)$$

$$\frac{\Delta T(s)}{Q(s)} = R \quad (7.9)$$

Table 7.1: Effort, flow, accumulators and dissipators in thermal systems

	Thermal system	SI
effort e	temperature T	K or °C
flow f	heat flow rate q	W
effort accumulator	—	—
accumulated effort $e_a = \int e dt$	—	—
relation between accumulated effort and flow $e_a = \varphi(f)$	—	—
accumulated energy $E_e = \int e_a df$	—	—
flow accumulator	heat accumulator with mass m and specific heat C_p	$\text{kg} \times \text{J kg}^{-1} \text{K}^{-1}$
accumulated flow $f_a = \int f dt$	heat $H = \int q dt$	J
relation between accumulated flow and effort $f_a = \varphi(e)$	heat $H = m C_p T$	—
accumulated energy $E_f = \int f_a de$	—	—
dissipator	thermal resistance with resistance R	K/W
relation between effort and flow $e = \varphi(f)$	$T = Rq$	—
dissipated energy $E_d = \int f de$	—	—

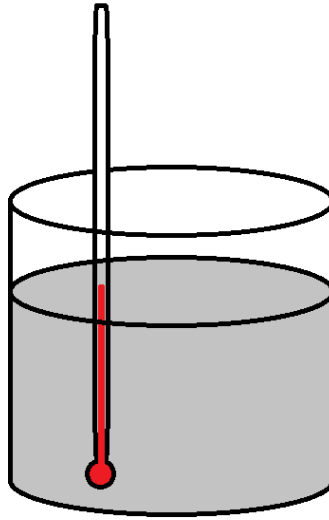


Figure 7.1: A thermometer immersed in a fluid.

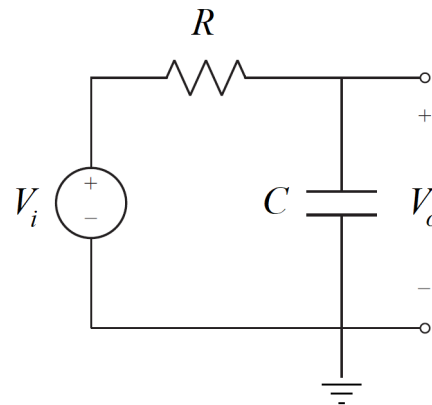


Figure 7.2: Electrical circuit equivalent to the thermal system in Figure 7.1.

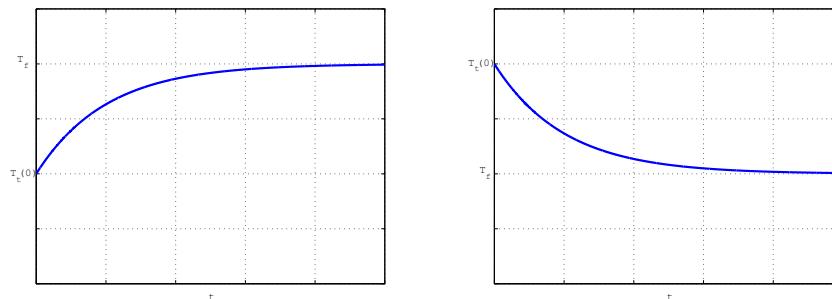


Figure 7.3: Evolution of temperature $T_t(t)$ in the system Figure 7.1 when $T_t(t) < T_f$ (left) and when $T_t(t) > T_f$ (right).

Example 7.1. The reading of a mercury or alcohol thermometer at time $t = 0$ is $T_t(0)$. At that instant, it is immersed in a fluid at temperature T_f (see Figure 7.1). Let the thermal resistance of the glass be R , and the specific heat of mass m of mercury or alcohol be C_p . How does $T_t(t)$ evolve?

This can be seen using an equivalent electrical circuit. Both temperatures become tensions; the thermal resistance of the glass becomes a resistance R and the heat accumulator becomes a capacitor C . So the equivalent circuit is the one in Figure 7.2. The corresponding transfer function — remember (5.16) — is

$$\frac{V_o(s)}{V_i(s)} = \frac{1}{1 + RCs} \quad (7.10)$$

Temperature T_f is constant and is applied from $t = 0$ on, so the Laplace transform of the input is

$$\mathcal{L}[(T_f - T_t(0)) H(t)] = \frac{T_f}{s} \quad (7.11)$$

Notice that the amplitude of the temperature change is $T_f - T_t(0)$; since the Heaviside function begins at 0 and ends in 1, we must multiply it by $T_f - T_t(0)$ and take into account the different initial value also when calculating the output we want to know:

$$\begin{aligned} T_t(t) - T_t(0) &= \mathcal{L}^{-1} \left[\frac{T_f}{s} \frac{1}{1 + RmC_p s} \right] \\ &= T_f \mathcal{L}^{-1} \left[\frac{\frac{1}{RmC_p}}{s \left(\frac{1}{RmC_p} + s \right)} \right] = T_f \left(1 - e^{-\frac{1}{RmC_p} t} \right) \end{aligned} \quad (7.12)$$

We conclude that temperature $T_t(t)$ begins at $T_t(0)$, ends at T_f , and changes exponentially with time. This is illustrated in Figure 7.3. \square

Glossary

Aber in Phantásien waren fast alle Wesen, auch die Tiere, mindestens zweier Sprachen mächtig: Erstens der eigenen, die sie nur mit ihresgleichen redeten und die kein Außenstehender verstand, und zweitens einer allgemeinen, die man Hochphantásisch oder auch die Große Sprache nannte. Sie beherrschte jeder, wenngleich manche sie in etwas eigentümlicher Weise benützten.

Michael ENDE (1929 — †1995), *Die unendliche Geschichte von A bis Z* (1979), 2

conduction condução
convection convecção
forced convection convecção forçada
free convection convecção livre
heat accumulator acumulador de calor
heat flow rate fluxo de calor
radiation radiação
specific heat calor específico
thermal resistance resistência térmica
temperature temperatura

Exercises

1. Consider the Wave Energy Converter of Figure 3.2, when submerged in sea water at constant temperature T_{sea} . Assume that the air inside the device has a homogeneous temperature $T_{air}(t)$ and is heated by the device's Power Take-Off (PTO) mechanism, at temperature $T_{PTO}(s)$, that delivers an electrical power $P(t)$ to the electrical grid with an efficiency η . (See Figure 7.4.) Heat transfer from the PTO to the air inside the device takes place by convection with a constant coefficient h and for area

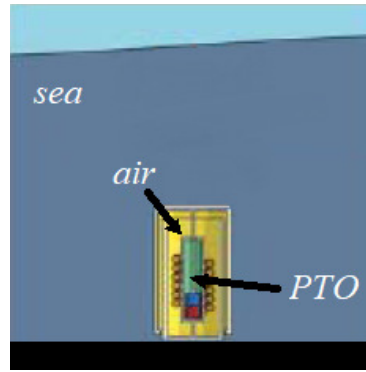


Figure 7.4: The Wave Energy Converter of Exercise 1.

A; neglect the thermal capacity of the metallic WEC itself; consider that C_{air} and C_{PTO} are the thermal capacities of the air and the PTO, that have masses m_{air} and m_{PTO} . Let $T(t) = T_{air}(t) - T_{PTO}(t)$. Find transfer function $\frac{T(s)}{P(s)}$.

2. Explain which of the electrical systems in Figure 5.23 have a thermal equivalent and which do not, and why.

Chapter 8

Modelling interconnected and non-linear systems

“You are right in thinking that he is under the British government. You would also be right in a sense if you said that occasionally he *is* the British government.”

“My dear Holmes!”

“I thought I might surprise you. Mycroft draws four hundred and fifty pounds a year, remains a subordinate, has no ambitions of any kind, will receive neither honour nor title, but remains the most indispensable man in the country.”

“But how?”

“Well, his position is unique. He has made it for himself. There has never been anything like it before, nor will be again. He has the tidiest and most orderly brain, with the greatest capacity for storing facts, of any man living. The same great powers which I have turned to the detection of crime he has used for this particular business. The conclusions of every department are passed to him, and he is the central exchange, the clearinghouse, which makes out the balance. All other men are specialists, but his specialism is omniscience. We will suppose that a minister needs information as to a point which involves the Navy, India, Canada and the bimetallic question; he could get his separate advices from various departments upon each, but only Mycroft can focus them all, and say offhand how each factor would affect the other.”

Sir Arthur CONAN DOYLE (1859 — †1930), *His last bow* (1917), The Adventure of the Bruce-Partington Plans (1908)

This chapter presents an overview of the modelling process.

8.1 Energy, effort and flow

Table 8.1 presents the impedances of all the flow accumulators, effort accumulators, and energy dissipators, summing up Tables 4.1, 5.1, 6.1, and 7.1, and showing clearly the existing parallelism between systems of different types. *Impedances*

This is the place to notice that effort variables are measured in relation to an arbitrary value that serves as zero:

- In Table 8.1 this is explicit for thermal systems, since temperature is denoted as ΔT , as what matters is the temperature difference.
- In the case of electrical systems, what matters is always the electrical tension at the extremities of the component.
- In the case of pipe flow, resistance and inductance depend on the pressure difference at the extremities. Reservoirs with a free surface also depend on a pressure difference, between the pressure of the liquid at the bottom and the atmospheric pressure.

Table 8.1: Effort, flow, accumulators and dissipators in different types systems

Type of system	Mechanical, translation	Mechanical, rotation	Electrical	Fluidic	Thermal
effort e flow f	velocity \dot{x} force F	angular velocity $\dot{\omega}$ torque τ	voltage U current I	pressure P volume flow rate Q	temperature T heat flow rate q
effort accumulator impedance	spring $\frac{sX(s)}{F(s)} = \frac{s}{K}$	angular spring $\frac{s\Omega(s)}{T(s)} = \frac{s}{\kappa}$	inductor $\frac{U(s)}{I(s)} = Ls$	fluidic inductance $\frac{P(s)}{Q(s)} = Ls$	— —
flow accumulator impedance	mass $\frac{sX(s)}{F(s)} = \frac{1}{Ms}$	moment of inertia $\frac{s\Omega(s)}{T(s)} = \frac{1}{Js}$	capacitor $\frac{U(s)}{I(s)} = \frac{1}{Cs}$	reservoir $\frac{P(s)}{Q(s)} = \frac{1}{Cs}$	heat accumulator $\frac{\Delta T(s)}{Q(s)} = \frac{1}{mC_p s}$
dissipator impedance	dampner $\frac{sX(s)}{F(s)} = \frac{1}{b}$	rotary dampner $\frac{s\Omega(s)}{T(s)} = \frac{1}{b}$	resistor $\frac{U(s)}{I(s)} = R$	fluidic resistance $\frac{P(s)}{Q(s)} = R$	thermal resistance $\frac{\Delta T(s)}{Q(s)} = R$

- In the case of mechanical systems, the energy dissipated by a damper depends on the relative velocities of its extremities, and the energy accumulated by a spring depends on the relative position of its extremities.

Notice that there may be values of these variables that we can think of as absolute zeros, such as temperature $-273.15\text{ }^\circ\text{C} = 0\text{ K}$, pressure 0 Pa of complete vacuum, or position and velocity measured in an inertial frame of reference. Still, it is often far more practical to use other values, such as atmospheric pressure, room temperature, or resting position, as zero.

Dealing with initial conditions

Example 8.1. A 300 kg dirigible balloon flies at constant altitude $z = 200\text{ m}$, because its impulsion cancels its weight. It can move vertically thanks to two electrical propulsors, each of which provides a force given by $F_p(t) = \gamma U(t)$, where $U(t)$ is the tension applied (control input) and the gain is $\gamma = 15\text{ N/V}$ (the force is upwards when $U > 0$). When the balloon moves, there is a viscous drag force with coefficient $c = 30\text{ N s/m}$. How does the altitude change with time when a 20 V tension is applied during 10 s?

A balance of forces shows that

$$\underbrace{300\ddot{z}(t)}_{\text{mass} \times \text{acceleration}} = \underbrace{2 \times 15U(t)}_{\text{propulsors}} - \underbrace{30\dot{z}(t)}_{\text{drag force}} \quad (8.1)$$

We know that initial conditions are $z(0) = 200$ and $\dot{z}(0) = 0$, so we could be tempted to apply the Laplace transform as

$$300(Z(s)s^2 - 200s) = 30U(s) - 30(Z(s)s - 200) \quad (8.2)$$

Rearranging terms,

$$(300s^2 + 30s)Z(s) = 30U(s) + 6 \times 10^4 s + 6 \times 10^3 \quad (8.3)$$

$$\Leftrightarrow Z(s) = \frac{1}{(10s + 1)s}U(s) + \frac{2 \times 10^3 s}{(10s + 1)s} + \frac{2 \times 10^2}{(10s + 1)s} = \frac{1}{(10s + 1)s}U(s) + \frac{2 \times 10^2}{s}$$

Notice that it is impossible to find a transfer function $\frac{Z(s)}{U(s)}$ relating the (Laplace transforms of) the input and the output. To obtain a transfer function, make $z^*(t) = z(t) - (0)$, and then

No transfer function if initial conditions are not zero

$$300Z^*(s) = 30U(s) - 30Z^*(s)s \Leftrightarrow \frac{Z^*(s)}{U(s)} = \frac{1}{(10s + 1)s} \quad (8.4)$$

The result will of course be the same, but this allows us to use many results established for transfer functions, such as those in Chapters 9 and 11. It also allows us to use MATLAB to find the answer as follows:

```
>> G = tf(1, [10 1 0]);
>> Ts = 0.001; t = 0 : Ts : 50;
>> U = zeros(size(t)); U(1:10/Ts) = 20*ones(1, 10/Ts);
>> z = lsim(G, U, t); z = z + 200;
>> figure, plot(t,Z)
>> xlabel('t [s]'), ylabel('z [m]')
```

Notice how we had to add 200 to the result (or else we would have to bear in mind that the plot would show oscillations around 200 m). See Figure 8.1. \square

Remark 8.1. Remember that we already did something similar in Example 7.1. \square

8.2 System interconnection

Transfer functions are of great aid when modelling several interconnected systems, of the same or of different types.

Example 8.2. Consider the system in Figure 8.2. The force exerted by the inductance in the handle that undergoes displacement x_2 is given by $F_2(t) = \alpha i(t)$, where $i(t)$ is the current in the inductance. Find $\frac{X_1(s)}{V_1(s)}$.

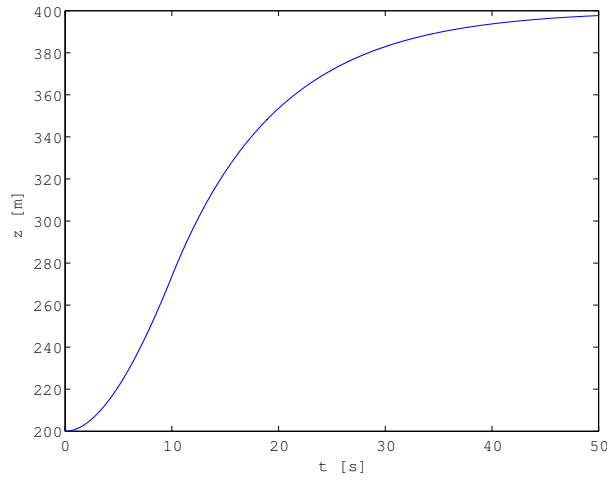


Figure 8.1: Results of Example 8.1.

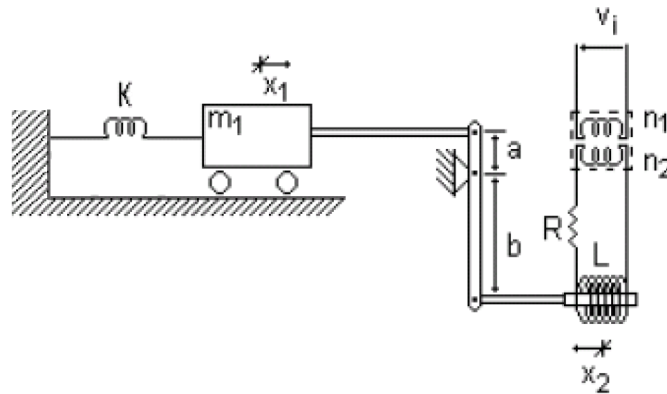


Figure 8.2: System of Example 8.2.

We can of course write all the equations, and obtain the desired result with successive replacements. Transfer functions allow us to model each system separately, making such replacements easier.

As to the electrical system, remembering (5.48),

$$R + Ls = \frac{\frac{n_2}{n_1} V_i(s)}{I(s)} \Leftrightarrow \frac{I(s)}{V_i(s)} = \frac{\frac{n_2}{n_1}}{R + Ls} \quad (8.5)$$

As to the lever, and letting F_1 be the force exerted on mass m_1 ,

$$F_1(t)a = F_2(t)b \Leftrightarrow \frac{F_1(s)}{F_2(s)} = \frac{b}{a} \quad (8.6)$$

As to the mass,

$$F_1(t) - Kx_1(t) = m_1 \ddot{x}_1(t) \Leftrightarrow \frac{X_1(s)}{F_1(s)} = \frac{1}{m_1 s^2 + K} \quad (8.7)$$

Finally,

$$\frac{X_1(s)}{V_i(s)} = \frac{X_1(s)}{F_1(s)} \frac{F_1(s)}{F_2(s)} \frac{F_2(s)}{I(s)} \frac{I(s)}{V_i(s)} = \frac{\frac{b}{a} \alpha \frac{n_2}{n_1}}{(m_1 s^2 + K)(R + Ls)} \quad (8.8)$$

This way, we are also able to study each transfer function separately, analysing its influence in the final result. \square

Bond graphs

Bond graphs are another tool that can be used to assist the modelling of interconnected systems. They consist in a graphical representation of what happens with energy in a system, based upon the concepts of effort and flux. These are written above and below arrows (of which, by convention, only half the

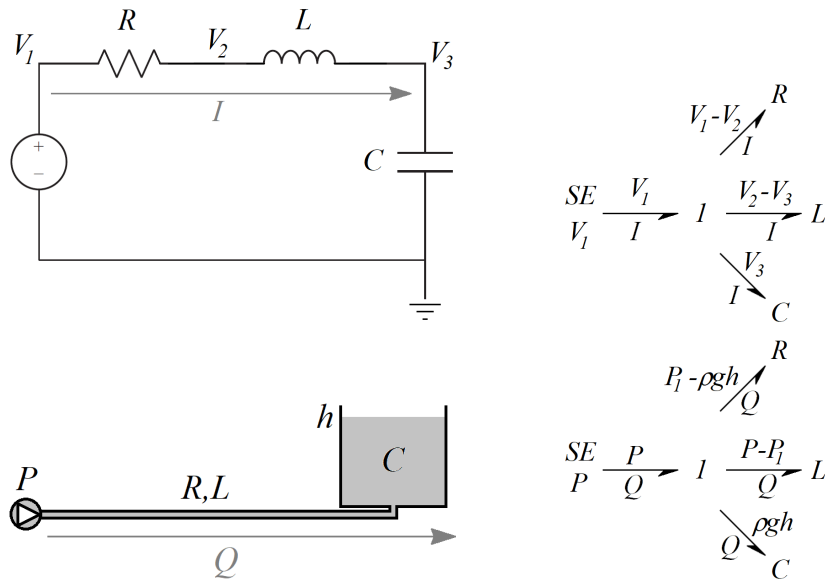


Figure 8.3: Two bond graphs of systems where elements have the same flux and there is an effort junction. Top: electrical circuit. Notice that $V_1 = (V_1 - V_2) + (V_2 - V_3) + V_3$. Bottom: fluidic system. Since the pipe has both resistance and inductance, the pressure change from the pump delivering a constant pressure P to the bottom of the reservoir where the hydraulic head is h and the pressure is ρgh is split into two, as if the fluid would first go through an inductance without resistance and then through a resistance without inductance, so that $P = (P - P_1) + (P_1 - \rho gh) + \rho gh$.

tip is drawn). Figure 8.3 shows two examples of bond graphs in which several elements have the same flux and different efforts; the corresponding junction of the several efforts in one system is by convention denoted by number 1. Figure 8.4 shows two examples of bond graphs in which several elements have the same effort and different flows; the corresponding junction of the several flows in one system is by convention denoted by number 0. Also notice how sources of energy are denoted by SE . Finally, the product of what is above and below each arrow (the effort and the flow) will be that element's instantaneous power, showing how energy is distributed over the system. We will not study bond graphs more complicated than these, nor further explore the ability of this graphical tool to assist in the modelling.

8.3 Dealing with non-linearities

Non-linearities can be classified as hard or soft, as they are respectively more or less severe. Though no uniform definition is universally accepted, we will say that

- a **soft non-linearity** is one that is continuous and differentiable, *Soft non-linearity*
- a **hard non-linearity** is linear almost everywhere, but is not differentiable, or even not continuous. *Hard non-linearity*

Figure 8.5 presents two examples.

Non-linearities are very common. They may be part of the design of a system, even of a control system. In Chapter 28 we will learn how to deal with non-linearities in control systems. What is important here is to notice that soft non-linearities can be approximated by a first order approximation around the operating point. Estimating how large the approximation error may be is important; we will do that in Chapters 12–14.

Example 8.3. In Figure 8.6, mass $m = 10$ kg rests on a non-linear spring and is pulled by force F applied simultaneously on a linear spring with $k = 10^3$ N/m and on a linear damper with $b = 500$ N s/m. The non-linear force of the spring is given by $F_k = 5000 - \frac{500}{\Delta y + 0.1}$ (SI), where the Δy is the variation of length

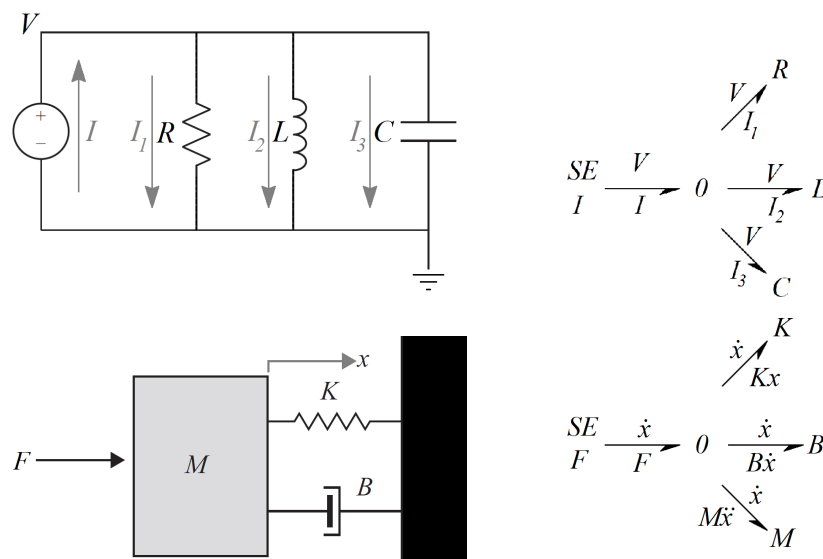


Figure 8.4: Two bond graphs of systems where elements have the same effort and there is a flux junction. Top: electrical circuit. Notice that $I = I_1 + I_2 + I_3$. Bottom: mechanical system. Notice that $F - Kx - B\dot{x} = M\ddot{x} \Leftrightarrow F = Kx + B\dot{x} + M\ddot{x}$.

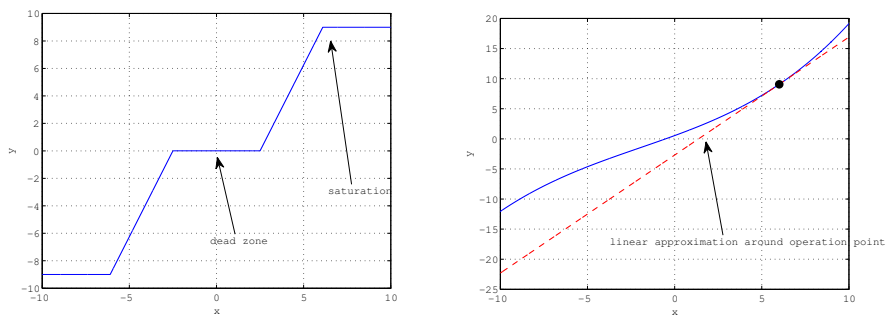


Figure 8.5: Left: hard non-linearities (dead zone and saturation; in practice limits need not be symmetric for positive and negative values, though this assumption is frequent). Right: soft non-linearity and one of its linear approximations.

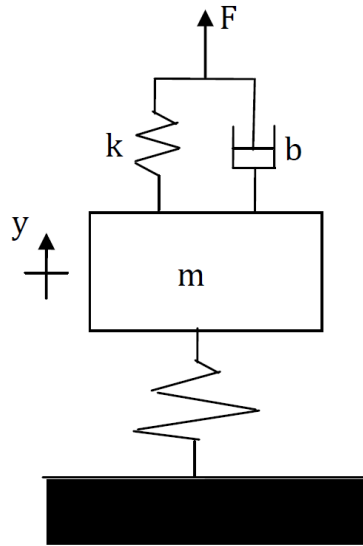


Figure 8.6: System of Example 8.3.

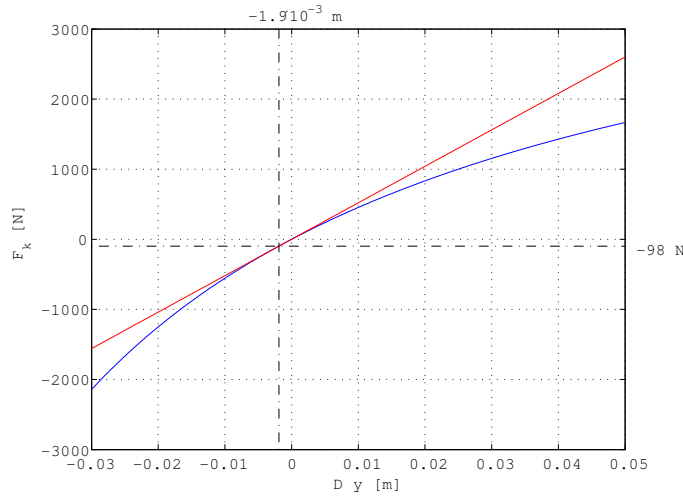


Figure 8.7: Non-linear force of Example 8.3.

around the uncompressed value. We want a linear model for this system around nominal conditions of rest when $F = 0$.

Figure 8.7 shows the non-linear force. When $F = 0$, the non-linear spring is compressed by the weight of m , which is -9.8×10 N (notice the minus sign, since the weight is downwards and the positive sign of y corresponds to an upwards direction), corresponding to

$$-98 = 5000 - \frac{500}{\Delta y + 0.1} \Leftrightarrow \Delta y = -1.9 \times 10^{-3} \text{ m} \quad (8.9)$$

The linearised law is

$$F_k \approx \left. \frac{dF_k}{d(\Delta y)} \right|_{\Delta y = -1.9 \times 10^{-3} \text{ m}} y = \left. \frac{500}{(\Delta y + 0.1)^2} \right|_{\Delta y = -1.9 \times 10^{-3} \text{ m}} y = 5.2 \times 10^4 y \text{ (SI)} \quad (8.10)$$

where $y = \Delta y + 1.9 \times 10^{-3}$ m, or, if you prefer, the variation of length around $\Delta y = -1.9 \times 10^{-3}$ m. Furthermore, the linear components are assumed to have no mass, and hence transmit force F to mass m . Thus

$$m\ddot{y} = F - 5.2 \times 10^4 y \text{ (SI)} \quad (8.11)$$

It should be stressed that linear model (8.11) is only an approximation. \square

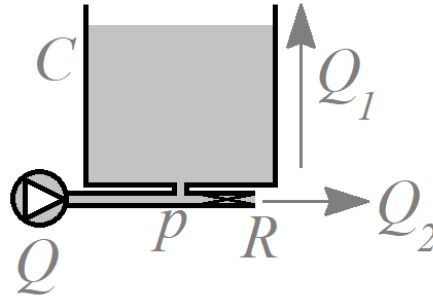


Figure 8.8: System of Exercise 1.

Glossary

Desejoso ainda o Fucarãdono, como mais douto ã os outros, de leuar a sua auante cõ pregũtas ã embarçassẽ o padre, lhe veyo arguindo de nouo ã porã razãõ punha nomes torpes ao Criador de todas as cousas, & aos Sãtos ã no ceo assistião em louuor seu, infamãdoo de mêtiroso, pois elle, como todos criaõ, era Deos de toda a verdade ? & para ã se entenda dõde naceo a este dizer isto, se ha de saber ã na lingoa do Iapaõ se chama a mêtira diusa, & porã o padre quãdo pregaua dezia ã aqella ley ã elle vinha denũciar era a verdadeira ley de Deos, o qual nome elles pela grossaria da sua lingoa nãõ podião pronũciar taõ claro como nos & por dizerẽ Deos deziãõ diũs, daquy veyo que estes seruos do diabo tomaraõ motiuo de dizerẽ aos seus que o padre era demonio em carne ã vinha infamar a Deos pãdo-lhe nome de mentiroso: (...) E porque tambem se saiba a razaõ porque lhe este bonzo disse que punha nomes torpes aos santos, foy, porque tinha o padre por costume quando acabaua de dizer missa rezar com todos hũa Ladaynha para rogar a N. Senhor pela augmẽtaçãõ da fé Catholica, & nesta ladainha dezia sempre, como nella se custuma, *Sancte Petre ora pro nobis, Sancte Paule ora pro nobis*, & assi dos mais Santos. E porã tambem este vocablo santi na lingoa Iapoa he torpe & infame, daquy veyo arguyr este ao padre ã punha maos nomes aos Sãtos, (...) & daly por diãte mãdou o padre ã se naõ dissesse mais *sancte*, senãõ *beate Petre, beate Paule*, & assi aos outros Santos, porque já dantes tinhaõ os bonzos todos perante el Rey feito peçonha disto.

Fernão MENDES PINTO (1509? — †1583), *Peregrinaçam* (1614, posth.),
CCXIII

bond graph grafo de ligação

hard non-linearity não-linearidade severa

linearisation, linearization (US) linearização

soft non-linearity não-linearidade suave

Exercises

1. Draw the bond graph of the system in Figure 8.8.
2. Draw the bond graph of the balloon from Example 8.1.
3. The system in Figure 8.9 is fed by a water pump with a characteristic curve given by $P(t) = 10^5 - 2 \times 10^6 Q(t)$, where P and Q are the pressure (Pa) and the volumetric flow (m^3/s) provided.

The pipe has a 0.01 m^2 cross section and a length of 50 m. Its flow resistance is neglectable; its inertance is not.

The tank has a free surface and 1 m^2 cross-section.

The valve is non-linear and verifies relation

$$Q_v(t) = 0.3 \times 10^{-4} N(t) \sqrt{P_v(t)} \quad (8.12)$$

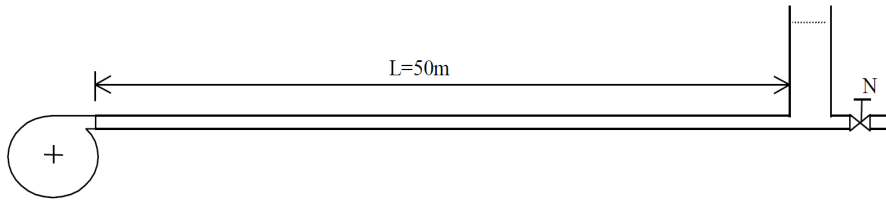


Figure 8.9: System of Exercise 3.

where Q_v is the flow through the valve (m^3/s), N is the opening of the valve (dimensionless), and P_v is the pressure (Pa) at the entrance of the valve, which is also the pressure at the bottom of the tank.

In nominal conditions, $\overline{P}_v = 8 \times 10^4$ Pa and $\overline{Q}_v = 0.01$ m^3/s .

- Show that the pipe's inertance is $L = 5 \times 10^6$ kg m^{-4} .
- Show that in nominal conditions the height of water in the tank is $\bar{h} = 8.16$ m.
- Show that the non-linear relation of the valve (8.12) can be linearised as

$$Q_v(t) = \overline{Q}_v + 0.085 \left(N(t) - \overline{N} \right) + 6.25 \times 10^{-7} \left(P_v(t) - \overline{P}_v \right) \quad (8.13)$$

- Show that the system can be modelled by (8.13) together with

$$\begin{cases} P_v(t) - \overline{P}_v = \rho g (h - \bar{h}) \\ (Q(t) - \overline{Q}) - (Q_v(t) - \overline{Q}_v) = A \frac{d(h - \bar{h})}{dt} \\ (P_b(t) - \overline{P}_b) - (P_v(t) - \overline{P}_v) = L \frac{d(Q(t) - \overline{Q})}{dt} \end{cases} \quad (8.14)$$

- Find transfer function $\frac{\Delta P_v(s)}{\Delta N(s)}$, relating variations around nominal conditions.
- In Figure 8.10, the lever with inertia I oscillates around the horizontal position (i.e. $\theta(t) = 0$) and is moved by torque τ_m . Mass m moves vertically, at distance d from the fulcrum of the lever, inside a cylinder with two springs of constant k , filled with incompressible oil. The pressure difference $\Delta p(t)$ between the two chambers of the cylinder moves the oil through fluidic resistance R . Thanks to oil lubrication, friction inside the cylinder is neglectable.
 - Write linearised equations for the dynamics of the system.
 - Find transfer function $\frac{\Theta(s)}{T_m(s)}$.
 - In Figure 8.11, the lever with inertia I is actuated by force $F(t)$ and supported on the other side by a spring and a damper. On the lever there is a car with mass m , moving to sideways due to gravity, without friction. When $F = 0$ and the car is on the fulcrum (i.e. its position is $x = 0$), the lever remains in the horizontal position. There is no friction at the fulcrum.
 - Write linearised equations for the dynamics of the system.
 - Find transfer function $\frac{X(s)}{F(s)}$.
 - The lever in Figure 8.12, with neglectable mass, is moved by a torque τ applied on the fulcrum, in the absence of which the lever is horizontal (i.e. $\theta = 0$). F is the force exerted on the fulcrum.

It is known that $m_1 = 1.5$ kg, $m_2 = 2.0$ kg, $d_1 = 0.6$ m, $d_2 = 0.4$ m, and $b = 20$ N s/m. The spring obeys the non-linear law in Figure 8.12, where δ is the length variation in mm (with $\delta > 0$ corresponding to compression), and F_m is the resulting force in N.

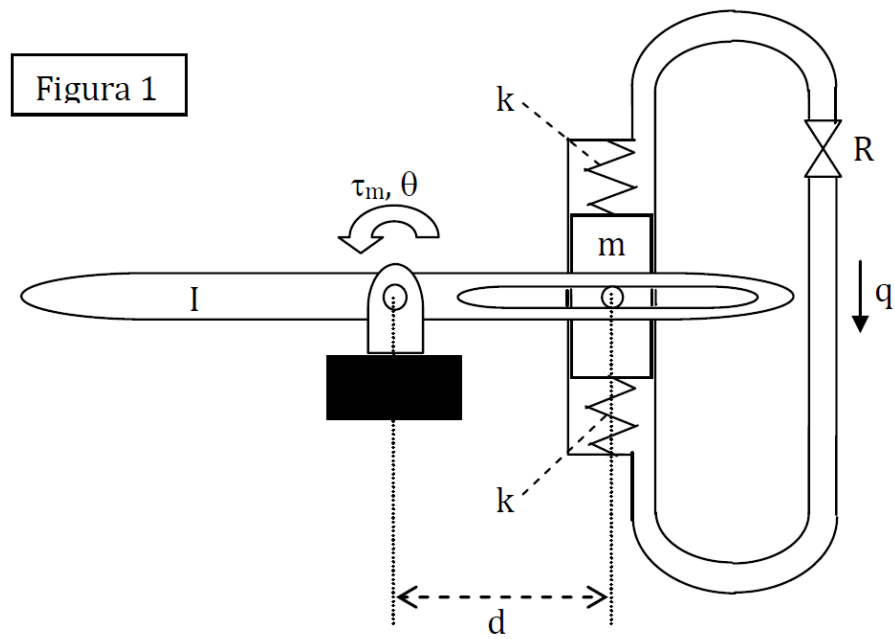


Figure 8.10: System of Exercise 4.

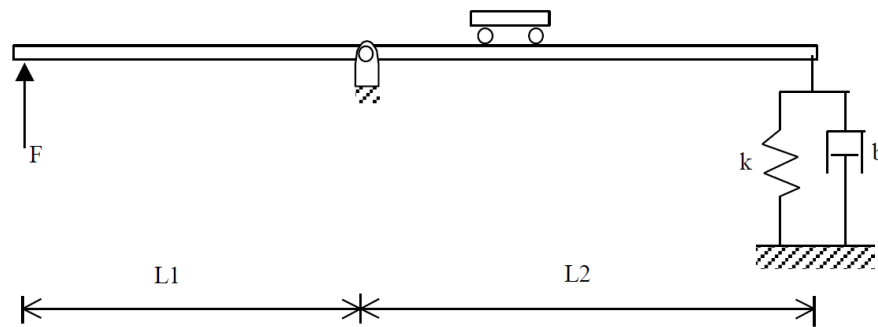


Figure 8.11: System of Exercise 5.

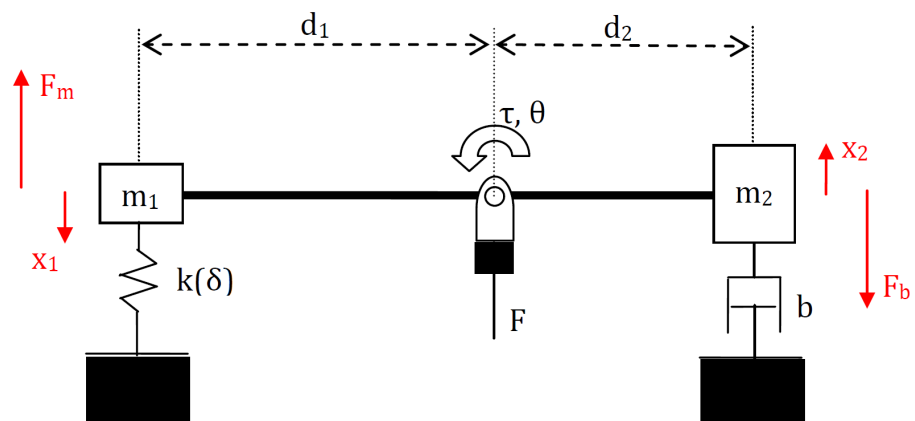


Figure 8.12: System of Exercise 6.

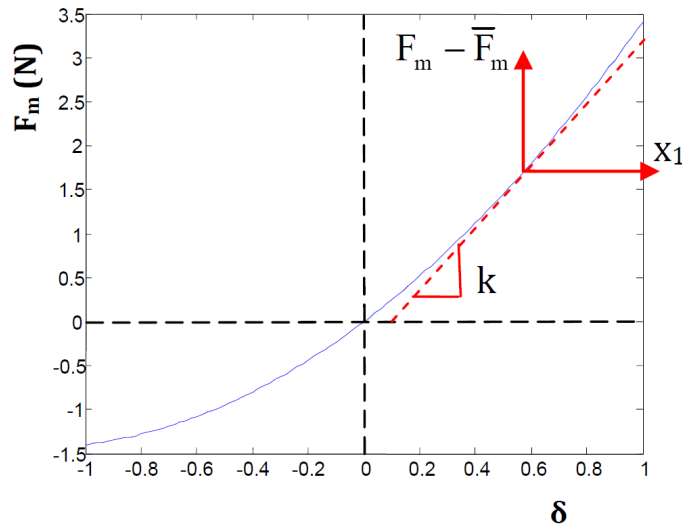


Figure 8.13: Non-linear law of the spring of Exercise 6.

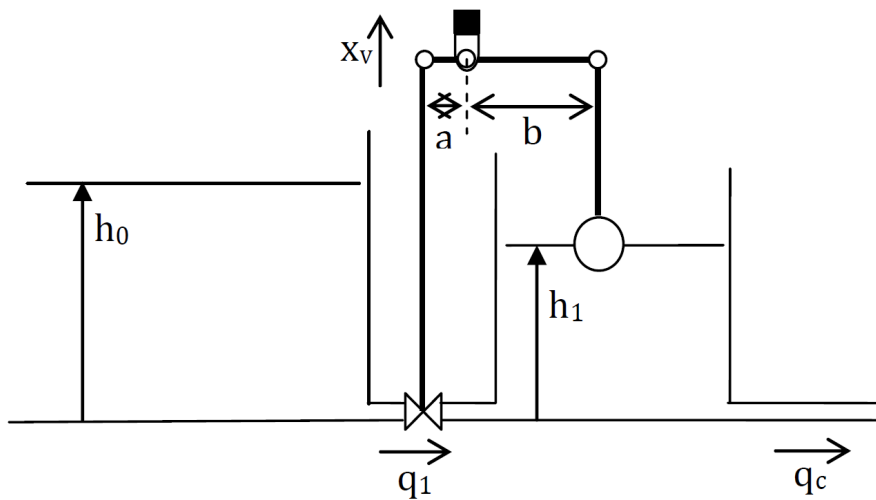


Figure 8.14: System of Exercise 7.

- (a) Show that, in nominal conditions, $\overline{F_m} = 1.63$ N.
- (b) Show from the plot in Figure 8.13 that the force of the spring can be linearised as $F_m = 1.63 + 3.26 \times 10^3 x_1$.
- (c) Write linearised equations for the dynamics of the system.
- (d) Find transfer function $\frac{F(s)}{T(s)}$.
7. In Figure 8.14, the fresh water ($\rho = 1000$ kg/m³) tank in the left is big enough to keep a constant liquid height $h_0 = 5$ m, while the tank in the right has a 10 m² cross-section and a variable liquid height $h_1(t)$. Flow $q_c(t)$ bleeds this tank and does not depend on pressure; flow $q_1(t)$ passes through a non-linear valve that verifies

$$q_1(t) = 0.15x_v(t)\sqrt{\Delta p(t)} \quad (\text{SI}) \quad (8.15)$$

where $\Delta p(t)$ is the pressure difference on both sides of the valve and $x_v(t)$ is mechanically actuated by $h_1(t)$ through a rigid lever with $a = 0.4$ m and $b = 4$ m.

In nominal conditions, $q_c(t) = q_1(t) = 0.2$ m³/s and $h_1(t) = 3$ m.

- (a) Show that the model of the flow through the valve (8.15) can be linearised around nominal conditions as

$$q_1(t) = 21x_v(t) + 5.09 \times 10^{-6} \Delta p \quad (\text{SI}) \quad (8.16)$$

- (b) Write linearised equations for the dynamics of the system.
- (c) Find transfer function $\frac{\Delta H_1(s)}{Q_c(s)}$.

Part II

Systems theory

To study and not think is a waste. To think and not study is dangerous.

CONFUCIUS (c. 551 BC — †c. 479 BC), *Analects* (5th c. BC?), II 15 (transl. A. Charles Muller, 2021)

In this part of the lecture notes:

Chapter 9 develops the very important notion of transfer function of a system through the representation of interconnected systems as blocks, in so-called block diagrams.

Chapter 10 is dedicated to the study of time and frequency responses in general.

Chapter 11 studies the time and frequency responses of different systems.

Here is what you need to know beforehand to follow these chapters:

- The Laplace and Fourier transforms, from Chapter 2;
- Transfer functions, from Sections 4.1 and 4.2 of Chapter 4.

Chapter 9

Transfer functions and block diagrams

In this chapter we will show how transfer functions can be used together with a graphic representation of system interconnection called block diagram. We conclude using block diagrams as a tool for a short introduction to control.

9.1 More on transfer functions

Remember Definition 4.1 about what is a transfer function of a SISO system modelled by a differential equation: it is the ratio between the Laplace transform of the output and the Laplace transform of the input, assuming initial conditions equal to zero. Also remember that behind each transfer function there is a differential equation, and that differential equations are models of real things.

Transfer functions are differential equations

Using transfer functions, we can easily study the behaviour of a system abstracting from its physical reality. This is the approach we will take from now on. This said, notice that remembering the actual system that is being studied can be useful to check if results are possible or not. Remember that models approximate reality, not the other way round. Also remember Example 4.1 about the spring stretched to an impossible length (it breaks, of course), or Remark 6.5 about pipes where the flow cannot be negative. We would not have found that just by looking at our models, which are linear.

Theorem 9.1. The transfer function of a SISO LTI continuous in time can be expressed as the ratio of two polynomials in s . *Ratio of polynomials in s*

Proof. Let the input of the SISO system be $u(t)$ and its output be $y(t)$. Because the system is LTI and continuous in time, it is modelled by a linear differential equation:

$$a_0 y(t) + a_1 \frac{dy(t)}{dt} + a_2 \frac{d^2 y(t)}{dt^2} + a_3 \frac{d^3 y(t)}{dt^3} + \dots = b_0 u(t) + b_1 \frac{du(t)}{dt} + b_2 \frac{d^2 u(t)}{dt^2} + b_3 \frac{d^3 u(t)}{dt^3} + \dots$$

$$\Leftrightarrow \sum_{k=0}^n a_k \frac{d^k y(t)}{dt^k} = \sum_{k=0}^m b_k \frac{d^k u(t)}{dt^k} \quad (9.1)$$

In the last expression, n and m are the highest derivative orders of the equation. Assuming zero initial conditions and applying the Laplace transform, this becomes

$$a_0 Y(s) + a_1 Y(s)s + a_2 Y(s)s^2 + a_3 Y(s)s^3 + \dots = b_0 U(s) + b_1 U(s)s + b_2 U(s)s^2 + b_3 U(s)s^3 + \dots$$

$$\Leftrightarrow \sum_{k=0}^n a_k Y(s)s^k = \sum_{k=0}^m b_k U(s)s^k \quad (9.2)$$

Rearranging terms,

$$\frac{Y(s)}{U(s)} = \frac{b_0 + b_1 s + b_2 s^2 + b_3 s^3 + \dots}{a_0 + a_1 s + a_2 s^2 + a_3 s^3 + \dots} = \frac{\sum_{k=0}^m b_k s^k}{\sum_{k=0}^n a_k s^k} \quad \square \quad (9.3)$$

□

Remark 9.1. Some authors change the order of the coefficients, and instead of (9.3) write

$$\frac{Y(s)}{U(s)} = \frac{\sum_{k=0}^m b_{m-k} s^k}{\sum_{k=0}^n a_{n-k} s^k} \quad (9.4)$$

This is a mere detail of notation. □

Normalising transfer function coefficients

Remark 9.2. (9.3) corresponds to an infinite number of representations of a same transfer function. It suffices to multiply both numerator and denominator by a constant. But it is common to normalise coefficients so that $a_0 = 1$, or $a_n = 1$, or $b_0 = 1$. □

Example 9.1. Consider the microprecision control setup test in Figure 3.9 from Example 3.19. The transfer function from one of the actuators to the position of the mass has been identified as

$$G(s) = \frac{9602}{s^2 + 4.27s + 7627} \quad (9.5)$$

This was normalised so that $a_2 = 1$, $n = 2$. We could also normalise a_0 or b_0 :

$$\begin{aligned} G(s) &= \frac{1.2589}{131.11 \times 10^{-6} s^2 + 559.85 \times 10^{-6} s + 1} \\ &= \frac{1}{104.14 \times 10^{-6} s^2 + 444.70 \times 10^{-6} s + 0.7943} \quad \square \end{aligned} \quad (9.6)$$

Getting the transfer function back

It is easy to find the differential equation from a transfer function. When the transfer function is represented merely by a letter, meaning that it is a function of s , as in (9.5) above, it still corresponds to the ratio of the Laplace transforms of output and input.

Example 9.2. (9.5) can be rewritten as

$$\begin{aligned} \frac{Y(s)}{U(s)} &= \frac{9602}{s^2 + 4.27s + 7627} \Leftrightarrow Y(s)(s^2 + 4.27s + 7627) = 9602U(s) \\ \Leftrightarrow Y(s)s^2 + 4.27Y(s)s + 7627Y(s) &= 9602U(s) \end{aligned} \quad (9.7)$$

which is the Laplace transform of the differential equation governing the plant:

$$y''(t) + 4.27y'(t) + 7627y(t) = 9602u(t) \quad \square \quad (9.8)$$

Proper transfer function

Definition 9.1. A transfer function is **proper** if the order of the polynomial in the numerator is equal to or less than the order of the polynomial in the denominator.

Strictly proper transfer function

A transfer function is **strictly proper** if the order of the polynomial in the numerator is less than the order of the polynomial in the denominator.

In the notation of (9.3), the transfer function is proper if $m \leq n$, and strictly proper if $m < n$. □

For reasons we shall address in Chapters 11 and 25, we will be working almost always with proper transfer functions, and most of the times with strictly proper transfer functions.

Order of a transfer function

Definition 9.2. The **order** of a transfer function is the highest order of the polynomials in the numerator and the denominator. If the transfer function is proper, its order is the order of the denominator. □

Remark 9.3. The order of a transfer function is also the order of the differential equation from which it was formed. In fact, s^k corresponds to a derivative of order k . □

Remark 9.4. Notice that some transfer functions can be simplified because numerator and denominator have common factors. Eliminating them reduces the order of the transfer function. □

Example 9.3. Here are examples of proper transfer functions of:

- Order 0

$$G_a(s) = 20 \quad (9.9)$$

- Order 1

$$G_b(s) = \frac{19}{s + 18} \quad (9.10)$$

$$G_c(s) = \frac{17s + 16}{s + 15} \quad (9.11)$$

$$G_d(s) = \frac{14}{s} \quad (9.12)$$

$$G_e(s) = \frac{13s + 12}{s} \quad (9.13)$$

- Order 2

$$G_f(s) = \frac{11}{s^2 + 10s + 9} \quad (9.14)$$

$$G_g(s) = \frac{8s + 7}{s^2 + 6s + 5} \quad (9.15)$$

$$G_h(s) = \frac{4s^2 + 3s + 2}{s^2 + s - 1} \quad (9.16)$$

$$G_i(s) = \frac{s^2 - 2s + 1}{s^2} \quad (9.17)$$

$$G_j(s) = \frac{s^2 - 3s - 4}{s^2 - 5s - 6} \quad (9.18)$$

They have all been normalised so that the coefficient of the highest order monomial in the denominator is 1 (i.e. $a_n = 1$). Transfer functions $G_b(s)$, $G_d(s)$, $G_f(s)$, $G_g(s)$, and $G_i(s)$ are strictly proper; the other ones are not.

$G_j(s)$ is of order 2 but can be simplified and become of order 1:

$$G_j(s) = \frac{s^2 - 3s - 4}{s^2 - 5s - 6} = \frac{(s - 4)(s + 1)}{(s - 6)(s + 1)} = \frac{s - 4}{s - 6} \quad \square \quad (9.19)$$

Transfer functions are often put in the following form, that explicitly shows the **zeros** of the transfer function (i.e. the zeros of the polynomial in the numerator) and the **poles** of the transfer function (i.e. the zeros of the polynomial in the denominator):

$$\frac{Y(s)}{U(s)} = \frac{b_m(s - z_1)(s - z_2)(s - z_3) \dots}{a_n(s - p_1)(s - p_2)(s - p_3) \dots} = \frac{b_m \prod_{k=1}^m (s - z_k)}{a_n \sum_{k=0}^n (s - p_k)} \quad (9.20)$$

Here the zeros are z_k , $k = 1, 2, \dots, m$ and the poles are p_k , $k = 1, 2, \dots, n$. Because both inputs and outputs are real, transfer function coefficients are real, and consequently the poles and zeros are either real or pairs of complex conjugates. (Remember Remark 2.6.) So in (9.20) it is usual to multiply such pairs, presenting a second order term instead of two complex terms.

Example 9.4. The second order transfer functions in Example 9.3 can be rewritten as

$$G_f(s) = \frac{11}{(s + 9)(s + 1)} \quad (9.21)$$

$$G_g(s) = \frac{8s + 7}{(s + 5)(s + 1)} \quad (9.22)$$

$$G_h(s) = \frac{4 \left(s + \frac{3 + \sqrt{23}j}{8} \right) \left(s + \frac{3 - \sqrt{23}j}{8} \right)}{\left(s + \frac{1 + \sqrt{5}}{2} \right) \left(s + \frac{1 - \sqrt{5}}{2} \right)} = \frac{4s^2 + 3s + 2}{\left(s + \frac{1 + \sqrt{5}}{2} \right) \left(s + \frac{1 - \sqrt{5}}{2} \right)} \quad (9.23)$$

$$G_i(s) = \frac{(s - 1)^2}{s^2} \quad (9.24)$$

For $G_j(s)$, see (9.19). Notice that, in the case of $G_h(s)$, only the second expression is usual; the first one, explicitly showing the two complex conjugate zeros, is not. \square

Remark 9.5. From Definition 9.2 results that the order of a proper transfer function is the number of its poles. \square

The following MATLAB functions use transfer functions in this form:

Transfer function from zeros, poles, gain

- `zpk` creates a transfer function from its zeros, poles, and the $\frac{b_m}{a_n}$ ratio in (9.20), here called gain k , and also converts a transfer function created with `tf` into this form;
- `pole` finds the poles of a transfer function;
- `tzero` finds the zeros of a transfer function.

Example 9.5. Transfer function (9.15) or (9.22)

- has one zero, $8s + 7 = 0 \Leftrightarrow s = -\frac{7}{8} = -0.875$,
- has two poles, $(s + 5)(s + 1) = 0 \Leftrightarrow s = -5 \vee s = -1$,
- verifies $k = \frac{b_m}{a_n} = \frac{8}{1} = 8$.

MATLAB's command `zpk`

It can be created, converted to a ratio of two polynomials as in (9.3), and converted back to the (9.20) form as follows:

```
>> G_g = zpk(-7/8, [-5 -1], 8)
```

```
G_g =
```

$$\frac{8 (s+0.875)}{(s+5) (s+1)}$$

Continuous-time zero/pole/gain model.

```
>> G_g = tf(G_g)
```

```
G_g =
```

$$\frac{8 s + 7}{s^2 + 6 s + 5}$$

Continuous-time transfer function.

```
>> G_g = zpk(G_g)
```

```
G_g =
```

$$\frac{8 (s+0.875)}{(s+5) (s+1)}$$

Continuous-time zero/pole/gain model.

MATLAB's commands
pole and tzero

Its poles and zeros can be found as follows:

```
>> tzero(G_g)
```

```
ans =  
-0.8750
```

```
>> pole(G_g)
```

```
ans =  
-5  
-1
```

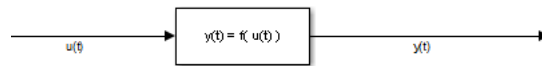



Figure 9.1: Generic block.

It does not matter whether a transfer function was created with `tf` or with `zpk` (or with any other function to create transfer functions that we did not study yet): `pole` and `tzero` work just the same.

Another way of finding the poles and the zeros is to access the numerator and the denominator, and then using `roots` to find the roots of these polynomials. The transfer function must be in the `tf` form this time, the only one that has the `num` and `den` fields:

```
>> G_g = tf(G_g);
>> G_g.num{1}
ans =
     0     8     7
>> roots(ans)
ans =
    -0.8750
>> G_g.den{1}
ans =
     1     6     5
>> roots(ans)
ans =
    -5
    -1
```

Notice that the `{1}` is necessary since MATLAB presumes that the transfer function is MIMO and thus has many transfer functions relating the many inputs with the many outputs. The cell array index accesses the first transfer function, which, as the system is SISO, is the only one. \square

A very important property of transfer functions for the rest of this chapter has already been mentioned in Section 8.2 and illustrated in Example 8.2: if two systems $G_1(s) = \frac{y_1(s)}{u_1(s)}$ and $G_2(s) = \frac{y_2(s)}{u_2(s)}$ are interconnected so that the output of one is the input of the other, $y_1(s) = u_2(s)$, then the resulting transfer function is

$$\frac{y_2(s)}{u_1(s)} = \frac{y_2(s)}{u_2(s)} \frac{y_1(s)}{u_1(s)} = G_1(s) G_2(s) \quad (9.25)$$

Multiplying transfer functions

Remark 9.6. Remember that the multiplication of two Laplace transforms does not correspond to the multiplication of the original functions, but rather to their convolution, as we have shown in (2.78). Operation convolution is defined in (2.76). (This is sometimes a source of confusion, because the sum of two Laplace transforms is the sum of the original functions, as \mathcal{L} is linear.) \square

Multiplication of \mathcal{L} is convolution in t

9.2 Block diagrams

Block diagrams are graphical representations of the relations between variables and functions. In our case, functions will be systems, and variables will be signals (which are themselves, as you remember, functions of time, or space). Figure 9.1 shows a generic system (represented by a block) relating two signals (represented by lines with arrows).

The practical thing to do for LTI systems is to represent them using their transfer functions, and consequently to represent signals by their Laplace transforms. The block in Figure 9.2 means that $Y(s) = G(s)U(s)$. This is yet another advantage of using the Laplace transform: the (Laplace transform of the) output is the product of the (transfer function of the) system and the (Laplace transform of the) input.

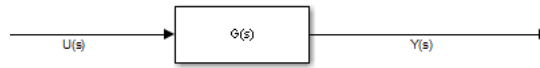


Figure 9.2: Linear block.

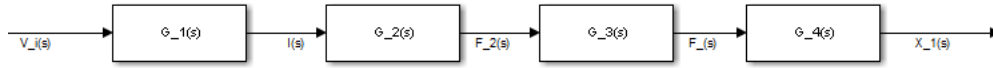


Figure 9.3: Block diagram of Example 9.6, corresponding to the mechatronic system in Figure 8.2 from Example 8.2.

Example 9.6. The mechatronic system in Example 8.2 had four transfer functions, as follows:

$$G_1(s) = \frac{I(s)}{V_i(s)} = \frac{\frac{n_2}{n_1}}{R + Ls} \tag{9.26}$$

$$G_2(s) = \frac{F_2(s)}{I(s)} = \alpha \tag{9.27}$$

$$G_3(s) = \frac{F_1(s)}{F_2(s)} = \frac{b}{a} \tag{9.28}$$

$$G_4(s) = \frac{X_1(s)}{F_1(s)} = \frac{1}{m_1s^2 + K} \tag{9.29}$$

The corresponding block diagram is shown in Figure 9.3. In fact,

$$I(s) = G_1(s)V_i(s) \tag{9.30}$$

$$F_2(s) = G_2(s)I(s) \tag{9.31}$$

$$F_1(s) = G_3(s)F_2(s) \tag{9.32}$$

$$X_1(s) = G_4(s)F_1(s) \quad \square \tag{9.33}$$

The Example above shows that several interconnected systems correspond to a sequence of blocks. By similarity with electrical circuits, blocks in such a sequence are said to be in **series** or in **cascade**. This is because of the property of transfer functions illustrated in (9.25). Clearly, two blocks A and B in series are equivalent to one block AB .

Adding signals is represented as shown in Figure 9.4, where

$$y = y_1 + y_2 = Au + Bu = (A + B)u \tag{9.34}$$

By similarity with electrical circuits, blocks A and B are said to be in **parallel**. Clearly, they are equivalent to one block $A + B$. Signal subtraction is indicated similarly.

The block configurations in Figure 9.5, wherein the input of a block depends on its output, is called **feedback loop** or just **feedback**: feedback, because the output is fed back to the block it originates from; and loop, because of the configuration of the diagram. In that Figure, A is called **direct branch** and B **feedback branch**. The two block diagrams only differ because of the sign affecting signal $d(s)$:

- when $b(s) = a(s) - d(s)$, there is **negative feedback**;
- when $b(s) = a(s) + d(s)$, there is **positive feedback**.

Negative feedback is far more common; when feedback is mentioned without specifying whether it is positive or negative, you can safely presume it is negative. Notice that, for both:

Blocks in series
Blocks in cascade

Blocks in parallel

Feedback

Loop

Direct branch

Feedback branch

Negative feedback

Positive feedback

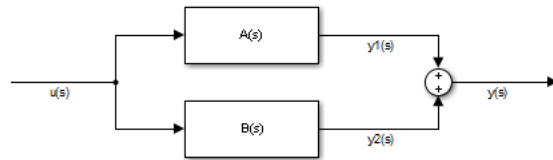


Figure 9.4: Block diagram with two blocks in parallel.

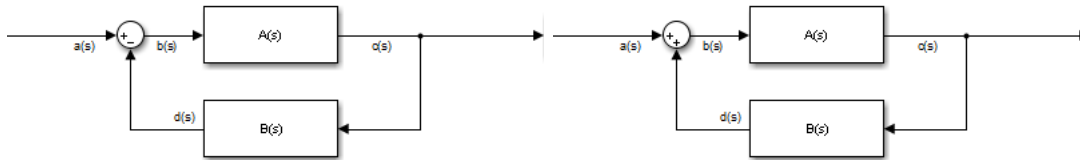


Figure 9.5: Block diagrams with feedback loops. Left: negative feedback. Right: positive feedback.

Input of the feedback loop

- the input of the loop is $a(s)$;
- the output of the loop is $c(s)$;
- the input of the direct branch is $b(s) = a(s) \mp d(s)$;
- the output of the direct branch is $a(s)$;
- the input of the feedback branch is $c(s)$;
- the output of the feedback branch is $d(s)$.

Output of the feedback loop

Consequently, for negative feedback,

$$\begin{aligned}
 c &= Ab = A(a - d) = A(a - Bc) = Aa - ABc \\
 \Rightarrow c + ABc &= Aa \Rightarrow c = a \frac{A}{1 + AB}
 \end{aligned}
 \tag{9.35}$$

and, for positive feedback,

$$\begin{aligned}
 c &= Ab = A(a + d) = A(a + Bc) = Aa + ABc \\
 \Rightarrow c - ABc &= Aa \Rightarrow c = a \frac{A}{1 - AB}
 \end{aligned}
 \tag{9.36}$$

Example 9.7. The centrifugal governor (see Figure 9.6) is a control system which had widespread use to control the pressure in boilers. It rotates because of the pressure of the steam. The faster it rotates, the more the two spheres go up, thereby opening a valve relieving steam pressure. Consequently the regulator spins slower, the balls go down, and this closes the valve, so pressure is no longer relieved and goes up again. This is negative feedback: an increase of any variable has as consequence the decrease of another variable that caused the original increase, and vice-versa. \square

Centrifugal governor

Example 9.8. Audio feedback (or “howl”) is an example of positive feedback. Surely you must have heard it often, whenever there is a sound system amplifying the sound detected by a microphone which is too close to the loudspeakers, so that even background noise is amplified to the point of being received again by the microphone and amplified further — see Figure 9.7. The amplitude of the resulting sound does not become infinite because at some point the amplifier and/or the loudspeakers saturate, but the “howl” can damage the equipment or, more importantly, the listeners’ auditory systems. \square



Figure 9.6: Centrifugal governor of a boiler in the former Barbadinhos water pumping station (currently the Water Museum), Lisbon.

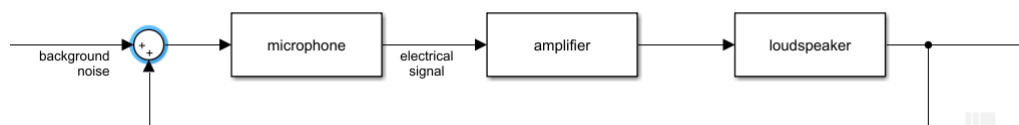


Figure 9.7: How audio feedback occurs.

Example 9.9. Biological processes provide numerous examples of both positive and negative feedback. We will go back to this in Chapter 43. \square

The best way to simplify block diagrams is to write the corresponding equations and do so analytically.

Example 9.10. In the block diagram of Figure 9.8 we make

$$G_1(s) = 2 \quad (9.37)$$

$$G_2(s) = \frac{s + 10}{s^2 + 0.5s + 5} \quad (9.38)$$

$$G_3(s) = \frac{1}{s + 1} \quad (9.39)$$

$$G_4(s) = \frac{20(s - 0.5)}{(s - 1)(s - 3)} \quad (9.40)$$

$$G_5(s) = \frac{1}{s} \quad (9.41)$$

$$(9.42)$$

(The block for $G_1(s)$ is triangular because SIMULINK, which we will mention below, uses triangles for constants, but this convention is unusual; when drawing blocks by hand, they are all usually rectangles.) Then

$$\begin{aligned} e = G_2c = G_2G_1b = G_2G_1(a - d) = G_1G_2(a - G_3e) &\Rightarrow \\ \Rightarrow (1 + G_1G_2G_3)e = G_1G_2a &\Rightarrow e = a \frac{G_1G_2}{1 + G_1G_2G_3} \end{aligned} \quad (9.43)$$

$$h = G_5g = G_5(e + f) = G_5 \left(a \frac{G_1G_2}{1 + G_1G_2G_3} + G_4a \right) = a \left(\frac{G_1G_2G_5}{1 + G_1G_2G_3} + G_4G_5 \right) \quad (9.44)$$

Finally, the whole block diagram corresponds to transfer function

$$\frac{h(s)}{a(s)} = \frac{2 \frac{s+10}{s^2+0.5s+5} \frac{1}{s}}{1 + 2 \frac{s+10}{s^2+0.5s+5} \frac{1}{s+1}} + \frac{20(s-0.5)}{s(s-1)(s-3)} \quad (9.45)$$

It is usually a good idea to put the result in one of the forms (9.3) or (9.20). Since calculations are rather complicated, we can use MATLAB:

```
>> s = tf('s');
>> (2/s*(s+10)/(s^2+0.5*s+5))/(1+2/(s+1)*(s+10)/(s^2+0.5*s+5))+...
20*(s-0.5)/((s-1)*(s-3)*s)
```

ans =

$$22 s^7 + 45 s^6 + 200 s^5 + 617.5 s^4 + 380.5 s^3 + 1960 s^2 - 950 s$$

$$s^9 - 2 s^8 + 8.25 s^7 - 10.75 s^6 - 55.25 s^5 + 33.75 s^4 - 350 s^3$$

$$+ 375 s^2$$

Continuous-time transfer function.

```
>> zpk(ans)
```

ans =

$$22 s (s+2.931) (s-0.4226) (s^2 + 0.5s + 5) (s^2 - 0.9629s + 6.972)$$

$$s^2 (s-3) (s+2.5) (s-1) (s^2 + 0.5s + 5) (s^2 - s + 10)$$

Continuous-time zero/pole/gain model.

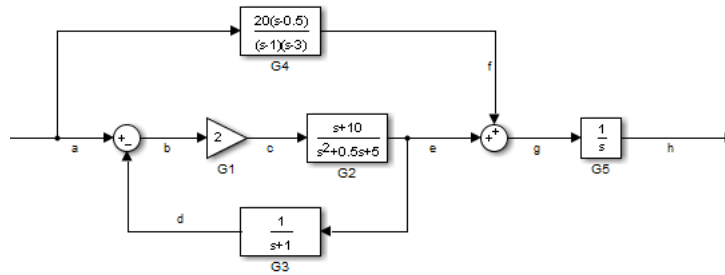


Figure 9.8: Block diagram of Example 9.10.

As you can see from the last result, it is possible to eliminate s and $s^2+0.5*s+5$ from both the numerator and the denominator. So $\frac{h(s)}{a(s)}$ is of sixth order. \square

In this way it is possible to find several generic equivalences in block diagrams, such as those of Figure 9.9, which may be used in block diagram simplification. The analytical simplification of block diagrams is however normally easier and less prone to errors.

MATLAB has commands to combine transfer functions:

- operators `+` and `*` add and multiply transfer functions (remember that two blocks in series correspond to the product of their transfer functions);
- `feedback` receives the direct and the feedback branches and gives the transfer function of the negative feedback loop.

MATLAB's
feedback

command **Example 9.11.** We can verify our calculations of Example 9.10 as follows:

```
>> G1 = 2;
>> G2 = (s+10)/(s^2+0.5*s+5);
>> G3 = 1/(s+1);
>> G4 = 20*(s-0.5)/((s-1)*(s-3));
>> G5 = 1/s;
>> loop_from_a_to_e = feedback(G1*G2, G3)
```

```
loop_from_a_to_e =
```

$$\frac{2 s^2 + 22 s + 20}{s^3 + 1.5 s^2 + 7.5 s + 25}$$

Continuous-time transfer function.

```
>> from_a_to_g = loop_from_a_to_e + G4
```

```
from_a_to_g =
```

$$\frac{22 s^4 + 34 s^3 + 73 s^2 + 411 s - 190}{s^5 - 2.5 s^4 + 4.5 s^3 - 0.5 s^2 - 77.5 s + 75}$$

Continuous-time transfer function.

```
>> from_a_to_h = from_a_to_g * G5
```

```
from_a_to_h =
```

$$\frac{22 s^4 + 34 s^3 + 73 s^2 + 411 s - 190}{s^6 - 2.5 s^5 + 4.5 s^4 - 0.5 s^3 - 77.5 s^2 + 75 s}$$

Transformation	Equation	Block diagram	Equivalent block diagram
1 Cascaded blocks	$Y = (P_1 P_2)X$		
2 Combining blocks in parallel	$Y = P_1 X \pm P_2 X$		
3 Removing a block from a forward loop	$Y = P_1 X \pm P_2 X$		
4 Eliminating a feedback loop	$Y = P_1(X \mp P_2 Y)$		
5 Removing a block from a feedback loop	$Y = P_1(X \mp P_2 Y)$		
6 Rearranging summing junctions	$Z = W \pm X \pm Y$		
7 Moving a summing junction in front of a block	$Z = PX \pm Y$		
8 Moving a summing junction beyond a block	$Z = P(X \pm Y)$		
9 Moving a take-off point in front of a block	$Y = PX$		
10 Moving a take-off point beyond a block	$Y = PX$		
11 Moving a take-off point in front of a summing junction	$Z = W \pm X$		
12 Moving a take-off point beyond a summing junction	$Z = X \pm Y$		

Figure 9.9: Block diagram simplification.



Figure 9.10: Commonly used blocks of SIMULINK.

Continuous-time transfer function.

```
>> zpk(from_a_to_h)
```

```
ans =
```

```

22 (s+2.931) (s-0.4226) (s^2 - 0.9629s + 6.972)
-----
s (s+2.5) (s-3) (s-1) (s^2 - s + 10)

```

Continuous-time zero/pole/gain model.

This is the same transfer function we found above, with the poles and zeros common to the numerator and denominator eliminated. \square

SIMULINK

MATLAB's most powerful tool for working with block diagrams is SIMULINK. All the block diagrams above have been created with SIMULINK, and then cropped so as not to show what SIMULINK calls source and sink, which are not part of what is shown in standard block diagrams (you must not include them when drawing block diagrams by hand). To use SIMULINK, access its library in MATLAB by clicking the corresponding button or typing `simulink`. The library looks like very different in different versions of MATLAB, but its organisation is similar: the most commonly used blocks are in one of the several subsets of the `Simulink` library; then there are libraries corresponding to the toolboxes you have installed. Figure 9.10 shows the blocks you will likely need:

Commonly used SIMULINK blocks

- The `Transfer Fcn` block, from the `Continuous` subset of the `Simulink` library, creates a transfer function like function `tf`.
- The `Zero-Pole` block, from the `Continuous` subset of the `Simulink` library, creates a transfer function like function `zpk`.
- The `LTI System` block, from the `Control System Toolbox` library, creates a transfer function using function `tf` or function `zpk`. It is also possible just to put there a variable with a transfer function, created in the command line.
- The `Sum` block (name hidden by default), from the `Math operations` subset of the `Simulink` library, is a sum point.
- The `Gain` block, from the `Math operations` subset of the `Simulink` library, multiplies a signal by a constant.
- The `From Workspace` block, from the `Sources` subset of the `Simulink` library, provides a signal to run a simulation. The signal is either a structure (see the block's dialogue for details) or a matrix with time instants in the first column and the corresponding values of the signal in the second column (these will be interpolated).
- The `Scope` block, from the `Sinks` subset of the `Simulink` library, plots the signal it receives. It can be configured to record the data to a variable in the workspace, which is most practical to reuse it later.
- The `Mux` block (from “multiplexer”, name hidden by default), from the `Signal Routing` subset of the `Simulink` library, joins two (or more) signals into one. The result is a vector-valued signal. If two real-values signals are multiplexed, the result is a vector-valued signal with dimension 2.



Figure 9.11: SIMULINK file of Example 9.12.

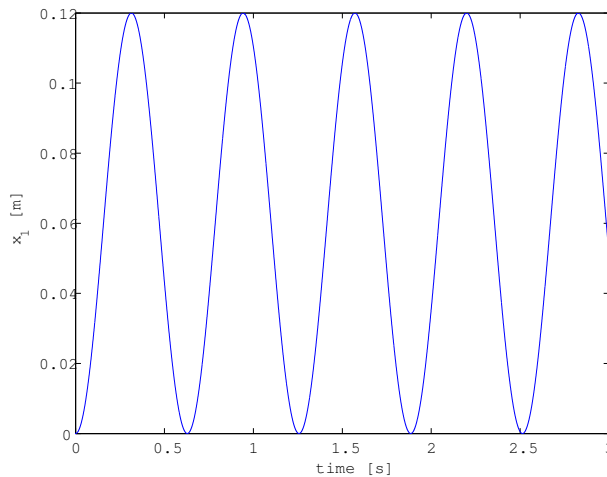


Figure 9.12: Output of Example 9.12.

To use a block, create an empty SIMULINK file and drag it there. Connect blocks with arrows by clicking and dragging from one block's output to another's input. Double-click a block to see a dialogue where you can fill in the arguments you would use in a MATLAB command written in the command line (i.e. after the `>>`). Most of the times you can use numbers or variables; you just have to create the variables before you create the model. Right-clicking a block shows a context menu with many options, among which those of showing or hiding the block's name, or rotating it. You can edit a block's name by clicking it, and add a label to a signal by double-clicking it.

To run a simulation, choose its duration in the box on the top of the window, or go to **Simulation > Model Configuration Parameters**. Then click the Play button, or use command `sim` with the name of the (previously saved) file with the block diagram model.

Example 9.12. Let us simulate the mechatronic system of Examples 8.2 and 9.6, given by (9.26)–(9.29). The SIMULINK file is as shown in Figure 9.11 and its running time was set to 3 s; variables have been used and must be defined before running the simulation, but this means that they are easier to change. Block **From Workspace** has matrix $\begin{bmatrix} 0 & 1 \end{bmatrix}$, meaning that at time 0 it will output value 1, and since no other value is provided this one will be kept. So we are finding the response of the system to a Heaviside function (2.5), or rather to a tension of 1 V being applied when the simulation begins. Block **Scope** is configured to save data to variable `Data`. The following commands create the variables, run the simulation, and plot again the results which you could also see in the Scope itself:

```
>> n2 = 200; n1 = 100; L = 1e-2; R = 100; alpha = 100; b = 0.3; a = 0.1;
m1 = 1; K = 100;
>> sim('prob3_ficha3_2011_modif_2')
>> figure, plot(Data.time,Data.signals.values(:,2)), xlabel('time [s]'),
ylabel('x_1 [m]')
```

See Figure 9.12. We could have expected these oscillations with constant amplitude, and you will know why in Chapter 11. \square

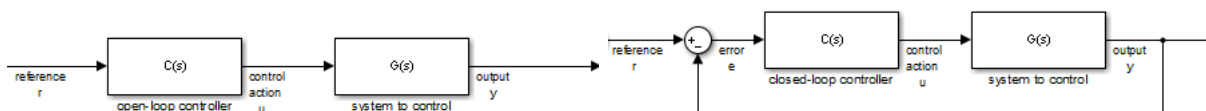


Figure 9.13: Left: open loop control. Right: closed loop control.

Remark 9.7. Notice that the input signal was specified in time and the output variable was obtained as a function of time, but the differential equations were specified as transfer functions, i.e. not as relations in variable t but in the Laplace transform variable s . This is the way SIMULINK works. However, do not forget that, since in a block diagram system dynamics is indicated by transfer functions, signals too must be given by their Laplace transforms, as functions of s . It is correct to say that $y(s) = G(s)u(s)$; it makes no sense at all to write $y(t) = G(s)u(t)$ mixing t and s . \square

The dialogue **Model Configuration Parameters**, which can also be accessed through a button, allows specifying many other things, among which:

- the numerical method used to solve the differential equations;
- the maximum and minimum time steps used by the numerical method;
- a tolerance that will not be exceeded by the numerical method's estimate of the errors incurred.

Notice that some numerical methods use fixed time steps. These may be used with differential equations, but are the only ones that can be used with difference equations (corresponding to digital models).

9.3 Control in open-loop and in closed-loop

There are two generic configurations for control systems: **open-loop control** and **closed-loop control**, shown in Figure 9.13. Every control system is a variation of one of these two configurations, or a combination thereof. Both add, to the system we want to control, another system called **controller**, intended to make the controlled system's output $y(t)$ follow some specified reference, or desired output, $r(t)$ (remember Section 3.1). In a perfectly controlled system, $y(t) = r(t)$, $\forall t$. The output of the controller is the system's input in the strict sense (the input must be a manipulated variable).

Open-loop control

In open-loop control the controller receives the reference that the system should follow, and decides from this desired output what control action to take. This control action will be the input of the system. It is not checked whether or not the system's output does follow the reference. So, if there is some unexpected deviation from the reference, this does not change the control action. Open-loop control only uses blocks in series.

In open-loop control,

$$y(s) = G(s)u(s) = G(s)C(s)r(s) \quad (9.46)$$

and since we want $y(s) = r(s)$ then we should have $C(s) = G^{-1}(s)$, i.e. the controller should be an **inverse model** of the system to control. Notice that if the model of the system is proper then the controller is not proper; you will learn why this brings problems in Chapter 11.

Closed-loop control

Closed-loop control uses negative feedback. The reference is compared with the system output. Ideally, the error should be zero. What the controller receives is this error, so the control action is based on the error.

Proportional control

The simplest closed-loop controller is proportional: $C(s) = K \in \mathbb{R}$. With **proportional control**, if the error is small, the control action is small too; if the error is large, the control action is also large. There are techniques to choose an appropriate value of K , and also to develop more complex controllers, with poles and zeros, which will be addressed in Part IV.

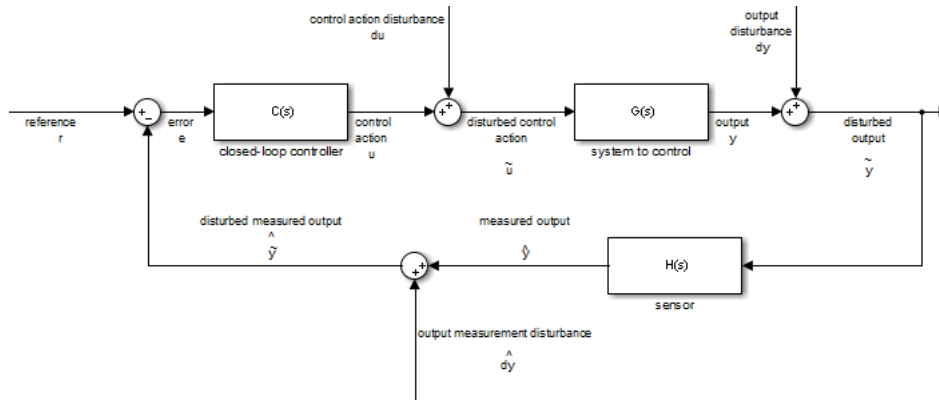


Figure 9.14: Closed-loop control with disturbances and sensor dynamics.

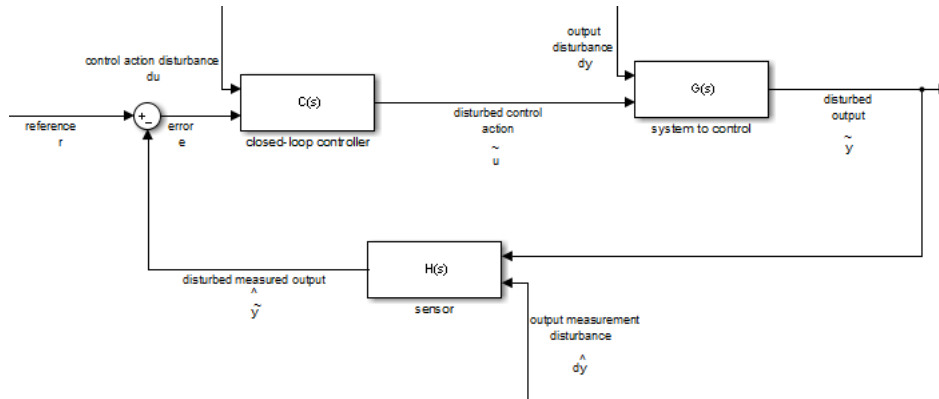


Figure 9.15: The same as Figure 9.14, but with MIMO systems.

Actually, no control system is that simple. Figure 9.14 shows a more realistic situation, including the following additions:

- $H(s)$ is the **sensor** that measures output y . A perfect sensor measures the output exactly: $\hat{y}(t) = y(t), \forall t$; and hence $H(s) = 1$. No sensor is perfect, but it is often possible to assume $H(s) = 1$ even so (in which case the block does not need to be there). If this is not the case, $H(s)$ must be explicitly taken into account. *Sensor dynamics*
- $d_u(t)$ is a disturbance that affects the control action. This means that the control action is not precisely received by the controlled system. For instance, if the control action is a force, this means that there are other forces acting upon the system. Or, if the control action is a current, there are unintended fluctuations of the value determined by the controller. *Control action disturbance*
- $d_y(t)$ is a disturbance that affects the system output. This means that the output is affected by something else other than the system. For instance, if the output is a flow, there is some other source of fluid, or some bleeding of fluid somewhere, that must be added or subtracted. Or, if the output is a position, there may be vibrations that have to be superimposed. *System output disturbance*
- $d_{\hat{y}}(t)$ is a disturbance that affects the sensor's measurement of the system output. Just like $u(t)$ can suffer a disturbance, so can $\hat{y}(t)$. *Output measurement disturbance*

Remark 9.8. Disturbances in Figure 9.14 follow what is called an additive model, since the disturbance is added to the signal it disturbs. Other models use multiplicative disturbances, that are multiplied rather than summed. Here we will stick to additive disturbances, which result in linear models. *Additive disturbances*
Multiplicative disturbances □

Remark 9.9. We saw in Chapter 3 that MIMO systems may have some inputs in the general sense that are disturbances and others that are manipulated variables. Figure 9.15 represents disturbances using MISO systems. The block diagram in Figure 9.14 reflects the same situation using only SISO systems. The

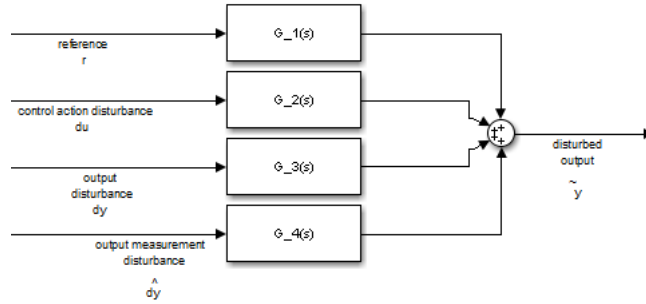


Figure 9.16: The same as Figure 9.14, but using transfer functions (9.48)–(9.51).

price to pay for using SISO systems is less freedom in establishing mathematical relations between disturbances and outputs. \square

The output of the block diagram in Figure 9.14 is

$$\begin{aligned}
 \tilde{y} &= d_y + y = d_y + G\tilde{u} = d_y + G(d_u + u) = d_y + Gd_u + GCe \\
 &= d_y + Gd_u + GC(r - \hat{y}) = d_y + Gd_u + GCr - GC(d_{\hat{y}} + \hat{y}) \\
 &= d_y + Gd_u + GCr - GCd_{\hat{y}} - GCH\tilde{y} \\
 \Rightarrow (1 + GCH)\tilde{y} &= d_y + Gd_u + GCr - GCd_{\hat{y}} \quad (9.47) \\
 \Rightarrow \tilde{y} &= \frac{1}{1 + GCH}d_y + \frac{G}{1 + GCH}d_u + \frac{GC}{1 + GCH}r + \frac{-GC}{1 + GCH}d_{\hat{y}}
 \end{aligned}$$

Because of the linearity of the relations involved, (9.47) gives the same result as if four transfer functions were involved as seen in Figure 9.16:

$$G_1 = \frac{\tilde{y}}{r} = \frac{GC}{1 + GCH} \quad (9.48)$$

$$G_2 = \frac{\tilde{y}}{d_u} = \frac{G}{1 + GCH} \quad (9.49)$$

$$G_3 = \frac{\tilde{y}}{d_y} = \frac{1}{1 + GCH} \quad (9.50)$$

$$G_4 = \frac{\tilde{y}}{d_{\hat{y}}} = \frac{-GC}{1 + GCH} \quad (9.51)$$

Notice that each of the four transfer functions above can be obtained assuming that all inputs but one of them are zero. If the system were not linear, that would not be the case.

Glossary

D'altra parte gli aveva detto la sera prima che lui possedeva un'dono: che gli bastava udire due che parlavano in una lingua qualsiasi, e dopo un poco era capace di parlare come loro. Dono singolare, che Niceta credeva fosse stato concesso solo agli apostoli.

Umberto Eco (1932 — †2016), *Baudolino* (2000), 2

block diagram diagrama de blocos

blocks in cascade blocos em cascata

blocks in parallel blocos em paralelo

closed-loop anel fechado, malha fechada

direct branch ramo direto

disturbance perturbação

feedback retroação

feedback branch ramo de retroação

feedback loop anel de retroação, malha de retroação

inverse model modelo inverso

open-loop anel aberto, malha aberta

order ordem

proper transfer function função de transferência própria

proportional control controlo proporcional

strictly proper transfer function função de transferência estritamente própria

Exercises

1. For each of the transfer functions below, answer the following questions:

- What are its poles?
- What are its zeros?
- What is its order?
- Is it a proper transfer function?
- Is it a strictly proper transfer function?
- What is the differential equation it corresponds to?

(a) $\frac{s}{s^2 + 12s + 20}$

(b) $\frac{s + \frac{1}{5}}{s - 5}$

(c) $\frac{s^2 + 2s + 10}{s^3 - 5s^2 + 15.25s}$

(d) $\frac{10}{(s + 1)^2(s^2 + 5s + 6)}$

(e) $\frac{s^2 + 2}{s^2(s + 3)(s + 50)}$

(f) $\frac{(s^4 + 6s^3 + 8.75s^2)}{(s^2 + 4s + 4)^2}$

2. Find the following transfer functions for the block diagram in Figure 9.17:

(a) $\frac{y(s)}{d(s)}$

(b) $\frac{y(s)}{r(s)}$

(c) $\frac{y(s)}{m(s)}$

(d) $\frac{y(s)}{n(s)}$

(e) $\frac{u(s)}{d(s)}$

(f) $\frac{u(s)}{r(s)}$

(g) $\frac{u(s)}{m(s)}$

(h) $\frac{u(s)}{n(s)}$

(i) $\frac{e(s)}{d(s)}$

(j) $\frac{e(s)}{r(s)}$

(k) $\frac{e(s)}{m(s)}$

(l) $\frac{e(s)}{n(s)}$

3. Figure 9.18 shows a variation of closed-loop control called internal model control (IMC). It has this name because it requires knowing a model of the system to control, as well as an inverse model of the system to control. In that block diagram:

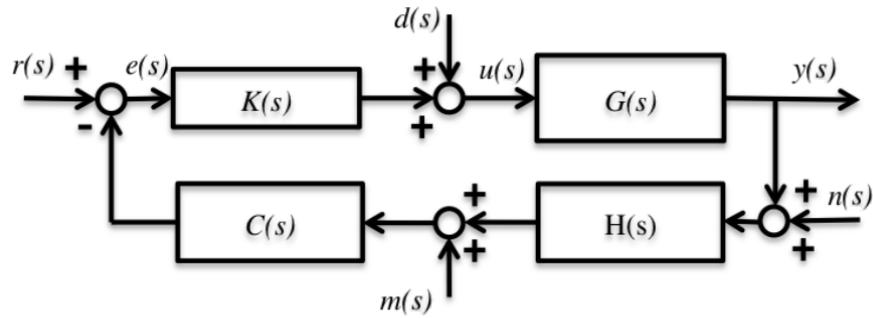


Figure 9.17: Block diagram of Exercise 2.

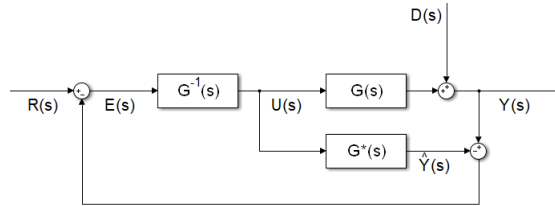


Figure 9.18: Internal model control (IMC).

- $G(s)$ is the plant to control,
- $G^*(s)$ is the model of the plant to control,
- $G^{-1}(s)$ is the inverse model of the plant to control.

- (a) Show that, if the model is perfect, i.e. if $G^*(s) = G(s)$, then the error is given by $E(s) = R(s) - D(s)$.
 - (b) Show that, if, additionally, the inverse model is perfect, i.e. $G^{-1}(s)G(s) = 1$, then the output is $Y(s) = R(s)$.
 - (c) Show that, whether the models are perfect or not, the block diagram of IMC in Figure 9.18 is equivalent to the block diagram of closed-loop control in Figure 9.13, if $C(s) = \frac{G^{-1}(s)}{1 - G^{-1}(s)G^*(s)}$.
4. Figure 9.19 shows a variation of closed-loop control called cascade control (or master-slave control, though that designation is out of favour nowadays). In that block diagram, the plant to control is $G(s) = G_1(s)G_2(s)$, and it possible to measure both $Y_1(s)$ and $Y_2(s)$. Each of the two parts of the system to control is controlled separately.
 - (a) Find transfer function $\frac{Y_1(s)}{U_2(s)}$.
 - (b) Use that result to find transfer function $\frac{Y_2(s)}{R(s)}$.

5. Redraw the block diagram of Figure 9.11 from Example 9.12 as follows:
 - use the values of the variables given in Example 9.12,

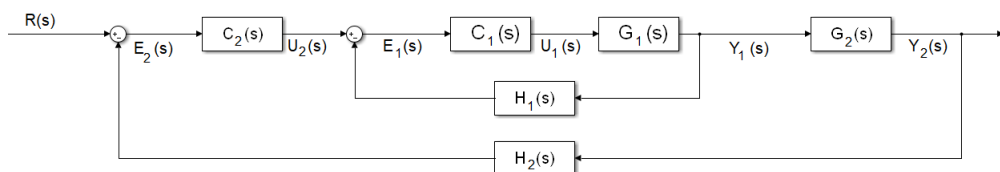


Figure 9.19: Cascade (or master-slave) control.

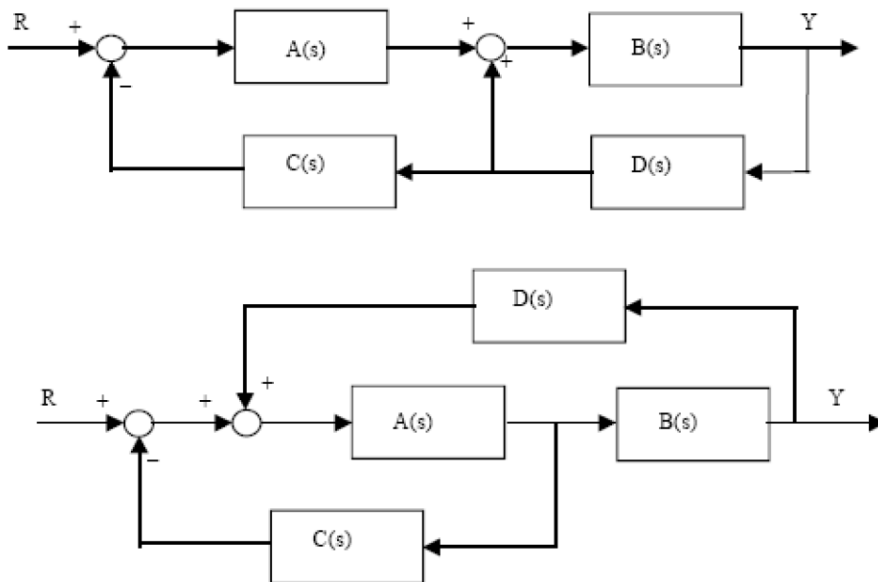


Figure 9.20: Block diagrams of Exercise 6.

- let the input $V_i(s)$ be a manipulated variable,
- let there be some reference $r(t)$ for $x_1(t)$ to follow,
- add proportional control K .

Then find transfer function $\frac{X_1(s)}{R(s)}$ as a function of K .

6. For each of the two block diagrams in Figure 9.20:

- Find transfer function $\frac{Y(s)}{R(s)}$.
- Let $A(s) = \frac{1}{s}$, $B(s) = \frac{10}{s+1}$, $C(s) = 2$, $D(s) = \frac{s+0.1}{s+2}$. Find the value of $\frac{Y(s)}{R(s)}$.

7. Prove all the equivalences of block diagrams shown in Figure 9.9.

Chapter 10

Time and frequency responses

Schirm und Robert fliegen dort
Durch die Wolken immer fort.
Und der Hut fliegt weit voran,
Stößt zuletzt am Himmel an.
Wo der Wind sie hingetragen,
Ja! das weiß kein Mensch zu sagen.

Heinrich HOFFMANN (1809 — †1894), *Der Struwwelpeter*, Die Geschichte vom fliegenden Robert (1847)

We already know that we can use the Laplace transform (and its inverse) to find out the output of any transfer function for any particular input. In this chapter we study several usual particular cases. This allows us to find approximate responses in many cases, and to characterise with simplicity more complex responses. It also paves the way to the important concept of frequency responses.

10.1 Time responses: steps and impulses as inputs

The following inputs are routinely used to test systems:

- The **impulse**:

Impulse

$$u(t) = \delta(t) \quad (10.1)$$

$$\mathcal{L}[u(t)] = 1 \quad (10.2)$$

- The **step**, with amplitude d :

Step

$$u(t) = dH(t) \quad (10.3)$$

$$\mathcal{L}[u(t)] = \frac{d}{s} \quad (10.4)$$

- In particular, the **unit step**, with amplitude 1:

Unit step

$$u(t) = H(t) \quad (10.5)$$

$$\mathcal{L}[u(t)] = \frac{1}{s} \quad (10.6)$$

- The **ramp**, with slope d :

Ramp

$$u(t) = dt \quad (10.7)$$

$$\mathcal{L}[u(t)] = \frac{d}{s^2} \quad (10.8)$$

- In particular, the **unit ramp**, with slope 1:

Unit ramp

$$u(t) = t \quad (10.9)$$

$$\mathcal{L}[u(t)] = \frac{1}{s^2} \quad (10.10)$$

- The **parabola**, with second derivative $2d$:

Parabola

$$u(t) = dt^2 \quad (10.11)$$

$$\mathcal{L}[u(t)] = \frac{2d}{s^3} \quad (10.12)$$

Unit parabola

- In particular, the **unit parabola**, with second derivative 2:

$$u(t) = t^2 \quad (10.13)$$

$$\mathcal{L}[u(t)] = \frac{2}{s^3} \quad (10.14)$$

You can either find the Laplace transforms above in Table 2.1, or calculate them yourself.

Remark 10.1. Notice that:

- the unit step is the integral of the impulse: $\int_0^t \delta(t) dt = H(t)$;
- the unit ramp is the integral of the unit step: $\int_0^t H(t) dt = t$;
- the unit parabola is not the integral of the unit ramp: $\int_0^t t dt = \frac{1}{2}t^2 \neq t^2$. \square

Properties of $\delta(t)$ $\delta(t)$ is not a function

Remark 10.2. Remember that while the Heaviside function $H(t)$ is a function, and so are t and t^2 , the Dirac delta $\delta(t)$ is not. It is a generalised function, and the limit of the following family of functions:

$$f(t, \epsilon) = \begin{cases} \frac{1}{\epsilon}, & \text{if } 0 \leq t \leq \epsilon \\ 0, & \text{if } t < 0 \vee t > \epsilon \end{cases} \quad (10.15)$$

$$\delta(t) = \lim_{\epsilon \rightarrow 0^+} f(t, \epsilon) \quad (10.16)$$

Since

$$\int_{-\infty}^{+\infty} f(t, \epsilon) dt = \int_0^\epsilon f(t, \epsilon) dt = \int_0^\epsilon \frac{1}{\epsilon} dt = 1, \quad \forall \epsilon \in \mathbb{R}^+ \quad (10.17)$$

Its integral in \mathbb{R} is 1

we also have

$$\int_{-\infty}^{+\infty} \delta(t) dt = 1 \quad (10.18)$$

Furthermore, for a continuous function $g(t)$,

$$\begin{aligned} f(t, \epsilon) \min_{0 \leq t \leq \epsilon} g(t) &\leq f(t, \epsilon)g(t) \leq f(t, \epsilon) \max_{0 \leq t \leq \epsilon} g(t) \\ \Rightarrow \int_0^\epsilon f(t, \epsilon) \min_{0 \leq t \leq \epsilon} g(t) dt &\leq \int_0^\epsilon f(t, \epsilon)g(t) dt \leq \int_0^\epsilon f(t, \epsilon) \max_{0 \leq t \leq \epsilon} g(t) dt \\ \Leftrightarrow \min_{0 \leq t \leq \epsilon} g(t) \int_0^\epsilon f(t, \epsilon) dt &\leq \int_0^\epsilon f(t, \epsilon)g(t) dt \leq \max_{0 \leq t \leq \epsilon} g(t) \int_0^\epsilon f(t, \epsilon) dt \\ \Rightarrow \min_{0 \leq t \leq \epsilon} g(t) &\leq \int_0^\epsilon f(t, \epsilon)g(t) dt \leq \max_{0 \leq t \leq \epsilon} g(t) \end{aligned} \quad (10.19)$$

where we used (10.17). Making $\epsilon \rightarrow 0^+$, we get

$$g(0) \leq \int_0^\epsilon f(t, \epsilon)g(t) dt \leq g(0) \Leftrightarrow \int_0^\epsilon f(t, \epsilon)g(t) dt = g(0) \quad (10.20)$$

Integral of $\delta(t)$ multiplied by a function A consequence of this is that

$$\int_a^b g(t)\delta(t) dt = g(0) \quad (10.21)$$

as long as the integration interval includes and overtakes 0, i.e. $a \leq 0 < b$; in particular,

$$\mathcal{L}[\delta(t)] = \int_0^{+\infty} \delta(t)e^{-st} dt = e^{-s0} = 1 \quad \square \quad (10.22)$$

The reasons why (10.1)–(10.13) are routinely used as inputs to test systems are:

- They are simple to create.
- Calculations are simple, given their Laplace transforms.
- They can be used to model many real inputs exactly, and even more as approximations.

Example 10.1. The following situations can be modelled as steps:

- A metal workpiece is taken from an oven and quenched in oil at a lower temperature.
- A sluice gate is suddenly opened, letting water into an irrigation canal.
- A switch is closed and a tension is thereby applied to the motor that rotates the joint of a welding robot.
- A finished part is dropped onto a conveyer belt.
- A car advancing at constant speed descends a sidewalk onto the street pavement. \square

Example 10.2. The following situations can be modelled as ramps:

- A deep space probe moves out of the solar system at constant speed along a straight line in an inertial system of coordinates, due to inertia, far from the gravitational influence of any close celestial body.
- A high-speed train moves from one station to another at cruiser speed.
- A welding robot creates a welding joint at constant speed, to ensure a uniform thickness.

Notice that, save for the first example, the ramp is limited in time: sooner or later, the train and the welding robot will have to stop. In fact, unlimited ramps are seldom found. \square

Remark 10.3. The impulse is in fact impossible to create: there are no physical quantities applied during no time at all, with an infinite intensity. However, the impulse is a good approximation of inputs that have a very short duration. Figure 10.1 shows two inputs in that situation: a sequence of two steps

$$\begin{aligned} u_1(t) &= k H(t - t_0) - k H(t - t_1) \\ &= \begin{cases} k, & \text{if } t_0 \leq t \leq t_1 \\ 0, & \text{if } t < t_0 \vee t > t_1 \end{cases} \approx k(t_1 - t_0)\delta(t - t_0) \end{aligned} \quad (10.23)$$

and, even more realistically, a sequence of two ramps, approximated by

$$\begin{aligned} \delta(t) \int_0^{+\infty} u_2(t) dt &= \delta(t) \left[\frac{1}{2}k(t_1 - t_0) + k(t_2 - t_1) + \frac{1}{2}k(t_3 - t_2) \right] \\ &= \delta(t - t_0) \frac{k}{2}(t_3 + t_2 - t_1 - t_0) \end{aligned} \quad (10.24)$$

Of course, any input with a form such as that of Figure 10.2 can be approximated by an impulse (multiplied by the integral over time of the input). \square

Remark 10.4. Unit steps are almost exclusively used because amplitude 1 makes calculations easier. Since we are assuming linearity, if the amplitude of the step is d instead of 1, the output will be that for the unit step, multiplied by d . The same can be said for unit ramps and unit parabolas. When steps (or ramps, or parabolas) are applied experimentally, amplitude 1 may be too big or too small, and a different one will have to be used instead. \square

Example 10.3. Suppose you want to test a car's suspension, when the wheel climbs or descends a step. Obviously nobody with a sound mind would apply a 1 m step for this purpose (see Figure 10.3). A 10 cm step would for instance be far more reasonable. Of course, if our model is linear, we can apply a unit step, knowing well that the result will be nonsense, and then simply scale down the result. \square

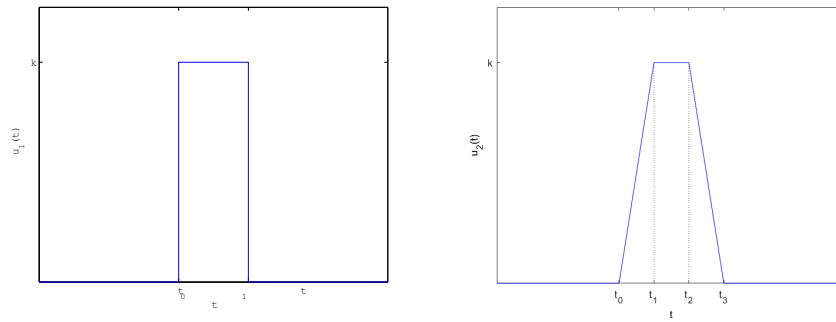


Figure 10.1: Two functions that can be approximated by an impulse if $t_0 \approx t_1$ (left) or $t_0 \approx t_3$ (right).

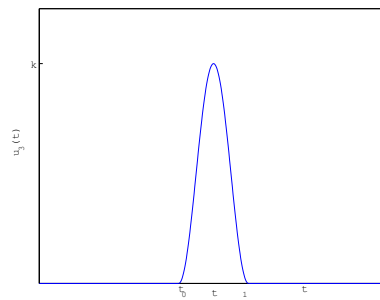


Figure 10.2: General form of a function that can be approximated by an impulse if $t_0 \approx t_1$.

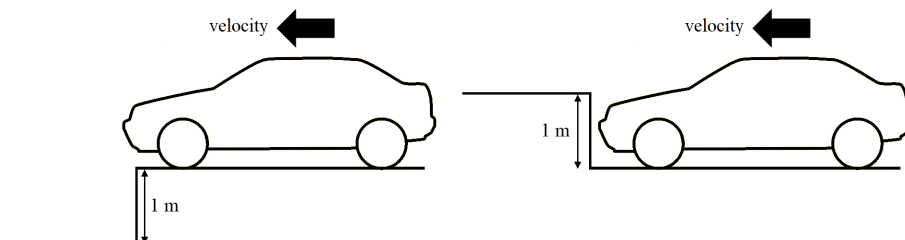


Figure 10.3: Would you test a car's suspension like this? (Source: Wikimedia, modified)



Figure 10.4: The Rasteirinho mobile robot, without the laptop computer with which it is controlled.

Example 10.4. The Rasteirinho (see Figure 10.4) is a mobile robot, of which about a dozen units are used at IST in laboratory classes of different courses. It is controlled by a laptop computer, fixed with velcro. Its maximum speed depends on the particular unit; in most, it is around 80 cm/s. Consequently, it is useless to try to make its position follow a unit ramp, which would correspond to a 1 m/s velocity. Once more, we could simulate its behaviour with a linear model for a unit ramp and then scale the output down. \square

Example 10.5. In the WECs of Figures 3.2 and 3.3, the air inside the device is compressed by the waves. A change of air pressure of 1 Pa is ludicrously small; it is useless even to try to measure it. But if our model of the WEC is linear we can simulate how much energy it produces when a unit step is applied in the air pressure and then scale the result up to a more reasonable value of the pressure variation. \square

In what follows we will concentrate on the impulse and unit step responses, and mention responses to unit ramps and steps with amplitudes which are not 1 whenever appropriate.

Theorem 10.1. The impulse response of a transfer function has a Laplace transform which is the transfer function itself.

Impulse response of a system

Proof. Since $G(s) = \frac{Y(s)}{U(s)}$, where $G(s)$ is a transfer function, $Y(s)$ is the Laplace transform of the output, and $U(s)$ is the Laplace transform of the input, and since the Laplace transform of an impulse is 1, the result is immediate. \square

Remark 10.5. This allows defining a system's transfer function as the Laplace transform of its output when the input is an impulse. This definition is an alternative to Definition 4.1 found in many textbooks. \square

Corollary 10.1. The output of a transfer function $G(s)$ for any input $u(t)$ is equal to the convolution of the input with the transfer function's impulse response $g(t)$:

$$y(t) = g(t) * u(t) = \int_0^t g(t - \tau)u(\tau) d\tau \quad (10.25)$$

Proof. This is an immediate result of Theorem 10.1 and of (2.78). \square

Remark 10.6. It is usually easier to calculate the Laplace transform of the input $U(s)$ to find the Laplace of the output as $Y(s) = G(s)U(s)$ and then finally the output as $y(t) = \mathcal{L}^{-1}[G(s)U(s)]$, than to calculate the output directly as $y(t) = g(t) * u(t)$. \square

The following MATLAB functions are useful to find time responses:

- `step` plots a system's response to a unit step (and can return the values plotted in vectors);
- `impulse` does the same for an impulse input;
- `lsim`, already studied in Section 4.2, can be used for any input.

Just like `lsim`, both `step` and `impulse` use numerical methods to find the responses, rather than analytical computations.

MATLAB's *command* `impulse` **Example 10.6.** The impulse, unit step and unit ramp responses of a plant are shown in Figure 10.5 and obtained as follows:

MATLAB's *command* `step`

```
>> s = tf('s'); G = 1/(s+1);
>> figure, impulse(G), figure, step(G)
>> t = 0 : 0.01 : 6; figure, plot(t, lsim(G, t, t))
>> xlabel('t [s]'), ylabel('output')
>> title('response to a unit ramp')
```

The time range is chosen automatically by `step` and `impulse`. \square

Example 10.7. The response of the transfer function from Example 10.6 to a step with amplitude 10 during 20 s can be found in two different manners, both providing, of course, the same result:

```
>> [stepresp, timevector] = step(G, 20);
>> t = 0 : 0.01 : 20;
>> figure, plot(t, lsim(G, 10*ones(size(t)), t), timevector, 10*stepresp)
>> xlabel('t [s]'), ylabel('output'), title('Step response')
```

There is, in fact, a slight difference in the two plots shown in Figure 10.6, because function `step` chooses the sampling time automatically, and it is different from the one explicitly fed to `lsim`. \square

10.2 Steady-state response and transient response

The impulse, unit step, and unit ramp responses of

$$G(s) = \frac{1}{s+1} \quad (10.26)$$

from Example 10.6, shown in Figure 10.5 as they are numerically calculated by Matlab, can be found analytically as follows:

- Impulse response:

$$y_i(t) = \mathcal{L}^{-1} \left[\frac{1}{s+1} \right] = e^{-t} \quad (10.27)$$

- Unit step response:

$$y_s(t) = \mathcal{L}^{-1} \left[\frac{1}{s+1} \frac{1}{s} \right] = 1 - e^{-t} \quad (10.28)$$

- Unit ramp response:

$$y_r(t) = \mathcal{L}^{-1} \left[\frac{1}{s+1} \frac{1}{s^2} \right] = \mathcal{L}^{-1} \left[-\frac{1}{s} + \frac{1}{s^2} + \frac{1}{s+1} \right] = t - 1 + e^{-t} \quad (10.29)$$

In each of them we can separate the terms that tend to zero as the time increases from those that do not. The first make up what we call the **transient response**. The latter make up what we call the **steady-state response**.

$$y_i(t) = \underbrace{0}_{\text{steady-state}} + \underbrace{e^{-t}}_{\text{transient}} \quad (10.30)$$

$$y_s(t) = \underbrace{1}_{\text{steady-state}} - \underbrace{e^{-t}}_{\text{transient}} \quad (10.31)$$

$$y_r(t) = \underbrace{t-1}_{\text{steady-state}} + \underbrace{e^{-t}}_{\text{transient}} \quad (10.32)$$

Transient
Steady-state

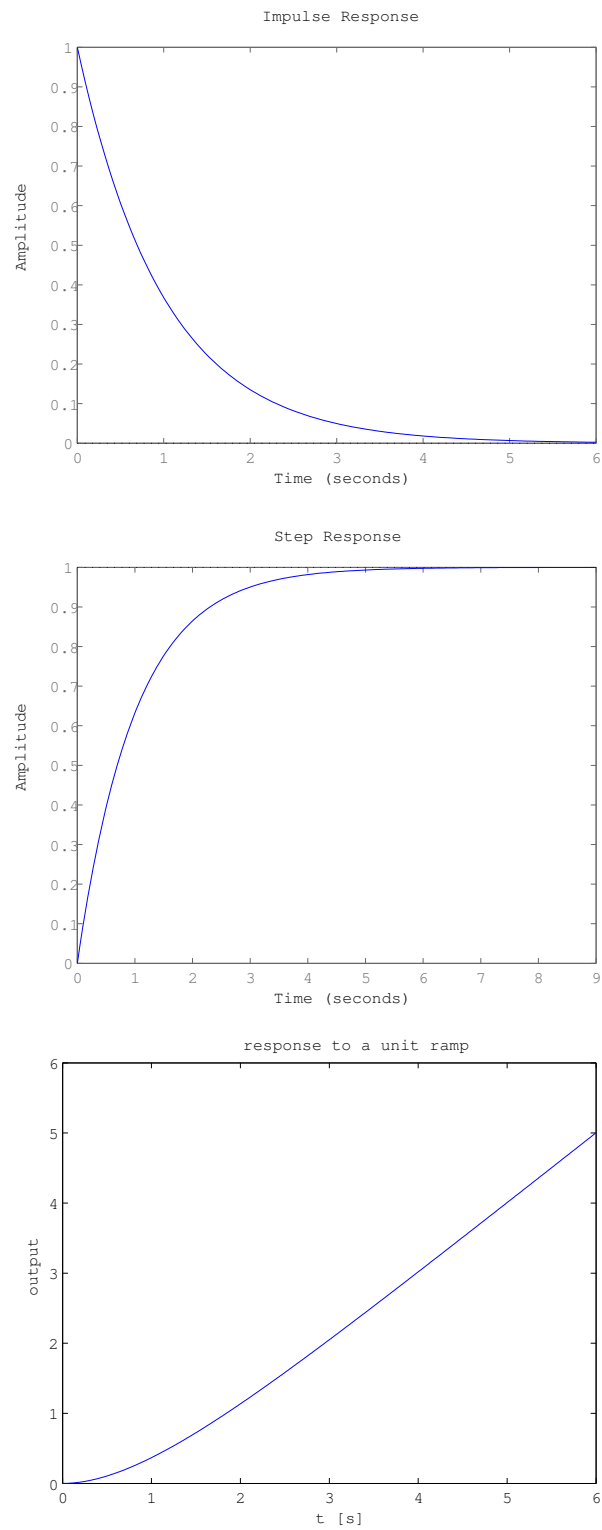


Figure 10.5: Impulse, unit step and unit ramp responses of $G(s) = \frac{1}{s+1}$, from Example 10.6.

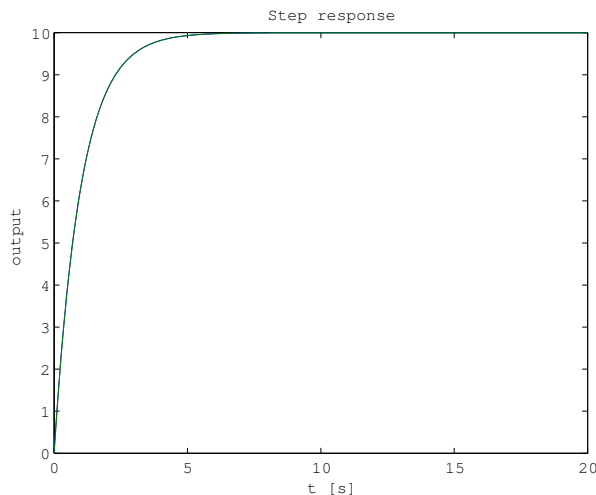


Figure 10.6: Response of $G(s) = \frac{1}{s+1}$ for a step with amplitude 10, from Example 10.7.

In other words, a time response $y(t)$ can be separated into two parts, the transient response $y_t(t)$ and the steady-state response $y_{ss}(t)$, such that

$$y(t) = y_t(t) + y_{ss}(t) \quad (10.33)$$

$$\lim_{t \rightarrow +\infty} y_t(t) = 0 \quad (10.34)$$

$$\lim_{t \rightarrow +\infty} y_{ss}(t) \neq 0 \vee y_{ss}(t) = 0, \forall t \quad (10.35)$$

We also call transient to the period of time in which the response is dominated by the transient response, and steady-state to the period of time in which the transient response is neglectable and the response can be assumed equal to the steady-state response. Whether a transient response can or cannot be neglected depends on how precise our knowledge of the response has to be. Below in Sections 11.2 and 11.3 we will see usual criteria for this.

The steady-state response can be:

- zero, as the impulse response of (10.26), shown in Figure 10.5;
- a non-null constant, as the unit step response of (10.26), shown in Figure 10.5;
- an oscillation with constant amplitude, as the step response of $\frac{1}{(s^2+1)(s+1)}$, shown in Figure 10.7;
- infinity, with the output increasing or decreasing monotonously, as the unit ramp response of (10.26), shown in Figure 10.5;
- infinity, with the output oscillating with increasing amplitude, as the impulse response of $\frac{s}{(s^2+1)^2}$, shown in Figure 10.7.

What the steady-state response is depends on what the system is and on what its input is.

Remark 10.7. Most systems never reach infinity. The probe of Example 10.2 can move away to outer space, but temperatures do not rise to infinite values (before that the heat source is exhausted, or something will burn), robots reach the end of their workspace, high electrical currents will activate a circuit breaker, etc.; in other words, for big values of the variables involved, the linear model of the system usually ceases in one way or another to be valid. \square

Over the next sections we will learn several ways to calculate steady-state responses without having to find an explicit expression for the output, and then calculating its limit. When the steady-state response is constant or infinity, it can be found from the final value theorem (Theorem 2.4), i.e. applying (2.72).

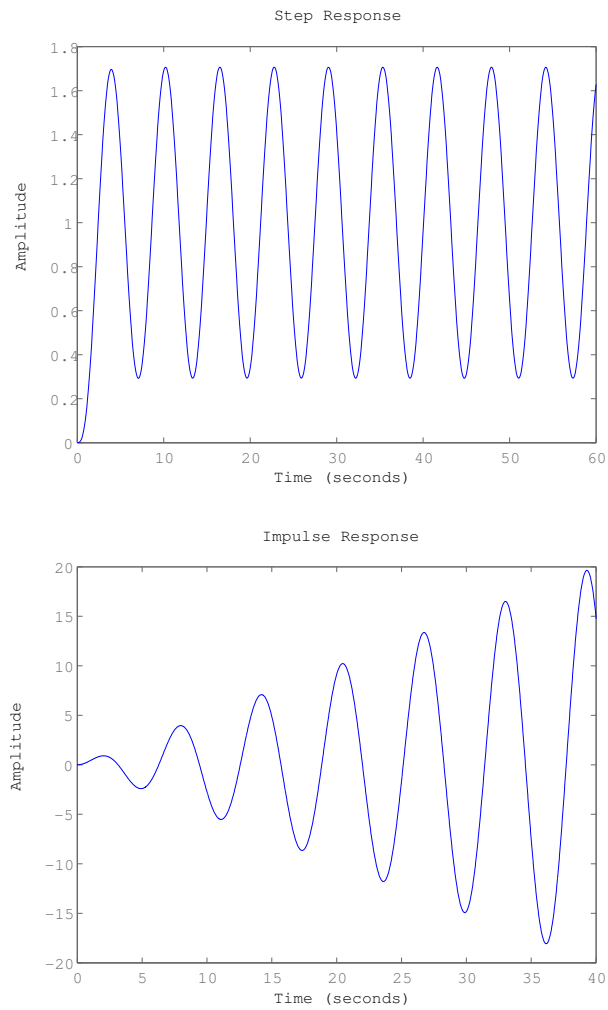


Figure 10.7: Time responses with oscillations: unit step response of $\frac{1}{(s^2+1)(s+1)}$ (top) and impulse response of $\frac{s}{(s^2+1)^2}$ (bottom).

Example 10.8. The steady-states of the impulse, step and ramp responses (10.27)–(10.29) are as follows:

$$\lim_{t \rightarrow +\infty} y_i(t) = \lim_{t \rightarrow +\infty} e^{-t} = 0 \quad (10.36)$$

$$\lim_{t \rightarrow +\infty} y_s(t) = \lim_{t \rightarrow +\infty} 1 - e^{-t} = 1 \quad (10.37)$$

$$\lim_{t \rightarrow +\infty} y_r(t) = \lim_{t \rightarrow +\infty} t - 1 + e^{-t} = +\infty \quad (10.38)$$

They can be found without the inverse Laplace transform using (2.72):

$$\lim_{t \rightarrow +\infty} y_i(t) = \lim_{s \rightarrow 0} s \frac{1}{s+1} = 0 \quad (10.39)$$

$$\lim_{t \rightarrow +\infty} y_s(t) = \lim_{s \rightarrow 0} s \frac{1}{s+1} \frac{1}{s} = 1 \quad (10.40)$$

$$\lim_{t \rightarrow +\infty} y_r(t) = \lim_{s \rightarrow 0} s \frac{1}{s+1} \frac{1}{s^2} = +\infty \quad \square \quad (10.41)$$

Example 10.9. Remember that (2.72) applies when the limit in time exists. Figure 10.7 shows two cases where this limit clearly does not exist because of oscillations with an amplitude that does not decrease. But the two corresponding limits are

$$\lim_{t \rightarrow +\infty} y(t) = \lim_{s \rightarrow 0} s \frac{1}{(s^2+1)(s+1)} \frac{1}{s} = 1 \quad (10.42)$$

$$\lim_{t \rightarrow +\infty} y(t) = \lim_{s \rightarrow 0} s \frac{s}{(s^2+1)^2} = \infty \quad (10.43)$$

In the first case we got the average value of the steady-state response; in the second, infinity. Neither case is a valid application of the final value theorem. We need to know first if the time limit exists. \square

Static gain

Definition 10.1. The constant steady-state output of the unit step response of a stable system $G(s) = \frac{Y(s)}{U(s)}$ is called the **static gain** of $G(s)$:

$$\lim_{t \rightarrow +\infty} y(t) = \lim_{s \rightarrow 0} s \underbrace{\frac{b_0 + b_1s + b_2s^2 + b_3s^3 + \dots}{a_0 + a_1s + a_2s^2 + a_3s^3 + \dots}}_{G(s)} \underbrace{\frac{1}{s}}_{U(s)} = \frac{b_0}{a_0} \quad \square \quad (10.44)$$

10.3 Stability

The time responses from Section 10.2 illustrate the importance of the concept of stability.

Bounded signal

Definition 10.2. A signal $x(t)$ is **bounded** if $\exists K \in \mathbb{R}^+ : \forall t, |x(t)| < K$. \square

BIBO stability

Definition 10.3. A system is:

- **stable** if, for every input which is bounded, its output is bounded too;
- **not stable** if there is at least a bounded input for which its output is not bounded.

This definition of **stability** is known as bounded input, bounded output stability (BIBO stability). \square

All poles of stable transfer functions are on the left complex half-plane

Theorem 10.2. A transfer function is stable if and only if all its poles are on the left complex half-plane.

Proof. We will prove this in two steps:

- A transfer function $G(s)$ is stable if and only if its impulse response $g(t)$ is absolutely integrable, i.e. iff $\exists M \in \mathbb{R}^+$

$$\int_0^{+\infty} |g(t)| dt < M \quad (10.45)$$

- A transfer function's impulse response is absolutely integrable if and only if all its poles are on the left complex half-plane. \square

Lemma 10.1. A transfer function is stable if and only if its impulse response is absolutely integrable.

Proof. Let us suppose that the impulse response $g(t)$ is absolutely integrable, and that

$$\int_0^{+\infty} |g(\tau)| d\tau = K \quad (10.46)$$

Let us also suppose that the input $u(t)$ is bounded, as required by the definition of BIBO stability:

$$|u(t)| \leq U, \quad \forall t \quad (10.47)$$

From (10.25) we get

$$\begin{aligned} |y(t)| &= |g(t) * u(t)| = \left| \int_0^t g(\tau) u(t - \tau) d\tau \right| \\ &\leq \int_0^t |g(\tau) u(t - \tau)| d\tau \\ &\leq \int_0^t |g(\tau)| |u(t - \tau)| d\tau \\ &\leq U \int_0^t |g(\tau)| d\tau \leq UK \end{aligned} \quad (10.48)$$

So the output is bounded, proving that the condition (impulse response absolutely integrable) is sufficient.

Reductio ad absurdum proves that it is also necessary. Suppose that the impulse response $g(t)$ is not absolutely integrable; thus, there is a time instant $T \in \mathbb{R}^+$ such that

$$\int_0^T |g(\tau)| d\tau = +\infty \quad (10.49)$$

Now let the input $u(t)$ be given by

$$u(T - t) = \text{sign}(g(t)) \quad (10.50)$$

This is a bounded input, $-1 \leq u(t) \leq 1$, $\forall t$, and so, if the transfer function were stable, the output would have to be bounded. But in time instant T

$$y(T) = \int_0^T g(\tau) u(T - \tau) d\tau = \int_0^T |g(\tau)| d\tau = +\infty \quad (10.51)$$

and thus $y(t)$ is not bounded. This shows that the condition is not only sufficient but also necessary. \square

Lemma 10.2. A transfer function's impulse response is absolutely integrable if and only if all its poles are on the left complex half-plane.

Proof. A transfer function $G(s)$ has an impulse response given by $\mathcal{L}^{-1}[G(s)]$. Transfer function $G(s)$ can be expanded into a partial fraction expansion, where the fractions have the poles of $G(s)$ in the denominator. Poles can be divided into four cases.

- The pole is real, $p \in \mathbb{R}$, and simple. In this case the fraction $\frac{k}{s-p}$ (where $k \in \mathbb{R}$ is some real numerator) has the inverse Laplace transform $k e^{pt}$.
 - If $p = 0$, then $\lim_{t \rightarrow +\infty} k e^{pt} = k$. In this case the impulse response is not absolutely integrable, since

$$\int_0^{+\infty} |k| dt = \lim_{t \rightarrow +\infty} |k|t = +\infty \quad (10.52)$$

- If $p > 0$, the exponential tends to infinity: $\lim_{t \rightarrow +\infty} k e^{pt} = \pm\infty$ (depending on the sign of k). If in the last case the response was not absolutely integrable, even more so in this one.

- If $p < 0$, the exponential tends to zero: $\lim_{t \rightarrow +\infty} k e^{pt} = 0$. The impulse response is absolutely integrable, since

$$\int_0^{+\infty} |k e^{pt}| dt = k \int_0^{+\infty} e^{pt} dt = k \left[\frac{1}{p} e^{pt} \right]_0^{+\infty} = \frac{k}{p} (0 - 1) = -\frac{k}{p} \in \mathbb{R}^+ \quad (10.53)$$

- The pole is real and its multiplicity n is 2 or higher. In this case there will be, in the expansion, fractions of the form $\frac{k_n}{(s-p)^n}, \frac{k_{n-1}}{(s-p)^{n-1}}, \frac{k_{n-2}}{(s-p)^{n-2}} \dots \frac{k_1}{s-p}$. (Here the $k_i \in \mathbb{R}$, $i = 1 \dots n$ are the numerators in the expansion.) The corresponding inverse Laplace transforms are of the form $\frac{k_i}{(i-1)!} t^{i-1} e^{pt}$, $i = 1 \dots n$.

- If $p = 0$, then the exponential tends to 1, but the power does diverge to infinity: $\lim_{t \rightarrow +\infty} \frac{k_i}{(i-1)!} t^{i-1} e^{pt} = \pm\infty$ (depending on the sign of k), $\forall i \geq 2$. So in this case the impulse response is not absolutely integrable, as seen above.
- If $p > 0$, then $\lim_{t \rightarrow +\infty} \frac{k_i}{(i-1)!} t^{i-1} e^{pt} = \pm\infty$, $\forall i$. Again, the impulse response is not absolutely integrable.
- If $p < 0$, then $\lim_{t \rightarrow +\infty} \frac{k_i}{(i-1)!} t^{i-1} e^{pt} = 0$, $\forall i$, since the effect of the exponential prevails. For the same reason, the impulse response is absolute integrable, just as in (10.53).

- The pole is complex, $p = a + bj \in \mathbb{C} \setminus \mathbb{R}$, $a, b \in \mathbb{R}$, and simple. Remember once more that complex poles must appear in pairs of complex conjugates, since all polynomial coefficients are real (otherwise real inputs would case complex outputs). In this case the fraction $\frac{k}{s-p} = \frac{k}{s-(a+bj)}$ (where $k \in \mathbb{C}$ is some complex numerator) has the inverse Laplace transform

$$\begin{aligned} \mathcal{L}^{-1} \left[\frac{k}{s-p} \right] &= \mathcal{L}^{-1} \left[\frac{k}{s-(a+bj)} \right] = k e^{pt} = k e^{at} e^{bjt} = k e^{at} e^{bjt} \\ &= k e^{at} (\cos bt + j \sin bt) \end{aligned} \quad (10.54)$$

and the fraction $\frac{\bar{k}}{s-\bar{p}} = \frac{\bar{k}}{s-(a-bj)}$ (where \bar{z} is the complex conjugate of $z \in \mathbb{C}$) has the inverse Laplace transform

$$\begin{aligned} \mathcal{L}^{-1} \left[\frac{\bar{k}}{s-\bar{p}} \right] &= \mathcal{L}^{-1} \left[\frac{\bar{k}}{s-(a-bj)} \right] = \bar{k} e^{\bar{p}t} = \bar{k} e^{at} e^{-bjt} = \bar{k} e^{at} e^{-bjt} \\ &= \bar{k} e^{at} (\cos(-bt) + j \sin(-bt)) = \bar{k} e^{at} (\cos bt - j \sin bt) \end{aligned} \quad (10.55)$$

Their effect on the impulse response is their sum:

$$\begin{aligned} \mathcal{L}^{-1} \left[\frac{k}{s-p} \right] + \mathcal{L}^{-1} \left[\frac{\bar{k}}{s-\bar{p}} \right] &= k e^{at} (\cos bt + j \sin bt) + \bar{k} e^{at} (\cos bt - j \sin bt) \\ &= (k + \bar{k}) e^{at} \cos bt = 2\Re(k) e^{at} \cos bt \end{aligned} \quad (10.56)$$

Notice that the imaginary parts cancel out, and we are left with oscillations having:

- period $\frac{2\pi}{b}$, where b is the positive imaginary part of the poles;
- amplitude $2\Re(k)e^{at}$, where a is the real part of the poles. The exponential is the important term, since it is the exponential that may cause this term to vanish or diverge.

So:

- If $a = 0$, then the amplitude of the oscillations remains constant; they do not go to zero neither do they diverge to an infinite amplitude.

This means that the impulse response is not absolutely integrable, since

$$\begin{aligned}
 \int_0^{+\infty} |2\Re(k) \cos bt| dt &= 2|\Re(k)| \int_0^{+\infty} |\cos bt| dt \\
 &= 2|\Re(k)| \lim_{n \rightarrow +\infty} n \int_0^{\frac{2\pi}{b}} |\cos bt| dt \\
 &= 4|\Re(k)| \lim_{n \rightarrow +\infty} n \int_0^{\frac{\pi}{b}} \sin bt dt \\
 &= \frac{4|\Re(k)|}{b} \lim_{n \rightarrow +\infty} n [-\cos bt]_0^{\frac{\pi}{b}} \\
 &= \frac{8|\Re(k)|}{b} \lim_{n \rightarrow +\infty} n = +\infty \quad (10.57)
 \end{aligned}$$

- If $a > 0$, the amplitude of the oscillations tends to infinity. Consequently the impulse response will not be absolutely integrable.
- If $a < 0$, the exponential tends to zero, and so will the oscillations. In this case the impulse response is absolutely integrable, since

$$\int_0^{+\infty} |2\Re(k) e^{at} \cos bt| dt \leq 2|\Re(k)| \int_0^{+\infty} e^{at} dt \quad (10.58)$$

and we end up with a case similar to (10.53).

- The pole is complex and its multiplicity n is 2 or higher. This case is a mixture of the last two. There will be terms of the form $\frac{k_i}{(s-(a+bj))^i} + \frac{\bar{k}_i}{(s-(a-bj))^i}$, $i = 1 \dots n$. The corresponding inverse Laplace transform is $\frac{2\Re(k_i)}{(i-1)!} t^{i-1} e^{at} \cos bt$. So:
 - If $a = 0$, then $e^{at} = 1$ but the amplitude of the oscillations still grows to infinity, because of the power function, if $i \geq 2$. So in this case the impulse response will not be absolutely integrable.
 - If $a > 0$, the amplitude of the oscillations tends to infinity. The same conclusion follows.
 - If $a < 0$, the exponential tends to zero, and for large times its effect prevails; so the the impulse response will be absolutely integrable.

It is clear that one single term not tending exponentially to zero suffices to prevent the impulse response from being absolutely integrable. Consequently, the only way for the impulse response to tend to zero is that all poles should have negative real parts; in other words, that all poles should lie on the left complex half-plane. \square

While some authors call unstable to all systems that are not stable, the following distinction is current.

Definition 10.4. A system is:

- **unstable** if its impulse response is not bounded;
- **marginally stable** if it is not stable and its impulse response is bounded. \square

Unstable and marginally stable systems

Unstable systems

Marginal stability

Theorem 10.3. Marginally stable systems have no poles on the right complex half-plane, and one or more simple poles on the imaginary axis.

Marginally stable systems have simple poles on the imaginary axis

Proof. It is clear from the proof of Lemma 10.2 that simple poles on the imaginary axis correspond to:

- impulse responses which are bounded:
 - a pole at the origin has a constant impulse response;
 - a pair of complex conjugate imaginary poles has constant amplitude sinusoidal oscillations as impulse response;

- responses to bounded inputs which are not bounded, since systems with such poles are not stable.

A single pole p on the right complex half-plane makes a system unstable, since, whatever the input may be, in the partial fraction expansion of the output there will be a fraction of the form $\frac{k}{s-p}$, and the proof of Lemma 10.2 shows that such terms always diverge exponentially to infinity.

The same happens with multiple poles on the imaginary axis:

- multiple poles at the origin cause a polynomial impulse response, which diverges to infinity (check Table 2.1, or (11.206) below);
- pure imaginary multiple poles also cause the impulse response to diverge, as argued in the proof of Lemma 10.2.

□

The effect of each pole on the stability of a system justifies the following nomenclature.

Stable, marginally stable, and unstable poles **Definition 10.5.** Poles are:

- **stable**, when located on the left complex half-plane;
- **marginally stable**, when simple and located on the imaginary axis, i.e. $s = j\omega$, $\omega \in \mathbb{R}$, and notice that this includes the origin, $s = 0$;
- **unstable**, when multiple and located on the imaginary axis, or when located on the right complex half-plane. □

Stability depends on pole location

A system is:

- **stable**, when all its poles are stable;
- **marginally stable**, when it has no unstable poles, and one or more of its poles are marginally stable;
- **unstable**, when it has one or more unstable poles.

Example 10.10. From the location of the poles, we can conclude the following about the stability of these transfer functions:

- $\frac{s+4}{(s+1)(s+2)(s+3)}$; poles: $-1, -2, -3$; stable transfer function
- $\frac{s-5}{s^2+6}$; poles: $\pm\sqrt{6}j$; marginally stable transfer function
- $\frac{s+7}{(s^2+8)^2}$; poles: $\pm\sqrt{8}j$ (double); unstable transfer function
- $\frac{(s-12)(s+13)}{(s+9)^2(s^2+20s+221)}$; poles: -9 (double), $-10 \pm 11j$; stable transfer function
- $\frac{14}{s-15}$; poles: 15 ; unstable transfer function
- $\frac{16}{s}$; poles: 0 ; marginally stable transfer function
- $\frac{-17}{s^2}$; poles: 0 (double); unstable transfer function
- $\frac{18}{s(s^2+19)}$; poles: $0, \pm\sqrt{19}j$; marginally stable transfer function
- $\frac{24}{(s+20)(s+21)(s+22)(s-23)}$; poles: $-20, -21, -22, 23$; unstable transfer function □

Poles, not zeros, determine stability **Remark 10.8.** Never forget that zeros have nothing to do with stability. □

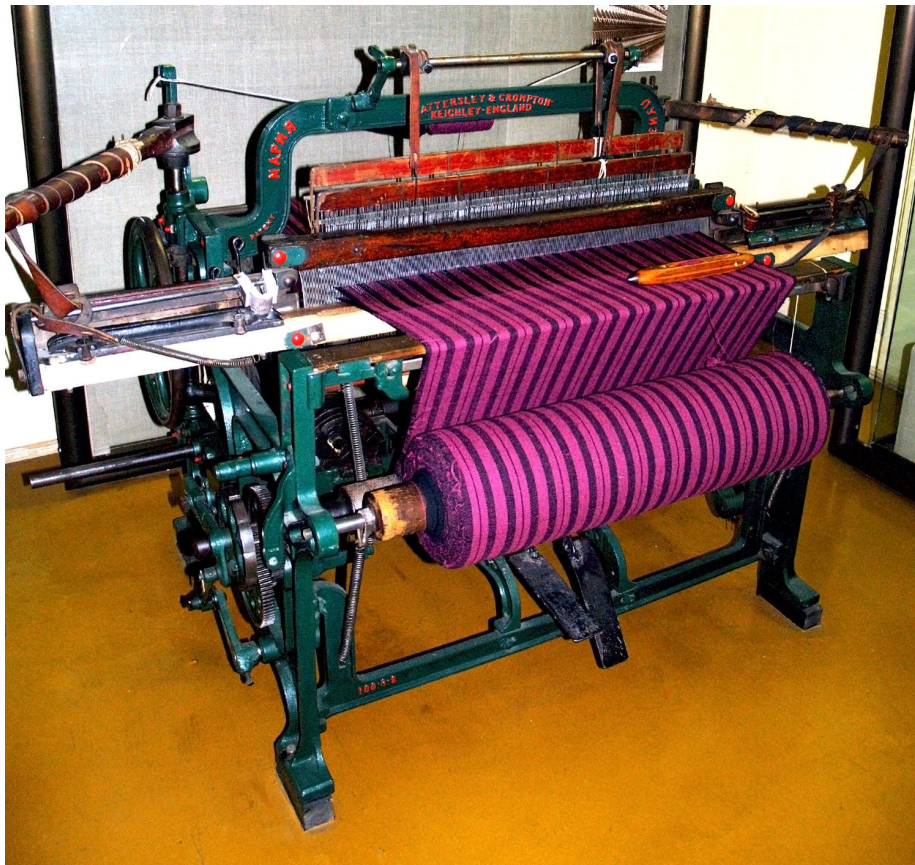


Figure 10.8: A weaving loom (source:Wikimedia).

10.4 Time responses: periodic inputs

Consider the weaving loom in Figure 10.8. The shuttle that carries the yarn that will become the weft thread moves without cease from the left to the right and then back. Meanwhile, half the warp threads are pulled up by a harness, which will then lower them while the other half goes up, and this too without cease. The corresponding references are similar to those in Figure 10.9. They are called **square wave** and **triangle wave**, and are examples of **periodic signals**.

Square wave
Triangle wave
Periodic signals

Definition 10.6. A **periodic signal** is one for which $\exists \mathfrak{T} \in \mathbb{R}^+$

$$f(t + \mathfrak{T}) = f(t), \quad \forall t \quad (10.59)$$

$T = \min \mathfrak{T}$ is the **period** of signal $f(t)$. □

Remark 10.9. Notice that the different values of \mathfrak{T} are in fact the integer multiples of T , i.e.

$$f(t + T) = f(t), \quad \forall t \Rightarrow f(t + nT) = f(t), \quad \forall t, n \in \mathbb{N} \quad \square \quad (10.60)$$

Triangle waves are also a useful alternative to ramps, since they avoid the inconvenience of an infinitely large signal. Square waves are useful in experimental settings for another reason: they allow seeing successive step responses, and consequently allow measuring parameters several times in a row. For this purpose, the period must be large enough for the transient regime to disappear.

Example 10.11. We can find the output of $G(s) = \frac{15}{s+20}$ to a square wave with period 1 s and amplitude 1 using MATLAB as follows:

MATLAB's *square* *command*

```
>> t = 0 : 0.001 : 3;
>> u = square(t*2*pi);
>> figure, plot(t,u, t,lsim(15/(s+20),u,t))
>> axis([0 3 -1.5 1.5])
>> xlabel('t [s]'), ylabel('input and output'), legend({'input','output'})
```

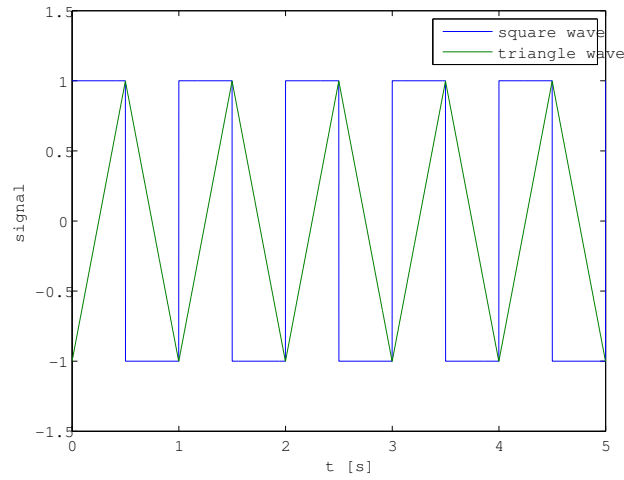


Figure 10.9: A square wave and a triangle wave (both with period 1 and amplitude 1).

Notice that the amplitude of the first step is 1 and the amplitude of the following steps is the peak to peak amplitude, twice as big, viz. 2. Also notice that there is a step every half period, i.e. every 0.5 s.

The period was appropriately chosen since (as we shall see in Section 11.2) the transient response is practically gone after 0.5 s. A period four times smaller would not allow seeing a complete step response. Both cases are shown in Figure 10.10. \square

Another useful periodic signal is the sinusoid, which appears naturally with any phenomena that are the projection onto a plane of a circular movement on a perpendicular plane. In practice, sinusoids are found (at least as approximations) when working with such different things as tides, motor vibrations, or daily thermal variations.

Sinusoidal inputs cause sinusoidal outputs in steady state

Theorem 10.4. The stationary response $y(t)$ of a stable linear plant $G(s)$ subject to a sinusoidal input $u(t) = \sin(\omega t)$ is

$$y(t) = |G(j\omega)| \sin(\omega t + \angle G(j\omega)) \quad (10.61)$$

where $\angle z$ is the phase, or argument, of $z \in \mathbb{C}$ (also notated often as $\arg z$), so that $z = |z|e^{j\angle z}$.

Proof. The output is

$$y(t) = \mathcal{L}^{-1} [Y(s)] \quad (10.62)$$

and

$$Y(s) = G(s)U(s) = G(s)\mathcal{L}[\sin(\omega t)] = G(s)\frac{\omega}{s^2 + \omega^2} = G(s)\frac{\omega}{(s + j\omega)(s - j\omega)} \quad (10.63)$$

If all poles p_k , $k = 1, \dots, n$ of $G(s)$ are simple, we can perform a partial fraction expansion of $Y(s)$ as follows:

$$\begin{aligned} Y(s) &= \frac{b_0}{s + j\omega} + \frac{\bar{b}_0}{s - j\omega} + \sum_{k=1}^n \frac{b_k}{s - p_k} \\ \Rightarrow y(t) &= \underbrace{b_0 e^{-j\omega t} + \bar{b}_0 e^{j\omega t}}_{\text{steady-state response } y_{ss}(t)} + \underbrace{\sum_{k=1}^n b_k e^{p_k t}}_{\substack{\text{transient} \\ \text{response } y_t(t)}} \end{aligned} \quad (10.64)$$

We know that all terms in the transient response $y_t(t)$ belong there because the exponentials are vanishing, since the poles are on the left complex half-plane.

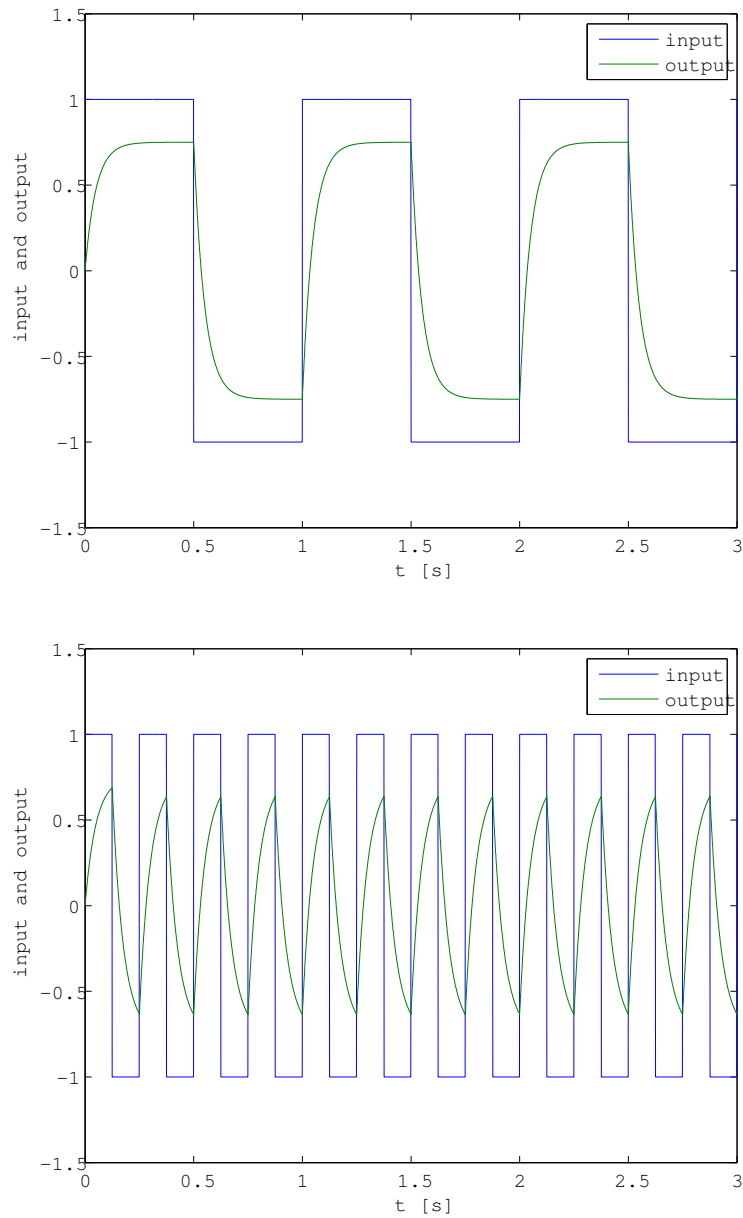


Figure 10.10: Response of $G(s) = \frac{15}{s + 20}$ to two square waves with different periods.

If there are multiple poles, the only difference is that there will be terms of the form $\frac{b_k}{(i-1)!} t^{i-1} e^{p_k t}$, $i \in \mathbb{N}$ in the transient response $y_t(t)$, which will still, of course, be vanishing with time. In either case, the steady-state response is the same.

From (10.63) we know that $Y(s) = G(s) \frac{\omega}{(s+j\omega)(s-j\omega)}$, and from (10.64) we know that $Y(s) = \frac{b_0}{s+j\omega} + \frac{\bar{b}_0}{s-j\omega} + \mathcal{L}[y_t(t)]$. We can multiply both by $s+j\omega$ and obtain

$$G(s) \frac{\omega}{s-j\omega} = b_0 + \left(\frac{\bar{b}_0}{s-j\omega} + \mathcal{L}[y_t(t)] \right) (s+j\omega) \quad (10.65)$$

Now we evaluate this equality at $s = -j\omega$:

$$G(-j\omega) \frac{\omega}{-2j\omega} = b_0 \quad (10.66)$$

Replacing $b_0 = G(-j\omega) \frac{1}{-2j}$ and $\bar{b}_0 = G(j\omega) \frac{1}{2j}$ in $y_{ss}(t) = b_0 e^{-j\omega t} + \bar{b}_0 e^{j\omega t}$, we obtain

$$\begin{aligned} y_{ss}(t) &= G(-j\omega) \frac{1}{-2j} e^{-j\omega t} + G(j\omega) \frac{1}{2j} e^{j\omega t} \\ &= |G(-j\omega)| e^{j\angle G(-j\omega)} \frac{1}{-2j} e^{-j\omega t} + |G(j\omega)| e^{j\angle G(j\omega)} \frac{1}{2j} e^{j\omega t} \\ &= -|G(j\omega)| e^{j(\angle G(-j\omega) - \omega t)} \frac{1}{2j} + |G(j\omega)| e^{j(\angle G(j\omega) + \omega t)} \frac{1}{2j} \\ &= \frac{1}{2j} |G(j\omega)| \left(e^{j(\angle G(j\omega) + \omega t)} - e^{j(\angle G(-j\omega) - \omega t)} \right) \\ &= \frac{1}{2j} |G(j\omega)| \left(\cos(\angle G(j\omega) + \omega t) + j \sin(\angle G(j\omega) + \omega t) \right. \\ &\quad \left. - \cos(-(\angle G(j\omega) + \omega t)) - j \sin(-(\angle G(j\omega) + \omega t)) \right) \\ &= \frac{1}{2j} |G(j\omega)| \left(\cos(\angle G(j\omega) + \omega t) + j \sin(\angle G(j\omega) + \omega t) \right. \\ &\quad \left. - \cos(\angle G(j\omega) + \omega t) + j \sin(\angle G(j\omega) + \omega t) \right) \\ &= \frac{1}{2j} |G(j\omega)| 2j \sin(\angle G(j\omega) + \omega t) \\ &= |G(j\omega)| \sin(\omega t + \angle G(j\omega)) \square \end{aligned} \quad (10.67)$$

Corollary 10.2. Since $G(s)$ is not only stable but also linear, if the input is $u(t) = A \sin(\omega t)$ instead, the output is

$$y(t) = A |G(j\omega)| \sin(\omega t + \angle G(j\omega)) \quad (10.68)$$

Example 10.12. Figure 10.11 shows the simulated vertical position of the Wave Energy Converter of Figure 3.2, the Archimedes Wave Swing, when subject to sinusoidal waves of different amplitudes. The device is in steady-state, as is clear both from the regularity of its movements and from the time already passed since the beginning of the simulation. As the input is sinusoidal, if the model were linear, the output should be sinusoidal too. But the shape of the output is not sinusoidal; it is not even symmetrical around its mean value; its amplitude does not increase linearly with the amplitude of the input. The model used to obtain these simulation results is obviously non-linear. \square

10.5 Frequency responses and the Bode diagram

(10.68) shows that, if a stable system $G(s)$ has a sinusoidal input, the steady-state output is related to the input through $G(j\omega)$, which is the Fourier transform (2.87) of the differential equation describing the system's dynamics:

Frequency, amplitude, and phase of output for sinusoidal inputs

- if the input is sinusoidal, the steady-state output is sinusoidal too;
- if the input has frequency ω , the steady-state output has frequency ω too;

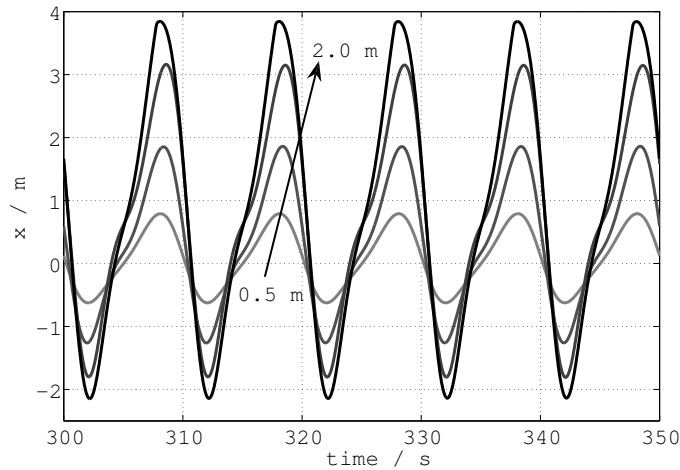


Figure 10.11: Vertical position of the AWS from Figure 3.2, simulated assuming sinusoidal sea waves of different amplitudes between 0.5 m and 2.0 m.

- if the input has amplitude A (or peak-to-peak amplitude $2A$), the steady-state output has amplitude $A|G(j\omega)|$ (or peak-to-peak amplitude $2A|G(j\omega)|$);
- if the input has phase θ at some time instant t , the steady-state output has phase $\theta + \angle G(j\omega)$ at that time instant t .

Remember that:

- the steady-state output is sinusoidal, but the transient is not: you must wait for the transient to go away to have a sinusoidal output; *The transient is not sinusoidal*
- unstable systems have transient responses that do not go away, so you will never have a sinusoidal output;
- ω is the frequency in radians per second. *ω is in rad/s*

Definition 10.7. Given a system $G(s)$:

- its **frequency response** is $G(j\omega)$, a function of ω ; *Frequency response*
- its **gain** at frequency ω is $|G(j\omega)|$; *Gain*
- its **gain in decibel** (denoted by symbol dB) is $20 \log_{10} |G(j\omega)|$ (gain $|G(j\omega)|$ is often called gain in absolute value, to avoid confusion with the gain in decibel); *Gain in dB*
Gain in absolute value
- its **phase** at frequency ω is $\angle G(j\omega)$. □ *Phase*

Remark 10.10. These definitions are used even if $G(s)$ is not stable. If the system is stable:

- the gain is the ratio between the amplitude of the steady-state output and the amplitude of the input;
- the phase is the difference in phase between the steady-state output sinusoid and the input sinusoid. □

Example 10.13. Figure 10.12 shows the output of $G(s) = \frac{300(s+1)}{(s+10)(s+100)}$ for a sinusoidal input of frequency 1 rad/s, found as follows:

```
>> s = tf('s');
>> G = 300*(s+1)/((s+10)*(s+100));
>> t = 0 : 0.001 : 30;
>> figure, plot(t,sin(t), t,lsim(G,sin(t),t))
>> xlabel('time [s]'), ylabel('output'), grid
```

The amplitude of the input is 1, by construction; the amplitude of the output is 0.4219. So the gain at 1 rad/s is $\frac{0.4219}{1} = 0.4219$ in absolute value, or $20 \log_{10} 0.4219 = -7.50$ dB. This maximum value is taking place at 26 s, while the corresponding maximum of the input takes place later, at $4 \times 2\pi + \frac{\pi}{2} = 26.7$ s. As the period is $2\pi = 6.28$ s, the phase is $\frac{26.7-26}{6.28} \times 360^\circ = 40^\circ$.

Figure 10.12 also shows the output of $G(s)$ when the frequency is 200 rad/s:

```
>> t = 0 : 0.0001 : 0.2;
>> figure, plot(t,sin(200*t), t,lsim(G,sin(200*t),t))
>> xlabel('time [s]'), ylabel('output'), grid
```

In that case, the amplitude of the input is still 1 and the amplitude of the output is 1.313. So the gain at 200 rad/s is $\frac{1.313}{1} = 1.313$ in absolute value, or $20 \log_{10} 1.313 = 2.37$ dB. This maximum value is taking place at 0.1703 s, while the corresponding maximum of the input takes place earlier, at $5 \times \frac{2\pi}{200} + \frac{\pi}{200} = 0.1649$ s. As the period is $\frac{2\pi}{200} = 0.0314$ s, the phase is $\frac{0.1649-0.1703}{0.0314} \times 360^\circ = -62^\circ$.

In both cases, it is visible that the first oscillations are not sinusoidal, because of both their shape and their varying amplitudes. In other words, the transient has not yet disappeared by then. \square

In the example above, the amplitude of the output was larger than that of the input in one case, and smaller in the other. Also in one case the extremes of the output sinusoid took place earlier than those of the input sinusoid, while in the other case it was the other way round.

Definition 10.8. Given

- a stable system $G(s)$,
- with sinusoidal input of frequency ω and amplitude A_u ,
- with steady-state sinusoidal output also of frequency ω and amplitude $A_y = A_u |G(j\omega)|$,

then:

- If the amplitude of the output is larger than the amplitude of the input, $A_y > A_u$, the system is **amplifying** its input:

Amplification

$$A_y > A_u \Rightarrow |G(j\omega)| = \frac{A_y}{A_u} > 1 \Rightarrow 20 \log_{10} |G(j\omega)| > 0 \text{ dB} \quad (10.69)$$

That is to say:

- the gain in absolute value is larger than 1;
- the gain in decibel is larger than 0 dB.

- If the amplitude of the output is smaller than the amplitude of the input, $A_y < A_u$, the system is **attenuating** its input:

Attenuation

$$A_y < A_u \Rightarrow |G(j\omega)| = \frac{A_y}{A_u} < 1 \Rightarrow 20 \log_{10} |G(j\omega)| < 0 \text{ dB} \quad (10.70)$$

That is to say:

- the gain in absolute value is smaller than 1;
- the gain in decibel is smaller than 0 dB.

- If the amplitude of the output and the amplitude of the input are the same, $A_y = A_u$, the system is neither amplifying nor attenuating its input:

$$A_y = A_u \Rightarrow |G(j\omega)| = \frac{A_y}{A_u} = 1 \Rightarrow 20 \log_{10} |G(j\omega)| = 0 \text{ dB} \quad (10.71)$$

That is to say:

- the gain in absolute value is 1;
- the gain in decibel is 0 dB.

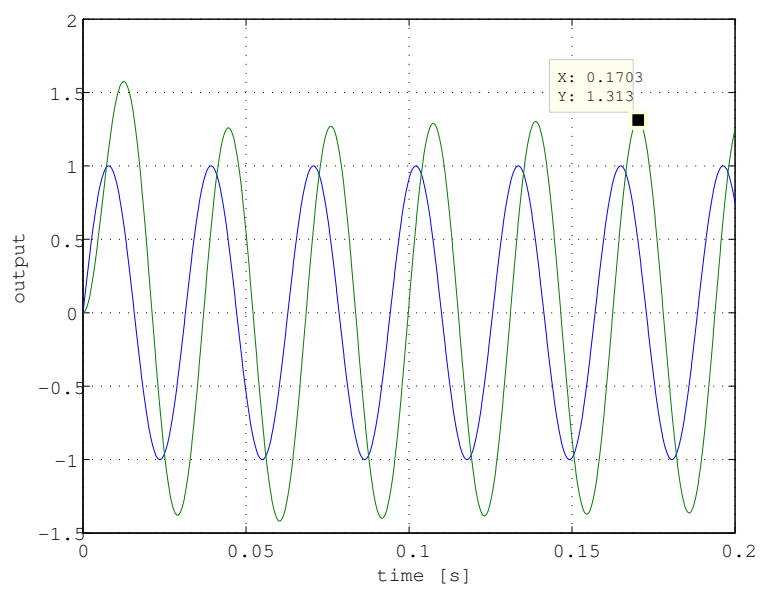
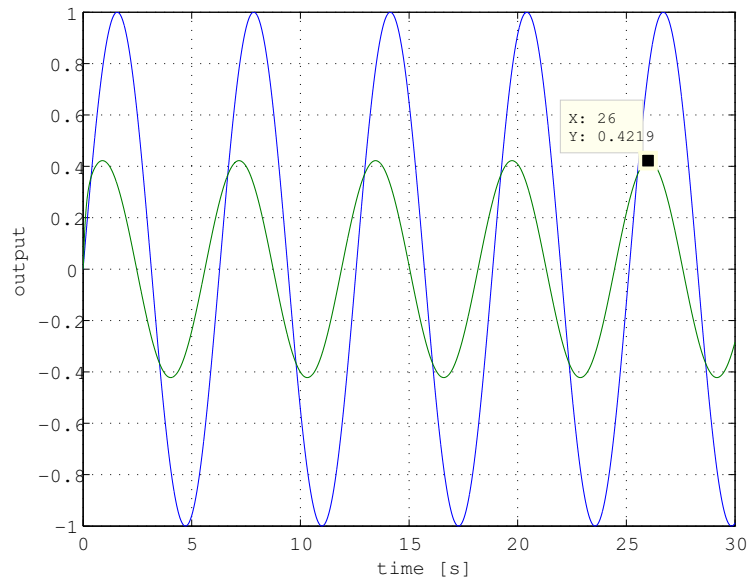


Figure 10.12: Response of $G(s) = \frac{300(s+1)}{(s+10)(s+100)}$ to two sinusoids with different periods.

Table 10.1: Gain values; A_u is the amplitude of the input sinusoid and A_y is the amplitude of the steady-state output sinusoid

	Gain in absolute value	Gain in decibel	Amplitudes
Minimum value	$ G(j\omega) = 0$	$20 \log_{10} G(j\omega) = -\infty$ dB	$A_y = 0$
Attenuation	$0 < G(j\omega) < 1$	$20 \log_{10} G(j\omega) < 0$ dB	$A_y < A_u$
Input and output with same amplitude	$ G(j\omega) = 1$	$20 \log_{10} G(j\omega) = 0$ dB	$A_y = A_u$
Amplification	$ G(j\omega) > 1$	$20 \log_{10} G(j\omega) > 0$ dB	$A_y > A_u$

Furthermore:

Phase lead

- If the extremes of the output take place earlier than the corresponding extremes of the input, the output **leads** in relation to the input; this means that

$$\angle G(j\omega) > 0 \quad (10.72)$$

Phase lag

- If the extremes of the output take place later than the corresponding extremes of the input, the output **lags** in relation to the input; this means that

$$\angle G(j\omega) < 0 \quad (10.73)$$

- If the extremes of the output and the corresponding extremes of the input take place at the same time, the output and the input are in phase; this means that

$$\angle G(j\omega) = 0 \quad (10.74)$$

Phase opposition

- If the maxima of the output and the minima of the input take place at the same time, and vice versa, the output and the input are in **phase opposition**; this means that

$$\angle G(j\omega) = \pm 180^\circ = \pm \pi \text{ rad} \quad \square \quad (10.75)$$

Remark 10.11. Notice that, since sinusoids are periodic, the phase is defined up to 360° shifts: a 90° phase is undistinguishable from a -270° phase, or for that matter from a 3690° phase or any $90^\circ + k360^\circ$, $k \in \mathbb{Z}$ phase. While each of these values can be in principle arbitrarily chosen, it usual to make the phase vary continuously (as much as possible) with frequency, starting from values for low frequencies determined as we will see below in Section 11.4. \square

Gain crossover frequency

Definition 10.9. Frequencies ω_{gc} at which the frequency response of a plant $G(s)$ verifies

$$|G(j\omega)| = 1 \Leftrightarrow 20 \log_{10}|G(j\omega)| = 0 \text{ dB} \quad (10.76)$$

are called **gain crossover frequencies**. \square

Phase crossover frequency

Definition 10.10. Frequencies at which input and output are in phase opposition are called **phase crossover frequencies**. \square

Gain values can be summed up as shown in Table 10.1.

Example 10.14. Consider the responses to sinusoidal inputs of $G(s) = \frac{1}{s^2 + 0.5s + 1}$ in Figure 10.13.

- For $\omega = 0.5$ rad/s:
 - The amplitude of the output is larger than that of the input, so we must have

$$|G(j0.5)| > 1 \Leftrightarrow 20 \log_{10}|G(j0.5)| > 0 \text{ dB} \quad (10.77)$$

- In fact, the gain is

$$\begin{aligned} |G(j0.5)| &= \left| \frac{1}{(j0.5)^2 + 0.5j0.5 + 1} \right| = \left| \frac{1}{1 - 0.25 + j0.25} \right| = \frac{1}{\sqrt{0.75^2 + 0.25^2}} = 1.26 \\ \Rightarrow 20 \log_{10} G(j0.5) &= 20 \log_{10} 1.26 = 2 \text{ dB} \end{aligned} \quad (10.78)$$

– The output is delayed in relation to the input, so we must have $\angle G(j0.5) < 0$.

– In fact, the phase is

$$\angle G(j0.5) = \angle \left(\frac{1}{0.75 + j0.25} \right) = \angle 1 - \angle(0.75 + j0.25) = 0^\circ - \arctan \frac{0.25}{0.75} = -18^\circ \quad (10.79)$$

• For $\omega = 1$ rad/s:

– The amplitude of the output is even larger now, so

$$|G(j)| > |G(j0.5)| = 1.26 \Leftrightarrow 20 \log_{10} |G(j)| > 20 \log_{10} |G(j0.5)| = 2 \text{ dB} \quad (10.80)$$

– In fact, the gain is

$$\begin{aligned} |G(j)| &= \left| \frac{1}{j^2 + 0.5j + 1} \right| = \left| \frac{1}{j0.5} \right| = \frac{1}{0.5} = 2 \\ \Rightarrow 20 \log_{10} |G(j)| &= 20 \log_{10} 2 = 6 \text{ dB} \end{aligned} \quad (10.81)$$

– The output is delayed in relation to the input. Furthermore, the output crosses zero as the input is already at a peak or at a through. So the phase is negative, and equal to -90° .

– In fact,

$$\angle G(j) = \angle \left(\frac{1}{j0.5} \right) = \angle 1 - \angle(j0.5) = 0^\circ - 90^\circ = -90^\circ \quad (10.82)$$

• For $\omega = 2$ rad/s:

– The amplitude of the input is larger than that of the output, so we must have

$$|G(j2)| < 1 \Leftrightarrow 20 \log_{10} |G(j2)| < 0 \text{ dB} \quad (10.83)$$

– In fact, the gain is

$$\begin{aligned} |G(j2)| &= \left| \frac{1}{(2j)^2 + 0.5j2 + 1} \right| = \left| \frac{1}{-3 + j} \right| = \frac{1}{\sqrt{9+1}} = 0.316 \\ \Rightarrow 20 \log_{10} |G(j2)| &= 20 \log_{10} 10^{-\frac{1}{2}} = -10 \text{ dB} \end{aligned} \quad (10.84)$$

– The output is delayed in relation to the input. Furthermore, input and output are almost in phase opposition, but not yet. So we must have $0^\circ < \angle G(j2) < -180^\circ$, but close to the latter value.

– In fact, the phase is

$$\angle G(j2) = \angle \left(\frac{1}{-3 + j} \right) = \angle 1 - \angle(-3 + j) = 0^\circ - \arctan \frac{1}{-3} = -162^\circ \quad \square \quad (10.85)$$

The **Bode diagram**, or Bode plot, is a graphical representation of the frequency response of a system, as a function of frequency. This diagram comprises two plots: *Bode diagram*

- a top plot, showing the gain in dB (y -axis) as a function of frequency in a semi-logarithmic scale (x -axis);
- a bottom plot, showing the phase in degrees (y -axis) as a function of frequency in a semi-logarithmic scale (x -axis).

Frequency is usually given in rad/s, but sometimes in Hz.

Definition 10.11. A frequency variation corresponding to a 10-fold increase or decrease is called **decade**. In a Bode diagram, since the frequency is shown in a logarithmic scale, decades are equally spaced. \square *Decade*

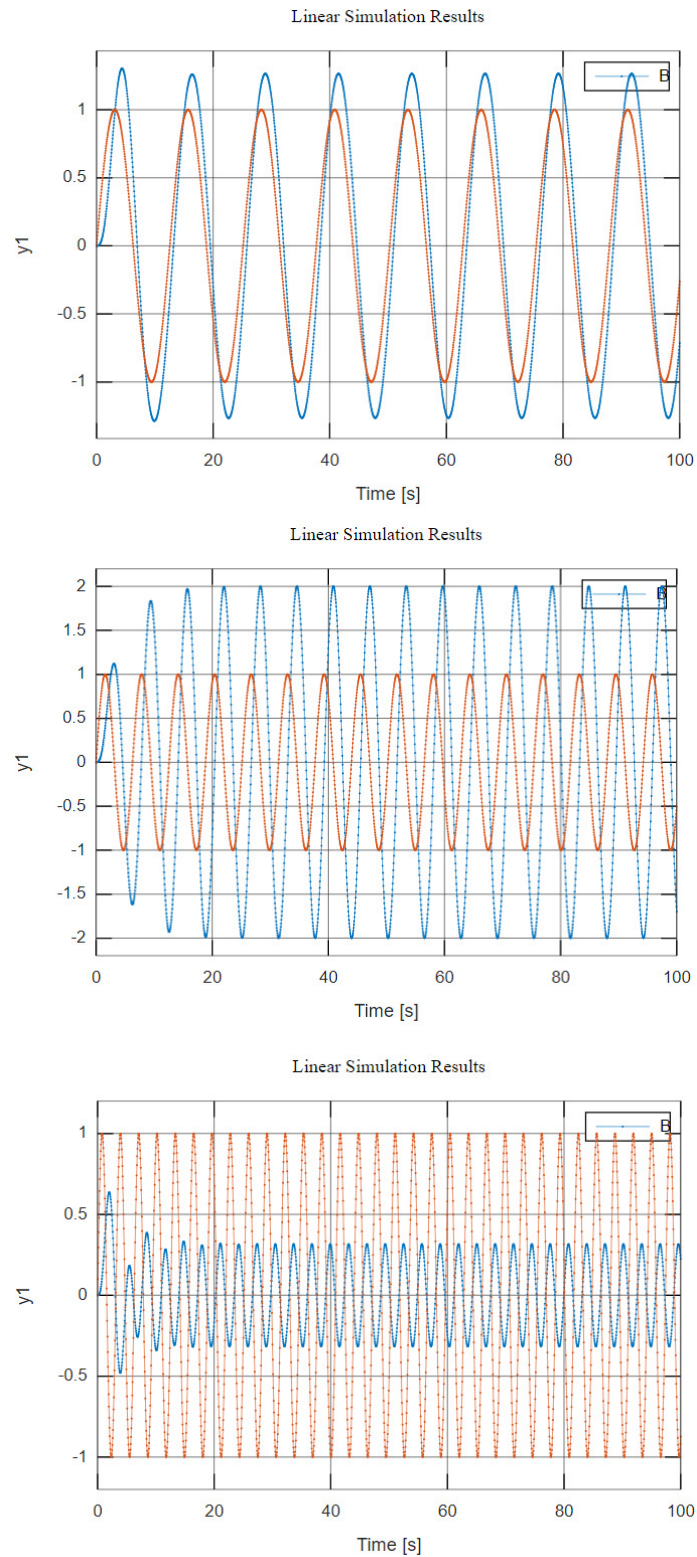


Figure 10.13: Responses of $G(s) = \frac{1}{s^2 + 0.5s + 1}$ (blue) to input sinusoids (red) with 0.5 rad/s (top), 1 rad/s (centre) and 2 rad/s (bottom).

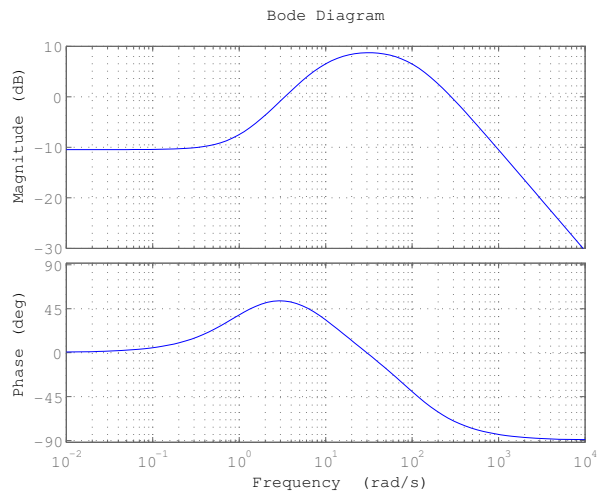


Figure 10.14: Bode diagram of $G(s) = \frac{300(s+1)}{(s+10)(s+100)}$.

In the following sections we will learn how to plot by hand the Bode diagram of any plant (or at least a reasonable approximation thereof); meanwhile, the following MATLAB commands can be used instead:

- `bode` plots the Bode diagram of a system;
- `freqresp` calculates the frequency response of a system.

Example 10.15. The Bode diagram in Figure 10.14 of $G(s) = \frac{300(s+1)}{(s+10)(s+100)}$ MATLAB's *command* `bode` from Example 10.13 is found as follows:

```
>> s = tf('s');
>> G = 300*(s+1)/((s+10)*(s+100));
>> figure, bode(G), grid
```

The gains and phases at $\omega = 1$ rad/s and $\omega = 200$ rad/s found in Example 10.13 can be observed in the diagram.

This way we first find the frequency response and then use it to plot the MATLAB's *command* `freqresp` Bode diagram:

```
>> [Gjw, w] = freqresp(G); % Gjw returned as a 3-dimensional tensor...
>> Gjw = squeeze(Gjw); % ...must now be squeezed to a vector
>> figure, subplot(2,1,1), semilogx(w, 20*log10(abs(Gjw)))
>> grid, xlabel('frequency [rad/s]'), ylabel('gain [dB]'), title('Bode diagram')
>> subplot(2,1,2), semilogx(w, rad2deg(unwrap(angle(Gjw))))
>> grid, ylabel('phase [degrees]') % unwrap avoids jumps of 360 degrees
```

To find the gains and phases to confirm those found in Example 10.13:

```
>> Gjw = freqresp(G, [1 200])
Gjw(:,:,1) =
    0.3294 + 0.2640i
Gjw(:,:,2) =
    0.6525 - 1.1704i
>> gains = 20*log10(abs(Gjw))
gains(:,:,1) =
   -7.4909
gains(:,:,2) =
    2.5420
>> phases = rad2deg(unwrap(angle(Gjw)))
phases(:,:,1) =
   38.7165
phases(:,:,2) =
  -60.8590
```

Here's another way to find the same values:

```

>> w = [1 200];
>> Gjw = 300*(1i*w+1)/((1i*w+10).*(1i*w+100));
Gjw =
    0.3271 + 0.2676i    0.6186 - 1.2212i
>> gains = 20*log10(abs(Gjw))
gains =
    -7.4909    2.5420
>> phases = rad2deg(unwrap(angle(Gjw)))
phases =
    38.7165   -60.8590

```

Notice the small differences due to numerical errors. □

Example 10.16. The Bode diagram in Figure 10.15 of $G(s) = \frac{1}{s^2+0.5s+1}$ from Example 10.14 shows the gains and phases found in that example, that can also be found as follows:

```
>> G = tf(1,[1 .5 1])
```

```
G =
      1
-----
s^2 + 0.5 s + 1
```

Continuous-time transfer function.

```

>> figure,bode(G),grid on
>> Gjw = squeeze(freqresp(G, [.5 1 2]))
Gjw =
    1.2000 - 0.4000i
    0.0000 - 2.0000i
   -0.3000 - 0.1000i
>> gains = 20*log10(abs(Gjw))
gains =
    2.0412
    6.0206
   -10.0000
>> phases = rad2deg(unwrap(angle(Gjw)))
phases =
   -18.4349
   -90.0000
  -161.5651

```

□

Example 10.17. From the Bode diagram in Figure 10.16, even without knowing what transfer function it belongs to, we can conclude the following:

- At $\omega = 0.1$ rad/s, the gain is 20 dB (i.e. $10^{\frac{20}{20}} = 10$ in absolute value) and the phase is $0^\circ = 0$ rad. So, if the input is

$$u(t) = 5 \sin(0.1t + \frac{\pi}{6}) \quad (10.86)$$

the steady-state output will be

$$y(t) = 5 \times 10 \sin(0.1t + \frac{\pi}{6}) = 50 \sin(0.1t + \frac{\pi}{6}) \quad (10.87)$$

- At $\omega = 10$ rad/s, the gain is 17 dB (i.e. $10^{\frac{17}{20}} = 7.1$ in absolute value) and the phase is $-45^\circ = -\frac{\pi}{4}$ rad. So, if the input is

$$u(t) = 5 \sin(10t + \frac{\pi}{6}) \quad (10.88)$$

the steady-state output will be

$$y(t) = 5 \times 7.1 \sin(10t + \frac{\pi}{6} - \frac{\pi}{4}) = 35.5 \sin(10t - \frac{\pi}{12}) \quad (10.89)$$

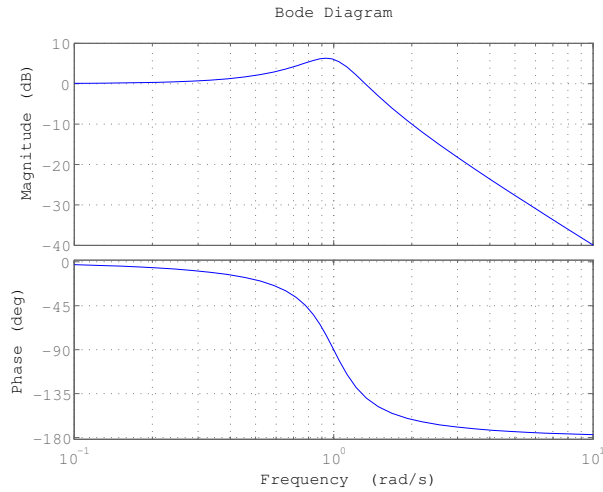


Figure 10.15: Bode diagram of $\frac{1}{s^2 + 0.5s + 1}$.

- At $\omega = 100$ rad/s, the gain is 0 dB (i.e. $10^0 = 1$ in absolute value) and the phase is $-85^\circ = -1.466$ rad. So, if the input is

$$u(t) = 5 \sin(100t + \frac{\pi}{6}) \quad (10.90)$$

the steady-state output will be

$$y(t) = 5 \times 1 \sin(100t + 0.524 - 1.466) = 5 \sin(100t - 0.942) \quad (10.91)$$

- At $\omega = 1000$ rad/s, the gain is -20 dB (i.e. $10^{-\frac{20}{20}} = 0.1$ in absolute value) and the phase is $-90^\circ = -\frac{\pi}{2}$ rad. So, if the input is

$$u(t) = 5 \sin(1000t + \frac{\pi}{6}) \quad (10.92)$$

the steady-state output will be

$$y(t) = 5 \times 0.1 \sin(1000t + \frac{\pi}{6} - \frac{\pi}{2}) = 0.5 \sin(1000t - \frac{\pi}{3}) \quad (10.93)$$

- The system is linear. So, if the input is

$$u(t) = 0.5 \sin(0.1t + \frac{\pi}{6}) + 25 \sin(1000t + \frac{\pi}{6}) \quad (10.94)$$

the steady-state output will be

$$\begin{aligned} y(t) &= 0.5 \times 10 \sin(0.1t + \frac{\pi}{6}) + 25 \times 0.1 \sin(1000t + \frac{\pi}{6} - \frac{\pi}{2}) \\ &= 5 \sin(0.1t + \frac{\pi}{6}) + 2.5 \sin(1000t - \frac{\pi}{3}) \end{aligned} \quad (10.95)$$

Notice how the frequency with the largest amplitude in the input now has the smallest. \square

Glossary

And he said: Behold, it is one people, and one tongue is to all: and they haue begunne to doe this, neyther wil they leaue of from their determinations, til they accomplish them indede. Come ye therefore, let vs goe downe, and there confound their tongue, that none may heare is neighbours voice.

MOSES (attrib.; 13th c. BC?), *Genesis*, xi 6–7, Douay-Rheims version (1609)

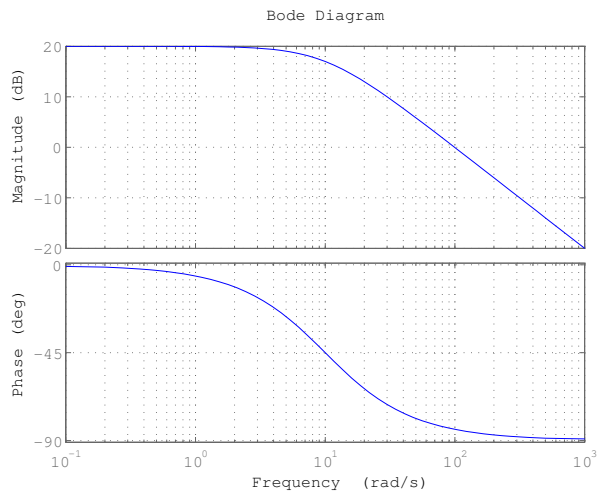


Figure 10.16: Bode diagram of Example 10.17.

amplification amplificação
attenuation atenuação
bounded signal sinal limitado
decade década
frequency response resposta em frequência
gain ganho
gain crossover frequency frequência de cruzamento de ganho
impulse impulso
marginally stable marginalmente estável
phase fase
phase crossover frequency frequência de cruzamento de fase
phase lag atraso de fase
phase lead avanço de fase
phase opposition oposição de fase
phase crossover frequency frequência de cruzamento de fase
ramp rampa
square wave onda quadrada
static gain ganho estacionário
steady-state estado estacionário
steady-state response resposta estacionária
stable estável
step degrau, escalão
transient response resposta transiente
triangle wave onda triangular
unit ramp rampa unitária
unit step degrau unitário
unstable instável

Exercises

- For each of the following pairs of a transfer function and an input:
 - find the Laplace transform of the input;
 - find the Laplace transform of the output;
 - find the value of the output for $t \gg 1$ without using the inverse Laplace transform;
 - find the output as a function of time;
 - separate that function of time into a transient and a steady state;
 - confirm the value of the output for $t \gg 1$ found previously.

(a) $G(s) = \frac{10}{s^2 + 21s + 20}$ and $u(t) = 0.4, t > 0$

- (b) $G(s) = \frac{5}{s + 0.1}$ and $u(t) = 2t$, $t > 0$
- (c) $G(s) = \frac{s}{s^2 + s + 1}$ and $u(t) = \delta(t)$
- (d) $G(s) = \frac{s}{s^2 + s + 1}$ and $u(t) = 0.4$, $t > 0$
- (e) $G(s) = \frac{7}{s}$ and $u(t) = 0.4$, $t > 0$
2. From the poles of the transfer functions of Exercise 1 of Chapter 9, explain which of them are stable, unstable, or marginally stable.
3. Figure 10.17 shows the Bode diagrams of some transfer functions. For each of them, read in the Bode diagram the values from which you can calculate the transfer function's steady state response to the following inputs:
- $u(t) = \sin(2t)$
 - $u(t) = \sin(2t + \frac{\pi}{2})$
 - $u(t) = \sin(1000t)$
 - $u(t) = 10 \sin(1000t)$
 - $u(t) = \frac{1}{3} \sin(0.1t - \frac{\pi}{4}) + \sin(2t + \frac{\pi}{2})10 \sin(1000t)$
4. For each of the following transfer functions:
- find the corresponding Fourier transform;
 - find the gain (both in absolute value and in decibel) and the phase (in radians or degrees, as you prefer) at the indicated frequencies.
- (a) $G(s) = \frac{5}{s + 0.1}$ and $\omega = 0.01, 0.1, 1$ rad/s
- (b) $G(s) = \frac{s}{s^2 + s + 1}$ and $\omega = 0.1, 1, 10$ rad/s
- (c) $G(s) = \frac{7}{s}$ and $\omega = 1, 10, 100$ rad/s
5. In naval and ocean engineering it is usual to call **Response Amplitude Operator** (RAO) to what we called gain. It is often represented in absolute value in a linear plot as a function of frequency. Figure 10.18 shows the RAO of four different heaving buoys. Suppose that each of them is subject to waves with an amplitude of 2 m and a frequency of 2π rad/s. What will be the amplitude of the oscillation of each buoy?
6. Consider transfer function $G(s) = \frac{10}{10s + 1}$.
- (a) When the input is a unit step, what will the steady state response be?
 - (b) When the input is a step with amplitude 3, what will the steady state response be?
 - (c) Without computing an expression for the output, give a rough estimate of how long it takes for the output to reach 20, when the input is a step with amplitude 3.
 - (d) Without computing an expression for the output, give a rough estimate of the 2% settling time, when the input is a step with amplitude 3.
 - (e) Calculate the output as a function of time, using an inverse Laplace transform, and find the exact values of the estimations from the last two questions.
 - (f) Suppose that the input is now a unit step again. What will the new value of the 2% settling time be? *Hint:* is the system linear or non-linear?
7. Show that, in (10.21), if $0 \notin [a, b]$ and function $g(t)$ is bounded, the integrand is zero everywhere, and the integral is zero.

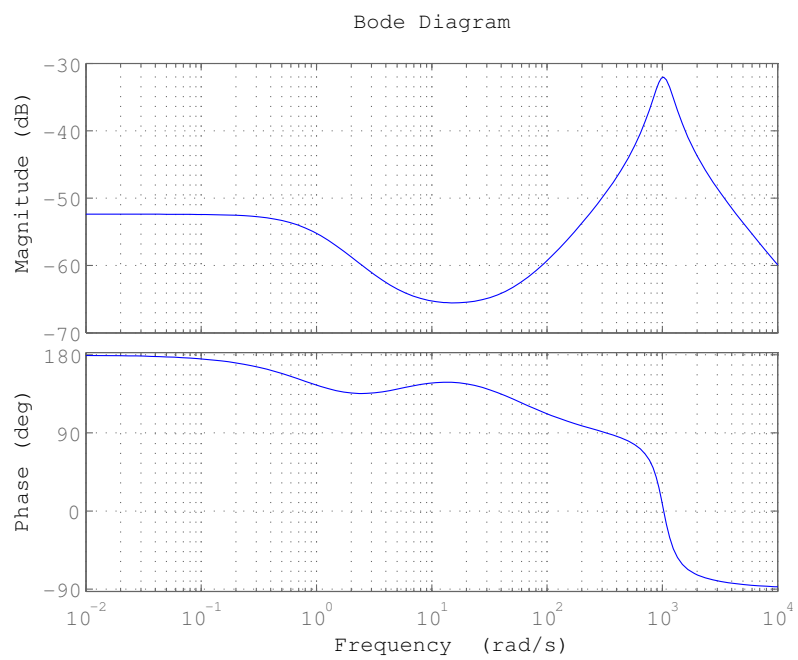
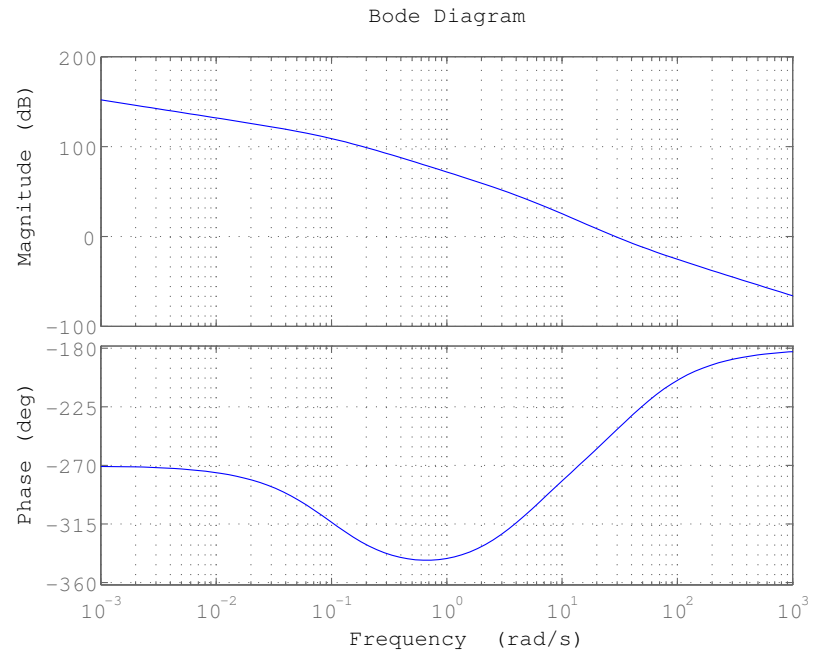


Figure 10.17: Bode diagrams of Exercise 3.

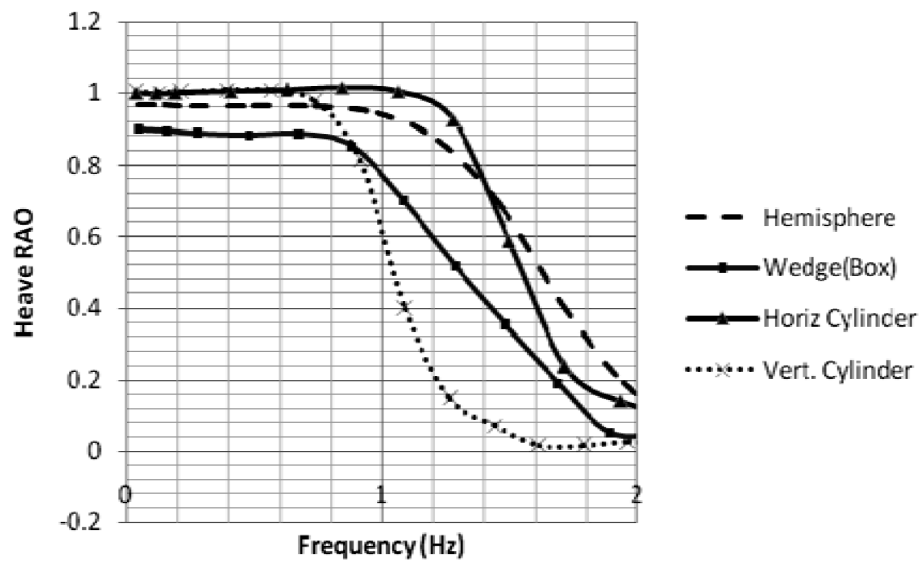


Figure 10.18: RAO of four heaving buoys of Exercise 5 (source: <http://marine-eng.ir/article-1-80-en.pdf>).

Chapter 11

Finding time and frequency responses

A rocket-driven spaceship can, obviously, only accelerate along its major axis — that is, ‘forwards’. Any deviation from a straight course demands a physical turning of the ship, so that the motors can blast in another direction. Everyone knows that this is done by internal gyros or tangential steering jets: but very few people know just how long this simple manoeuvre takes. The average cruiser, fully fuelled, has a mass of two or three thousand tons, which does not make for rapid footwork. But things are even worse than this, for it is not the mass, but the moment of inertia that matters here — and since a cruiser is a long, thin object, its moment of inertia is slightly colossal. The sad fact remains (though it is seldom mentioned by astronautical engineers) that it takes a good ten minutes to rotate a spaceship through 180 degrees, with gyros of any reasonable size. Control jets are not much quicker, and in any case their use is restricted because the rotation they produce is permanent and they are liable to leave the ship spinning like a slow-motion pin-wheel, to the annoyance of all inside.

In the ordinary way, these disadvantages are not very grave. One has millions of kilometres and hundreds of hours in which to deal with such minor matters as a change in the ship’s orientation. It is definitely against the rules to move in ten-kilometre-radius circles, and the commander of the *Doradus* felt distinctly aggrieved.

Arthur C. CLARKE (1917 — †2008), *Hide-and-peek*, Astounding Science Fiction, September 1949

In this chapter we systematically study the time and frequency responses of different systems, beginning with the simplest cases.

11.1 Time and frequency responses of a pole at the origin

In this section we study the behaviour of transfer functions of the form

$$G(s) = \frac{b}{s}, \quad b > 0 \quad (11.1)$$

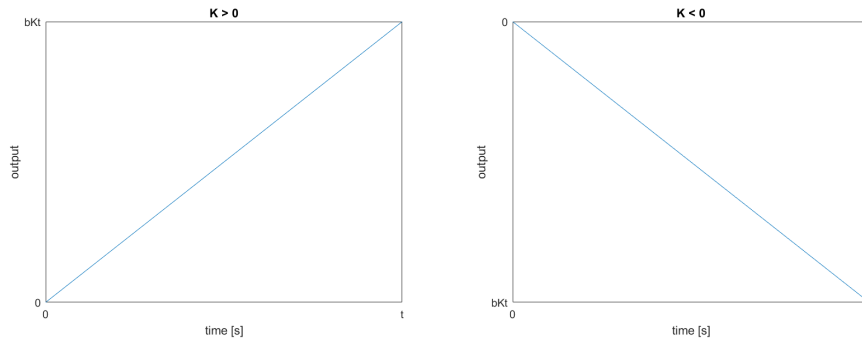
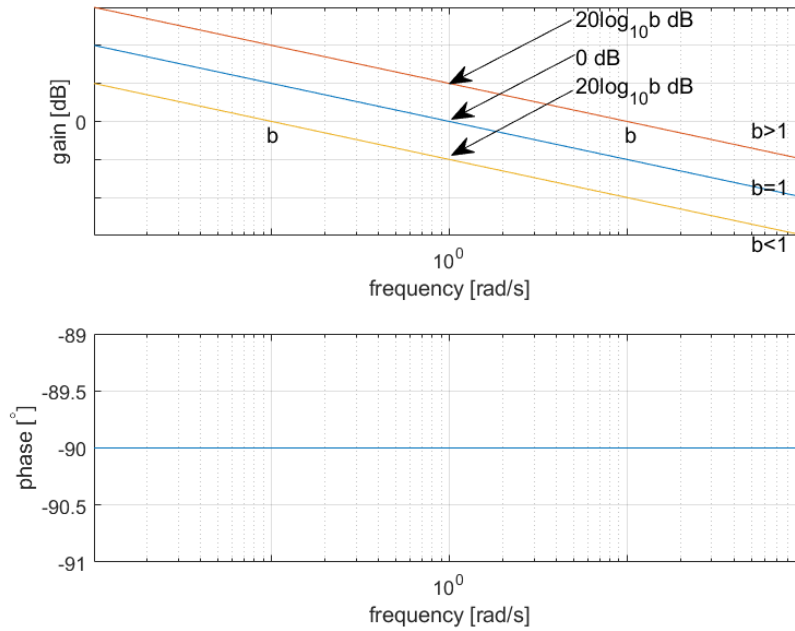
which have one pole at the origin, $s = 0$.

As to time responses, it is clear that, for any input $u(t)$, the output is $\frac{1}{s}$ integrates its input

$$y(t) = \mathcal{L}^{-1} \left[U(s) \frac{b}{s} \right] = b \int_0^t u(t) \quad (11.2)$$

where we have used (2.47). In particular, the unit step response of (11.1) is *Unit step response of $\frac{b}{s}$*

$$y(t) = \mathcal{L}^{-1} \left[\frac{b}{s^2} \right] = bt \quad (11.3)$$

Figure 11.1: Response of $\frac{b}{s}$ to a step of amplitude K .Figure 11.2: Bode diagram of $\frac{b}{s}$.

and in the more general case of a step with amplitude K the response is

$$y(t) = bKt \quad (11.4)$$

Step response of $\frac{b}{s}$

This response will go up if $K > 0$ and go down if $K < 0$, as seen in Figure 11.1.

As we know, (11.1) is marginally stable, since

- its response to a step, which is a bounded input, is not bounded;
- its impulse response, $y(t) = \mathcal{L}^{-1} \left[\frac{b}{s} \right] = bH(t)$, is bounded.

Frequency response of $\frac{b}{s}$

The frequency response of (11.1) is

$$G(j\omega) = \frac{b}{j\omega} \quad (11.5)$$

$$|G(j\omega)| = \left| \frac{b}{j\omega} \right| = \frac{b}{\omega} \quad (11.6)$$

$$20 \log_{10} |G(j\omega)| = \left| \frac{b}{j\omega} \right| = 20 \log_{10} b + 20 \log_{10} \omega \quad (11.7)$$

$$\angle G(j\omega) = \underbrace{\angle b}_{0^\circ} - \underbrace{\angle(j\omega)}_{90^\circ} = -90^\circ \quad (11.8)$$

As to the gain of (11.8), notice that:

- it is shown as a straight line in a Bode diagram, since it has a linear variation with $\log_{10} \omega$;

- the slope of that straight line is always -20 dB per decade;
- there is one gain crossover frequency, which is b rad/s;
- when compared with the case $b = 1$, the gain is shifted up if $b > 1$ or down if $0 < b < 1$, by $20 \log_{10} b$ in either case.

11.2 Time and frequency responses of a first-order system without zeros

In this section we study the behaviour of transfer functions of the form

$$G(s) = \frac{b}{s+a}, \quad a \neq 0, b > 0 \quad (11.9)$$

which have one pole, $s = -a \neq 0$.

The unit step response of (11.9) is, as can be seen in Table 2.1,

Step response of $\frac{b}{s+a}$

$$y(t) = \frac{b}{a} (1 - e^{-at}) \quad (11.10)$$

and, in the more general case of a step with amplitude K , the step response is

$$y(t) = \frac{bK}{a} (1 - e^{-at}) \quad (11.11)$$

As we know, (11.9) is

- stable if $a > 0$ (pole in $-a \in \mathbb{R}^-$), in which case the exponential in (11.10)–(11.11) vanishes when t increases;
- unstable if $a < 0$ (pole in $-a \in \mathbb{R}^+$), in which case the exponential in (11.10)–(11.11) keeps increasing when t increases.

Figure 11.3 shows the general shape of the step response. Notice that, if the system is stable:

- the steady-state response is constant and given by $y_{ss} = \frac{bK}{a}$;
- the response will go up if $K > 0$ and down if $K < 0$;
- the response is monotonous, and thus the output is never larger than the steady-state value;
- the response changes with time as follows:

$$y\left(\frac{0.7}{a}\right) = \frac{bK}{a} (1 - e^{-0.7}) = 0.50 y_{ss} \quad (11.12)$$

$$y\left(\frac{1}{a}\right) = \frac{bK}{a} (1 - e^{-1}) = 0.63 y_{ss} \quad (11.13)$$

$$y\left(\frac{2}{a}\right) = \frac{bK}{a} (1 - e^{-2}) = 0.86 y_{ss} \quad (11.14)$$

$$y\left(\frac{2.3}{a}\right) = \frac{bK}{a} (1 - e^{-2.3}) = 0.90 y_{ss} \quad (11.15)$$

$$y\left(\frac{3}{a}\right) = \frac{bK}{a} (1 - e^{-3}) = 0.95 y_{ss} \quad (11.16)$$

$$y\left(\frac{4}{a}\right) = \frac{bK}{a} (1 - e^{-4}) = 0.98 y_{ss} \quad (11.17)$$

$$y\left(\frac{4.6}{a}\right) = \frac{bK}{a} (1 - e^{-4.6}) = 0.99 y_{ss} \quad (11.18)$$

- $\frac{1}{a}$ is called **time constant**;
- the slope of the response at $t = 0$ is given by

Time constant

$$y'(0) = \frac{bK}{a} a e^{-a \cdot 0} = bK \quad (11.19)$$

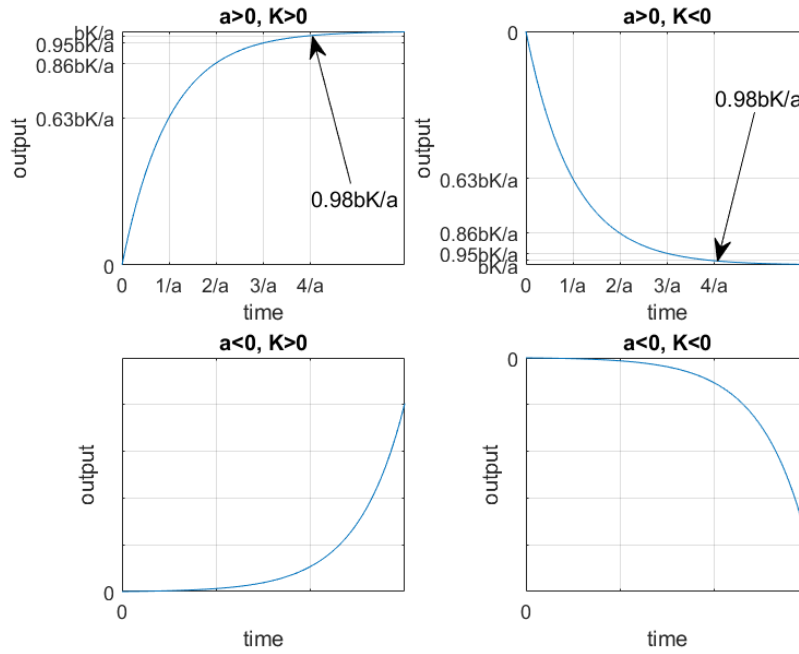


Figure 11.3: Response of $\frac{b}{s+a}$ to a step with amplitude K .

Definition 11.1. The $x\%$ -settling time $t_{s,x\%}$ is the minimum value of time *Settling time* for which

$$\forall t > t_{s,x\%}, \quad y_{ss} \left(1 - \frac{x}{100}\right) \leq y(t) \leq y_{ss} \left(1 + \frac{x}{100}\right) \quad (11.20)$$

The most usual are the 5%-settling time and the 2%-settling time, though other values, such as 10%, 15%, or 1%, are sometimes found. \square

Since step response (11.11) is monotonous, (11.12)–(11.18) show that

- its 10% settling time is $t_{s,10\%} = \frac{2.3}{a}$;
- its 5% settling time is $t_{s,5\%} = \frac{3}{a}$;
- its 2% settling time is $t_{s,2\%} = \frac{4}{a}$;
- its 1% settling time is $t_{s,1\%} = \frac{4.6}{a}$.

Frequency response of $\frac{b}{s+a}$

The frequency response of (11.9) is

$$G(j\omega) = \frac{b}{j\omega + a} = \frac{b(a - j\omega)}{a^2 + \omega^2} \quad (11.21)$$

$$|G(j\omega)| = \left| \frac{b}{j\omega + a} \right| = \frac{b}{\sqrt{a^2 + \omega^2}} \quad (11.22)$$

$$20 \log_{10} |G(j\omega)| = 20 \log_{10} b - 10 \log_{10} (a^2 + \omega^2) \text{ [dB]} \quad (11.23)$$

$$\angle G(j\omega) = 0 - \angle(j\omega + a) = -\arctan \frac{\omega}{a} \quad (11.24)$$

and thus, if $a > 0$ (stable system):

- for low frequencies,

$$\omega \ll a \Rightarrow G(j\omega) \approx \frac{b}{a} \quad (11.25)$$

$$\omega \ll a \Rightarrow |G(j\omega)| \approx \left| \frac{b}{a} \right| = \frac{b}{a} \quad (11.26)$$

$$\omega \ll a \Rightarrow 20 \log_{10} |G(j\omega)| \approx 20 \log_{10} \frac{b}{a} \text{ [dB]} \quad (11.27)$$

$$\omega \ll a \Rightarrow \angle G(j\omega) \approx \angle \left(\frac{b}{a} \right) = 0^\circ \quad (11.28)$$

- for frequency $\omega = a$,

$$G(ja) = \frac{b}{a(j+1)} \quad (11.29)$$

$$|G(ja)| = \left| \frac{b}{a(j+1)} \right| = \frac{b}{a} \frac{1}{\sqrt{2}} \quad (11.30)$$

$$20 \log_{10} |G(ja)| = 20 \log_{10} \frac{b}{a} - 20 \log_{10} \sqrt{1+1} \approx 20 \log_{10} \frac{b}{a} - 3 \text{ [dB]} \quad (11.31)$$

$$\angle G(ja) = 0 - \angle(j+1) = -45^\circ \quad (11.32)$$

- for high frequencies,

$$\omega \gg a \Rightarrow G(j\omega) \approx \frac{b}{j\omega} \quad (11.33)$$

$$\omega \gg a \Rightarrow |G(j\omega)| \approx \left| \frac{b}{j\omega} \right| = \frac{b}{\omega} \quad (11.34)$$

$$\omega \gg a \Rightarrow 20 \log_{10} |G(j\omega)| \approx 20 \log_{10} b - 20 \log_{10} \omega \text{ [dB]} \quad (11.35)$$

$$\omega \gg a \Rightarrow \angle G(j\omega) \approx \angle \left(\frac{b}{j\omega} \right) = -90^\circ \quad (11.36)$$

In other words:

- for low frequencies, $\frac{b}{s+a}$ is similar to constant $\frac{b}{a}$, with a zero phase and a constant gain;
- for high frequencies, $\frac{b}{s+a}$ is similar to $\frac{b}{s}$, with a -90° phase and a linear gain with a slope of -20 dB per decade;
- the phase goes down from 0° for low frequencies to -90° for high frequencies.

If $a < 0$, the gain is the same; only the phase changes:

- for low frequencies,

$$\omega \ll a \Rightarrow \angle G(j\omega) \approx \angle \left(\frac{b}{a} \right) = -180^\circ \quad (11.37)$$

- for frequency $\omega = |a|$,

$$G(j|a|) = \frac{b}{a + j|a|} = \frac{b}{|a|(-1 + j)} \quad (11.38)$$

$$\omega = |a| \Rightarrow \angle G(j\omega) = 0 - \angle(-1 + j) = -135^\circ \quad (11.39)$$

- for high frequencies,

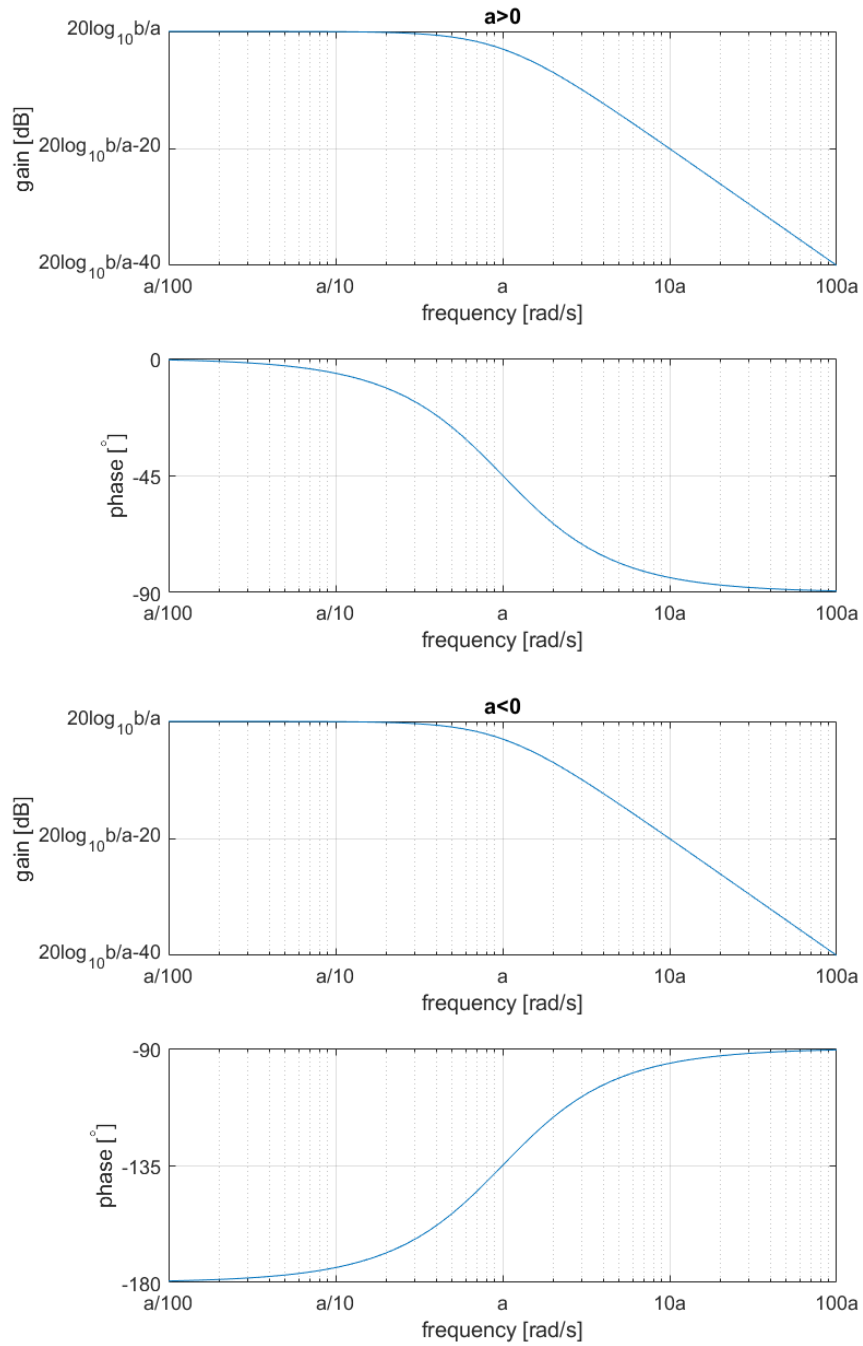
$$\omega \gg a \Rightarrow G(j\omega) \approx \frac{b}{j\omega} \omega \gg a \Rightarrow \angle G(j\omega) \approx \angle \left(\frac{b}{j\omega} \right) = -90^\circ \quad (11.40)$$

So in this case the phase goes up from -180° for low frequencies to -90° for high frequencies. Remember that since the system is unstable in this case the output will not be a sinusoid in steady state (it will have diverged exponentially to infinity).

Figure 11.4 shows the Bode diagram of (11.9).

Remark 11.1. The Bode diagram of (11.9) in Figure 11.4 is often approximated by its asymptotes shown in Figure 11.5. \square

Remark 11.2. The steady-state response of (11.9) to a unit step $\frac{b}{a}$ can be found from the gain for low frequencies (11.25). In fact, a low frequency corresponds to a large time span. \square

Figure 11.4: Bode diagram of $\frac{b}{s+a}$.

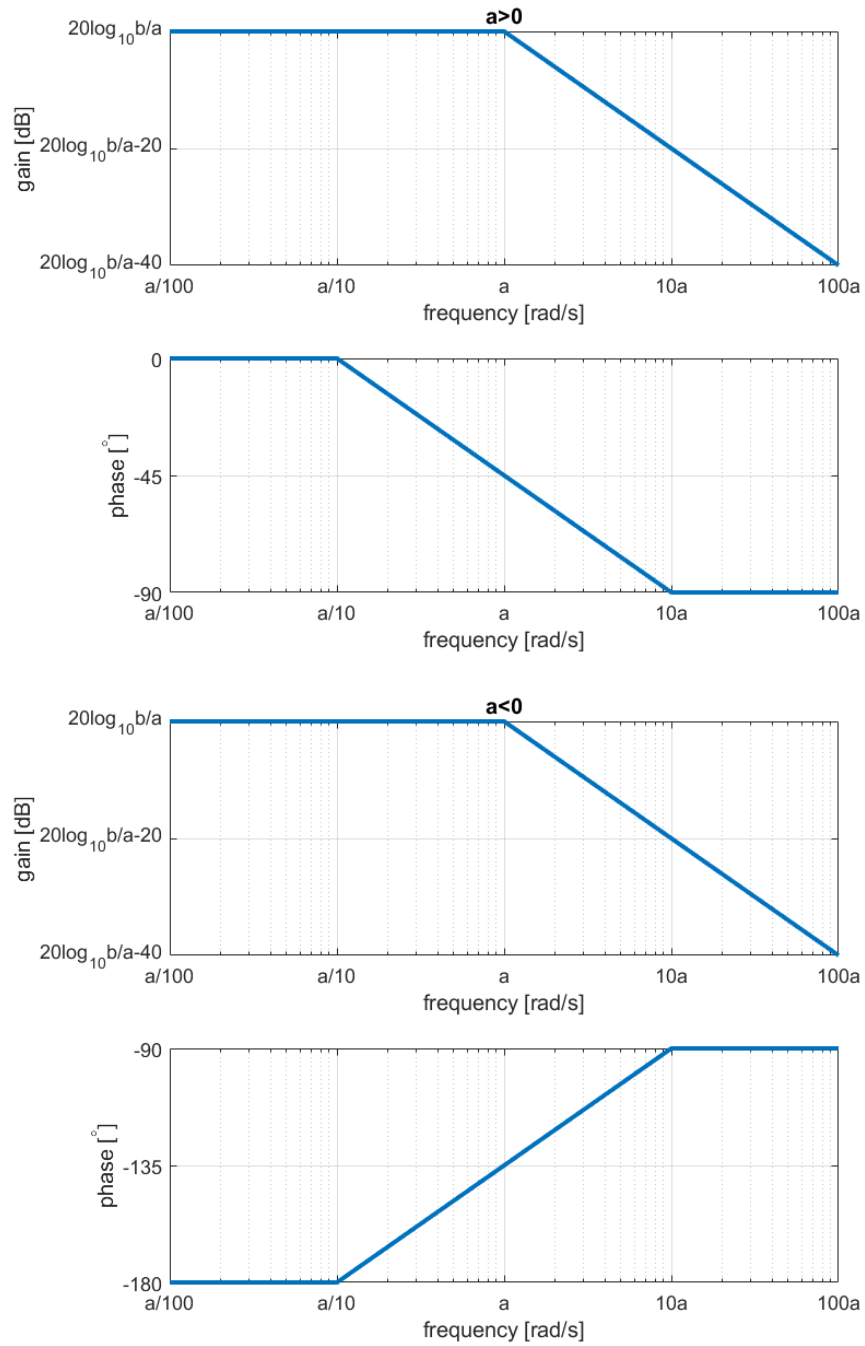


Figure 11.5: Asymptotes of the Bode diagram of $\frac{b}{s+a}$ in Figure 11.4.

11.3 Time and frequency responses of a second-order system without zeros

In this section we study the behaviour of transfer functions of the form

$$G(s) = \frac{b}{s^2 + a_1s + a_2}, \quad b > 0 \quad (11.41)$$

which have two poles. It is usual to rewrite (11.41) as

$$G(s) = \frac{b}{s^2 + 2\xi\omega_n s + \omega_n^2} = \frac{b}{(s + \xi\omega_n + \sqrt{\xi^2 - 1})(s + \xi\omega_n - \sqrt{\xi^2 - 1})}, \quad b > 0 \quad (11.42)$$

For reasons we will see later, ξ is called damping coefficient and ω_n natural frequency. The poles of (11.42) are

$$p_1 = -\xi\omega_n + \omega_n\sqrt{\xi^2 - 1} \quad (11.43)$$

$$p_2 = -\xi\omega_n - \omega_n\sqrt{\xi^2 - 1} \quad (11.44)$$

Notice that:

- (11.41) can always be put in the form (11.42), save
 - when $a_2 < 0$, in which case the transfer function is unstable, as can be easily seen in a way we will learn below in section 11.5),
 - or when $a_2 = 0 \neq a_1$, in which case there is one pole at the origin, and the transfer function is marginally stable;
- if either ξ or ω_n are negative the real part of the poles is positive and (11.42) is unstable too;
- if both ξ or ω_n are negative, (11.42) is just as if both were positive, so that case need not be considered;
- if $\omega_n = 0$ then $G(s) = \frac{b}{s^2}$, which is unstable;
- if $\xi \geq 1$ then the poles are real, in which case we can write

$$G(s) = \frac{b}{(s + p_1)(s + p_2)}, \quad b > 0, \quad p_1, p_2 \in \mathbb{R} \quad (11.45)$$

- if, in particular, $\xi = 1$ then the two real poles are both equal to $-\omega_n$:

$$G(s) = \frac{b}{(s + \omega_n)^2}, \quad b, \omega_n > 0 \quad (11.46)$$

- if $0 \leq \xi < 1$ the system is not unstable and the complex conjugate poles are given by

$$p_1 = -\xi\omega_n + j\omega_n\sqrt{1 - \xi^2} \quad (11.47)$$

$$p_2 = -\xi\omega_n - j\omega_n\sqrt{1 - \xi^2} \quad (11.48)$$

Five cases will be treated separately:

Overdamped system

- $\xi > 1$, i.e. the two poles are real and different, as in (11.45): in this case the system is called **overdamped**;

Critically damped system

- $\xi = 1$, i.e. there is a double pole, as in (11.46): in this case the system is called **critically damped**;

Underdamped system

- $0 < \xi < 1$, i.e. the two poles are complex conjugate and stable: in this case the system is called **underdamped**;

System with no damping

- $\xi = 0$, i.e. the two poles are complex conjugate and marginally stable: in this case the system is said to have **no damping**;

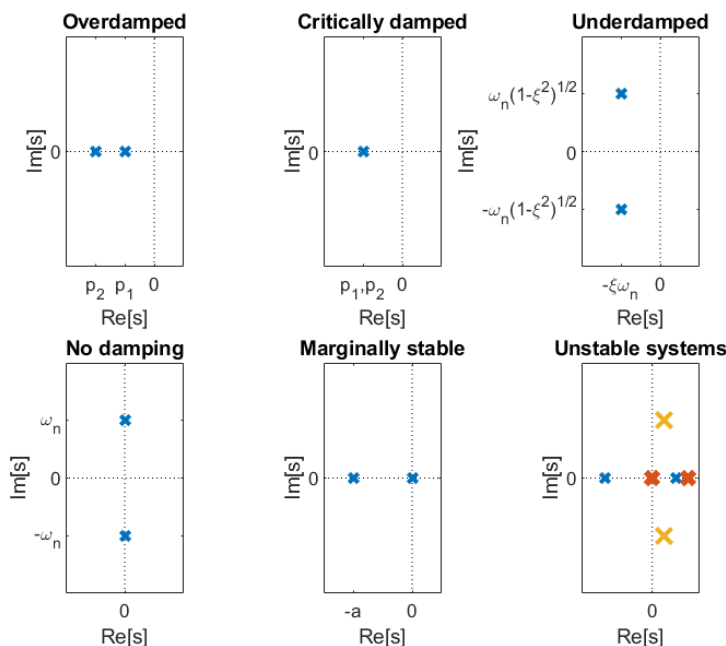


Figure 11.6: Location of the poles in the complex plane of second order systems. Notice that the bottom left undamped system is marginally stable, as well as the bottom centre one with a pole at the origin.

- $G(s) = \frac{b}{s^2+a_1s} = \frac{b}{s(s+a_1)}$, $b > 0$, i.e. the transfer function is marginally stable with one pole in the origin.

See Figure 11.6.

response of damped system If both poles of (11.41) are real, different from zero and different from each other, then its response to a step of amplitude K is, as can be seen in Table 2.1,

$$y(t) = \frac{bK}{p_1 p_2} \left(1 + \frac{1}{p_1 - p_2} (p_2 e^{-p_1 t} - p_1 e^{-p_2 t}) \right) \quad (11.49)$$

As expected, should either p_1 or p_2 be negative, (11.49) would diverge to infinity. Assuming that both p_1 and p_2 are positive, we can rewrite (11.49) using (11.43)–(11.44):

$$\begin{aligned} p_1 p_2 &= (-\xi \omega_n + \omega_n \sqrt{\xi^2 - 1})(-\xi \omega_n - \omega_n \sqrt{\xi^2 - 1}) \\ &= \xi^2 \omega_n^2 - \omega_n^2 (\xi^2 - 1) = \omega_n^2 \end{aligned} \quad (11.50)$$

$$p_1 - p_2 = -\xi \omega_n + \omega_n \sqrt{\xi^2 - 1} - (-\xi \omega_n - \omega_n \sqrt{\xi^2 - 1}) = 2\omega_n \sqrt{1 - \xi^2} \quad (11.51)$$

$$y(t) = \frac{bK}{\omega_n^2} \left(1 + \frac{1}{2\sqrt{1 - \xi^2}} \left((-\xi - \sqrt{\xi^2 - 1}) e^{(\xi - \sqrt{\xi^2 - 1})\omega_n t} - (-\xi + \sqrt{\xi^2 - 1}) e^{(\xi + \sqrt{\xi^2 - 1})\omega_n t} \right) \right) \quad (11.52)$$

Then:

- the steady-state response is constant and given by $y_{ss} = \frac{bK}{p_1 p_2} = \frac{bK}{\omega_n^2}$;
- the response begins at 0 with a horizontal slope, since

$$y'(t) = \frac{bK}{p_1 p_2} \frac{1}{p_1 - p_2} (-p_1 p_2 e^{-p_1 t} + p_1 p_2 e^{-p_2 t}) \quad (11.53)$$

and thus $y'(0) = 0$;

- the response is monotonous, since $y'(t) = \frac{bK}{p_1 - p_2} (e^{-p_2 t} - e^{-p_1 t}) = 0$ admits only one solution which is $t = 0$;

- there is an inflection point, which can be found as follows:

$$\begin{aligned} y''(t) &= \frac{bK}{p_1 - p_2} (p_1 e^{-p_1 t} - p_2 e^{-p_2 t}) = 0 \\ \Rightarrow \log p_1 - p_1 t &= \log p_2 - p_2 t \\ \Leftrightarrow t &= \frac{p_1 - p_2}{\log p_1 - \log p_2} \end{aligned} \quad (11.54)$$

(11.52) shows that $y(t)$ depends on t only in the argument of the two exponentials, where t appears multiplied by ω_n . This product $\omega_n t$ has no dimensions. This step response can be put as a function of $\omega_n t$, as in Figure 11.7, and then will vary only with ξ as shown.

Frequency response of overdamped system
 $\frac{b}{(s+p_1)(s+p_2)}$

The frequency response in this case can be easily found thanks to the following result:

Theorem 11.1. The gain in dB and the phase of the product of two transfer functions are the sum of their separate gains and phases:

$$20 \log_{10} |G_1(s)G_2(s)| = 20 \log_{10} |G_1(s)| + 20 \log_{10} |G_2(s)| \quad (11.55)$$

$$\angle [G_1(s)G_2(s)] = \angle [G_1(s)] + \angle [G_2(s)] \quad \square \quad (11.56)$$

So we can just add the frequency responses of two first-order systems and obtain the responses in Figure 11.8. Again, it is usual to plot the corresponding asymptotes, shown in Figure 11.9, instead. Notice that:

- For low frequencies, the gain is $20 \log_{10} \frac{b}{p_1 p_2} = 20 \log_{10} \frac{b}{\omega_n^2}$ dB, and the phase is 0° . Indeed, when $\omega \approx 0$, the system is similar to a constant:

$$G(j\omega) = \frac{b}{-\omega^2 + 2\xi\omega_n j\omega + \omega_n^2} \approx \frac{b}{\omega_n^2} \quad (11.57)$$

- For high frequencies, the gain is linear with a slope of -40 dB per decade, and the phase is -180° . Indeed, when $\omega \gg p_1, p_2$, the system is similar to a double integrator:

$$G(j\omega) = \frac{b}{-\omega^2 + 2\xi\omega_n j\omega + \omega_n^2} \approx -\frac{b}{\omega^2} \quad (11.58)$$

- When the two poles do not have the same order of magnitude, it is possible to notice their effect on gain and phase separately. But, if the values of p_1 and p_2 are close to each other, their effects on the frequency response merge.

Step response of critically damped system
 $\frac{b}{(s+\omega_n)^2}$

If both poles of (11.41) are equal, (11.50) shows that $p_1 = p_2 = \omega_n$, and the system's response to a step of amplitude K is, as can be seen in Table 2.1,

$$y(t) = \frac{bK}{\omega_n^2} (1 - e^{-\omega_n t} - \omega_n t e^{-\omega_n t}) \quad (11.59)$$

As expected, should p be negative, (11.59) would diverge to infinity. Assuming that p is positive, then:

- the steady-state response is constant and given by $y_{ss} = \frac{bK}{\omega_n^2}$;
- the response begins at 0 with a horizontal slope, since

$$y'(t) = \frac{bK}{\omega_n^2} (\omega_n e^{-\omega_n t} - \omega_n e^{-\omega_n t} + \omega_n^2 t e^{-\omega_n t}) = bK t e^{-\omega_n t} \quad (11.60)$$

and thus $y'(0) = 0$;

- the response is monotonous, since $y'(t) = bK t e^{-\omega_n t} = 0$ admits only one solution which is $t = 0$;
- there is an inflection point, which can be found as follows:

$$\begin{aligned} y''(t) &= bK (e^{-\omega_n t} - \omega_n t e^{-\omega_n t}) = bK (1 - \omega_n t) e^{-\omega_n t} = 0 \\ \Leftrightarrow t &= \frac{1}{\omega_n} \end{aligned} \quad (11.61)$$

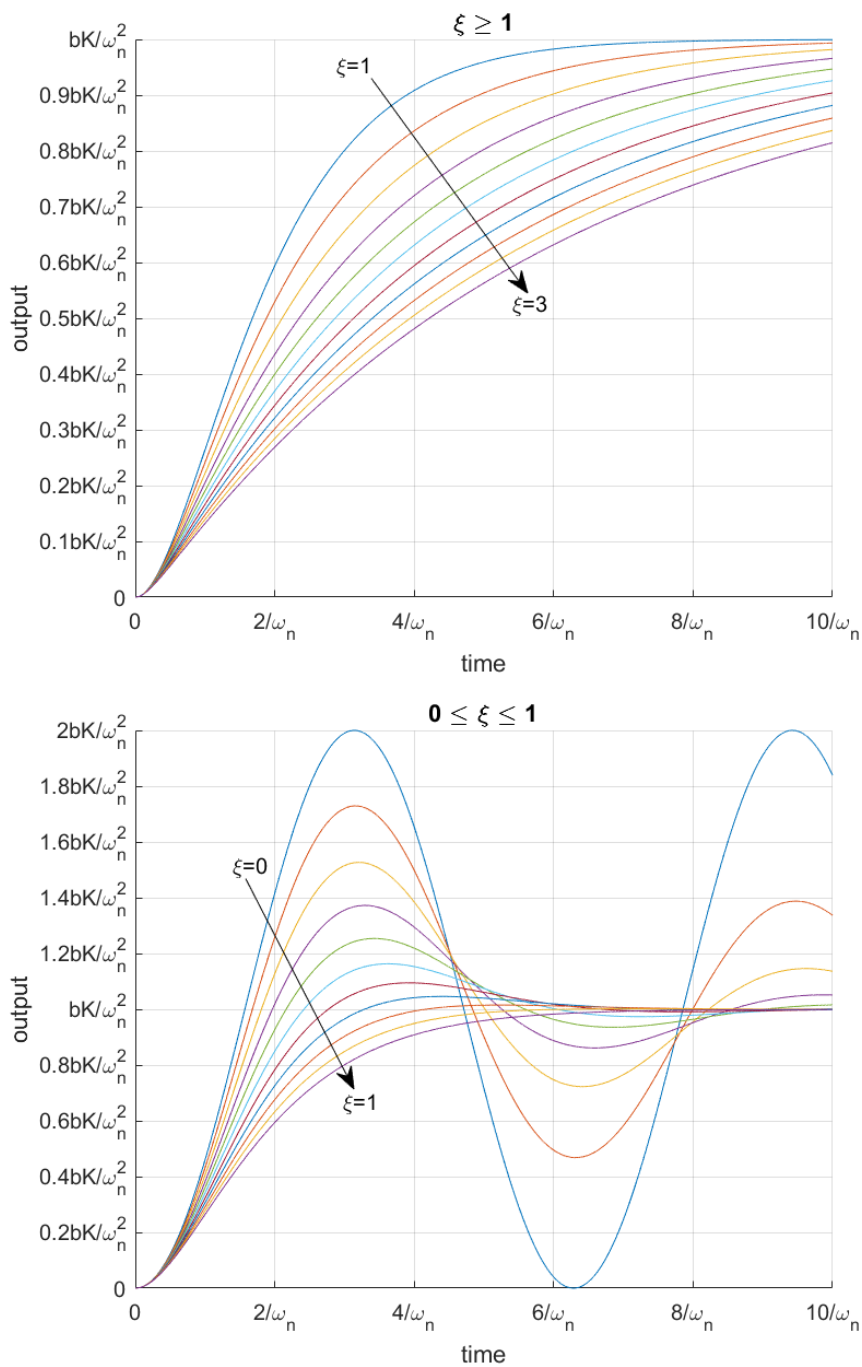


Figure 11.7: Response of $\frac{b}{s^2+2\xi\omega_n s+\omega_n^2}$, $b, \omega_n > 0$ to a step with amplitude $K > 0$. Top: $\xi = 1, 1.2, 1.4, 1.6, 1.8, 2, 2.2, 2.4, 2.6, 2.8, 3$. Bottom: $\xi = 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$.

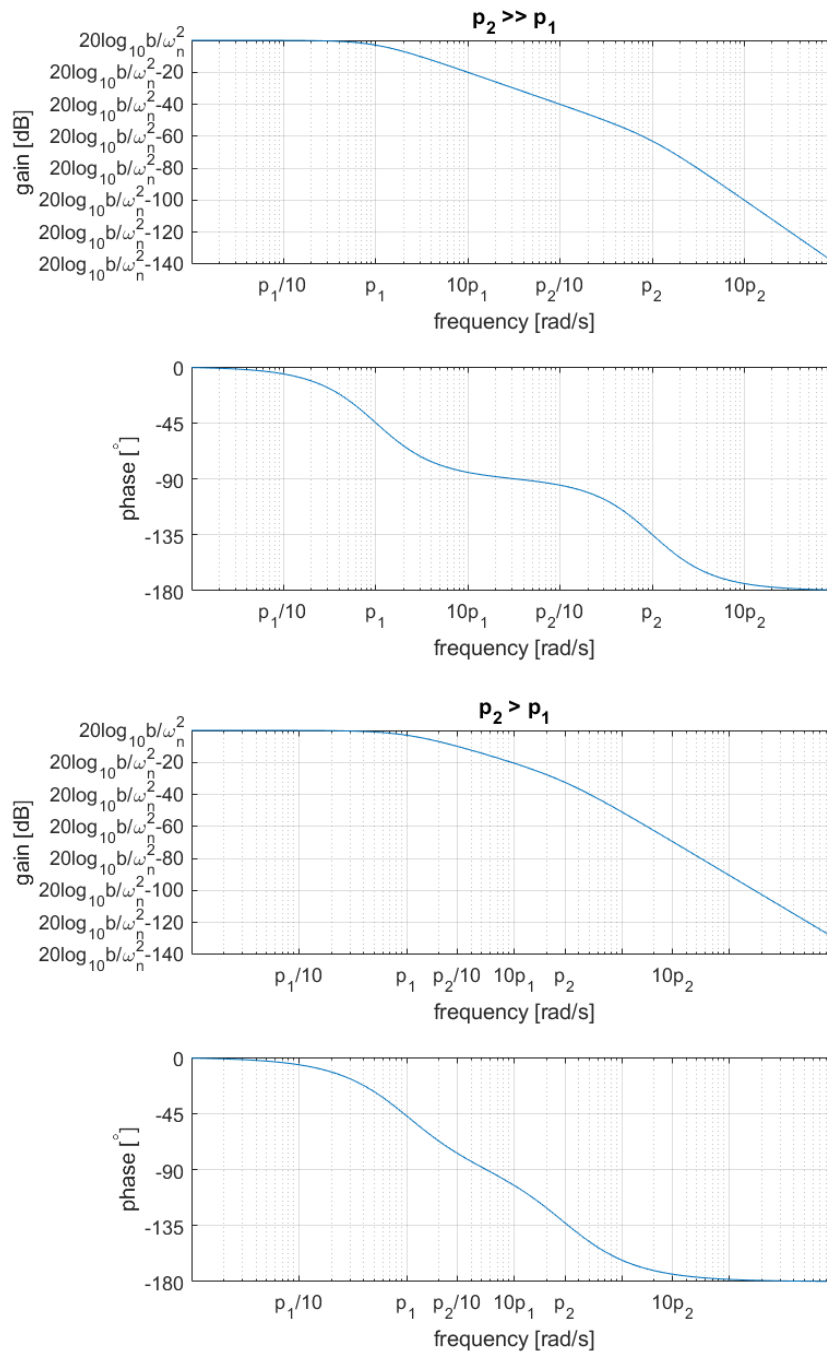


Figure 11.8: Bode diagram of $\frac{b}{s^2 + 2\xi\omega_n s + \omega_n^2}$, $b, \omega_n > 0$, $\xi > 1$, i.e. of $\frac{b}{(s+p_1)(s+p_2)}$, $b, p_1, p_2 > 0$. Top: $\xi \gg 1$. Bottom: $\xi > 1$.

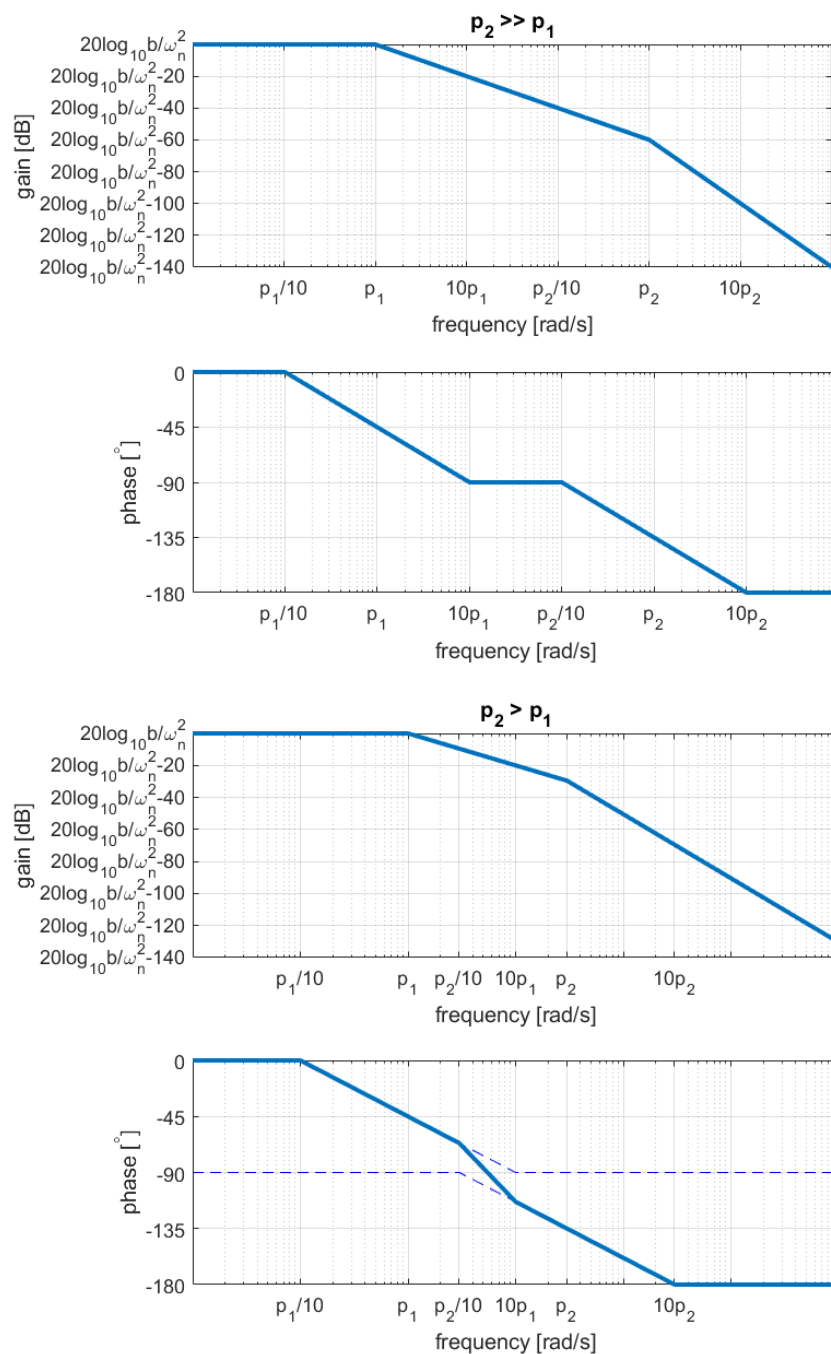


Figure 11.9: Asymptotes of the Bode diagram of $\frac{b}{s^2 + 2\xi\omega_n s + \omega_n^2}$, $b, \omega_n > 0$, $\xi > 1$, i.e. of $\frac{b}{(s+p_1)(s+p_2)}$, $b, p_1, p_2 > 0$, in Figure 11.8. Top: $\xi \gg 1$. Bottom: $\xi > 1$. Notice how in the latter case the approximation is poorer.

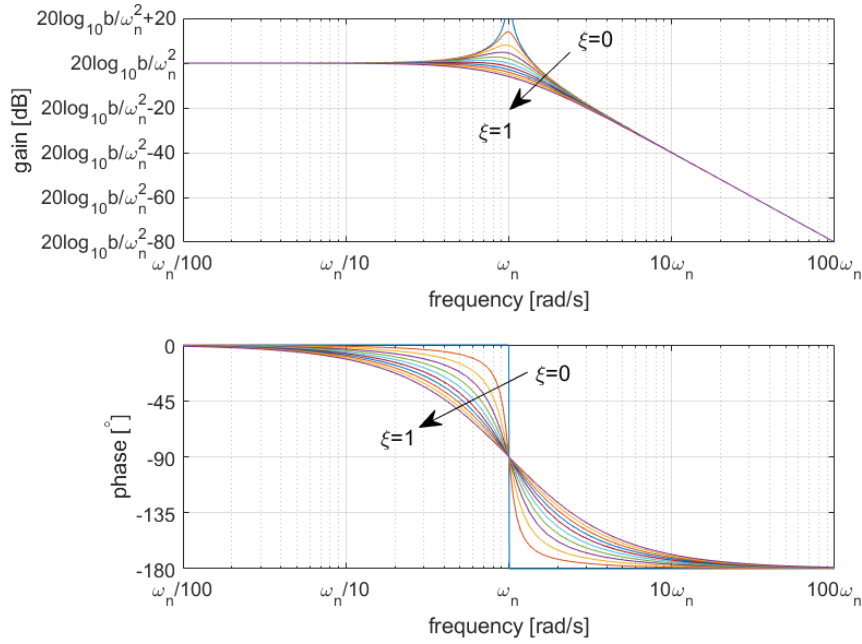


Figure 11.10: Bode diagram of $\frac{b}{s^2 + 2\xi\omega_n s + \omega_n^2}$, $b, \omega_n > 0$, for $\xi = 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$.

Once more, (11.59) shows that $y(t)$, given in Figure 11.7, depends on t only, by product $\omega_n t$.

Frequency response of critically damped system

The frequency response in this case can be found thanks to (11.55)–(11.56) and is shown in Figure 11.10. Again, it is usual to plot instead the corresponding asymptotes, shown in Figure 11.11. Notice that:

- For low and high frequencies, the frequency response is the same as in the overdamped case.
- For $\omega = \omega_n$, the gain and phase are

$$G(j\omega_n) = \frac{b}{(j\omega_n + \omega_n)^2} = \frac{b}{\omega_n^2(j+1)^2} = \frac{b}{2j\omega_n^2} \quad (11.62)$$

$$\begin{aligned} 20 \log_{10} |G(j\omega_n)| &= 20 \log_{10} \frac{b}{\omega_n^2} - 20 \log_{10} 2 \\ &= 20 \log_{10} \frac{b}{\omega_n^2} - 6 \text{ dB} \end{aligned} \quad (11.63)$$

$$\angle G(j\omega_n) = -90^\circ \quad (11.64)$$

Step response of underdamped system

If the poles of (11.41) are complex conjugate, the system's response to a step of amplitude K is, as can be seen in Table 2.1,

$$y(t) = \frac{bK}{\omega_n^2} \left(1 - \frac{1}{\sqrt{1-\xi^2}} e^{-\xi\omega_n t} \sin \left(\omega_n \sqrt{1-\xi^2} t + \arctan \frac{\sqrt{1-\xi^2}}{\xi} \right) \right) \quad (11.65)$$

See Figure 11.7. As already mentioned, should product $-\xi\omega$ be positive (i.e. if either ξ or ω_n are negative), (11.65) would diverge to infinity. Assuming that the system is stable, i.e. that $0 < \xi < 1$, then:

- the steady-state response is again constant and given by $y_{ss} = \frac{bK}{\omega_n^2}$;
- the response has oscillations caused by the sinusoid, and so is not monotonous;
- the frequency of the oscillations, $\omega_n \sqrt{1-\xi^2}$, is called damped frequency (as opposed to natural frequency ω_n);
- the oscillations have decreasing amplitudes thanks to the exponential;

Damped frequency

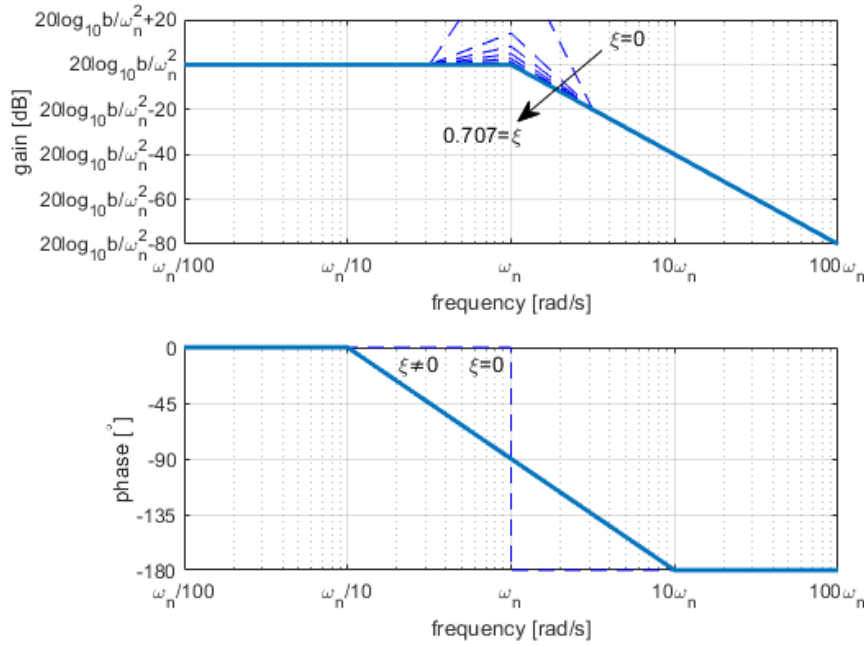


Figure 11.11: Asymptotes of the Bode diagram of $\frac{b}{s^2+2\xi\omega_n s+\omega_n^2}$, $b, \omega_n > 0$, for $\xi = 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6$, and $\frac{\sqrt{2}}{2} \leq \xi \leq 1$.

- indeed, since $-1 \leq \sin x \leq 1$, $\forall x$, the response $y(t)$ is limited by two exponential curves $\bar{y}(t)$ and $\underline{y}(t)$ that converge to y_{ss} (see Figure 11.12):

$$\underline{y}(t) = \frac{bK}{\omega_n^2} \left(1 - \frac{e^{-\xi\omega_n t}}{\sqrt{1-\xi^2}} \right) \leq y(t) \leq \frac{bK}{\omega_n^2} \left(1 + \frac{e^{-\xi\omega_n t}}{\sqrt{1-\xi^2}} \right) = \bar{y}(t) \quad (11.66)$$

- the oscillations take place around the steady-state, which means that the response will exceed that value at some instants: this is called **overshoot** (see Figure 11.12);
- the response begins at 0 with a horizontal slope, since

$$\begin{aligned} y'(t) &= -\frac{bK}{\omega_n^2 \sqrt{1-\xi^2}} \left[-\xi\omega_n e^{-\xi\omega_n t} \sin \left(\omega_n \sqrt{1-\xi^2} t + \arctan \frac{\sqrt{1-\xi^2}}{\xi} \right) \right. \\ &\quad \left. + \omega_n \sqrt{1-\xi^2} e^{-\xi\omega_n t} \cos \left(\omega_n \sqrt{1-\xi^2} t + \arctan \frac{\sqrt{1-\xi^2}}{\xi} \right) \right] \\ &= \frac{bK e^{-\xi\omega_n t}}{\omega_n \sqrt{1-\xi^2}} \left[\xi \sin \left(\omega_n \sqrt{1-\xi^2} t + \arctan \frac{\sqrt{1-\xi^2}}{\xi} \right) \right. \\ &\quad \left. - \sqrt{1-\xi^2} \cos \left(\omega_n \sqrt{1-\xi^2} t + \arctan \frac{\sqrt{1-\xi^2}}{\xi} \right) \right] \quad (11.67) \end{aligned}$$

and thus

$$\begin{aligned} y'(0) &= \frac{bK}{\omega_n \sqrt{1-\xi^2}} \left(\xi \overbrace{\sin \left(\arctan \frac{\sqrt{1-\xi^2}}{\xi} \right)}^{\sin} - \sqrt{1-\xi^2} \overbrace{\cos \left(\arctan \frac{\sqrt{1-\xi^2}}{\xi} \right)}^{\cos} \right) \\ &= \frac{bK}{\omega_n \sqrt{1-\xi^2}} \left(\xi \sqrt{1-\xi^2} - \sqrt{1-\xi^2} \xi \right) = 0 \quad (11.68) \end{aligned}$$

- that the response is not monotonous can also be seen equalling (11.67) to

zero, so as to find the time instants in which (11.65) changes direction:

$$\begin{aligned}
 y'(t) &= 0 & (11.69) \\
 \Leftrightarrow \xi \sin \left(\omega_n \sqrt{1-\xi^2} t + \arctan \frac{\sqrt{1-\xi^2}}{\xi} \right) \\
 -\sqrt{1-\xi^2} \cos \left(\omega_n \sqrt{1-\xi^2} t + \arctan \frac{\sqrt{1-\xi^2}}{\xi} \right) &= 0 \\
 \Leftrightarrow \tan \left(\omega_n \sqrt{1-\xi^2} t + \arctan \frac{\sqrt{1-\xi^2}}{\xi} \right) &= \frac{\sqrt{1-\xi^2}}{\xi} \\
 \Rightarrow \omega_n \sqrt{1-\xi^2} t + \arctan \frac{\sqrt{1-\xi^2}}{\xi} &= \arctan \frac{\sqrt{1-\xi^2}}{\xi} + k\pi \\
 \Leftrightarrow t &= \frac{k\pi}{\omega_n \sqrt{1-\xi^2}}, \quad k \in \mathbb{Z}
 \end{aligned}$$

- because the upper exponential curve in (11.66) limiting the response decreases with time, the first overshoot is the largest, and its maximum value is thus the the **maximum overshoot**;

Maximum overshoot

- the value M_p of this maximum overshoot is usually given as a percentage of the steady-state value:

$$M_p = \frac{\max y(t) - y_{ss}}{y_{ss}} \Leftrightarrow \max y(t) = y_{ss}(1 + M_p) \quad (11.70)$$

Peak time

- the time instant at which the maximum overshoot takes place is the **peak time** t_p , which is consequently given by (11.69) for $k = 1$:

$$t_p = \frac{\pi}{\omega_n \sqrt{1-\xi^2}} \quad (11.71)$$

- M_p can be found replacing (11.71) in (11.65) and then using (11.70):

$$\begin{aligned}
 y(t_p) &= \frac{bK}{\omega_n^2} \left(1 - \frac{1}{\sqrt{1-\xi^2}} e^{-\xi\omega_n \frac{\pi}{\omega_n \sqrt{1-\xi^2}}} \sin \left(\omega_n \sqrt{1-\xi^2} \frac{\pi}{\omega_n \sqrt{1-\xi^2}} + \arctan \frac{\sqrt{1-\xi^2}}{\xi} \right) \right) \\
 &= \frac{bK}{\omega_n^2} \left(1 - \frac{1}{\sqrt{1-\xi^2}} e^{\frac{-\xi\pi}{\sqrt{1-\xi^2}}} \underbrace{\sin \left(\pi + \arctan \frac{\sqrt{1-\xi^2}}{\xi} \right)}_{\substack{\sin \\ \cos \\ \sin \arctan \sqrt{1-\xi^2}/\xi}} \right) \\
 &= \frac{bK}{\omega_n^2} \left(1 + e^{\frac{-\xi\pi}{\sqrt{1-\xi^2}}} \right) \\
 \Rightarrow M_p &= \frac{\frac{bK}{\omega_n^2} \left(1 + e^{\frac{-\xi\pi}{\sqrt{1-\xi^2}}} \right) - \frac{bK}{\omega_n^2}}{\frac{bK}{\omega_n^2}} = e^{\frac{-\xi\pi}{\sqrt{1-\xi^2}}} \quad (11.72)
 \end{aligned}$$

Settling time

- (11.65) could be used to find settling times for different percentages of the steady-state value, but the sinusoid makes calculations difficult (the response is not monotonous) and numerical results have abrupt variations with ξ ; it is easier to obtain approximated (by excess) settling times from the (monotonous) limiting exponential curves in (11.66), e.g. $\bar{y}(t)$:

$$\begin{aligned}
 \bar{y}(t_{s,x\%}) &= y_{ss} \left(1 + \frac{x}{100} \right) \\
 \Leftrightarrow \frac{bK}{\omega_n^2} \left(1 + \frac{e^{-\xi\omega_n t}}{\sqrt{1-\xi^2}} \right) &= \left(1 + \frac{x}{100} \right) \frac{bK}{\omega_n^2} \\
 \Leftrightarrow e^{-\xi\omega_n t} &= \frac{x}{100} \sqrt{1-\xi^2} \\
 \Rightarrow t &= \frac{-\log \left(\frac{x}{100} \sqrt{1-\xi^2} \right)}{\xi\omega_n} \leq \frac{-\log \frac{x}{100}}{\xi\omega_n} \quad (11.73)
 \end{aligned}$$

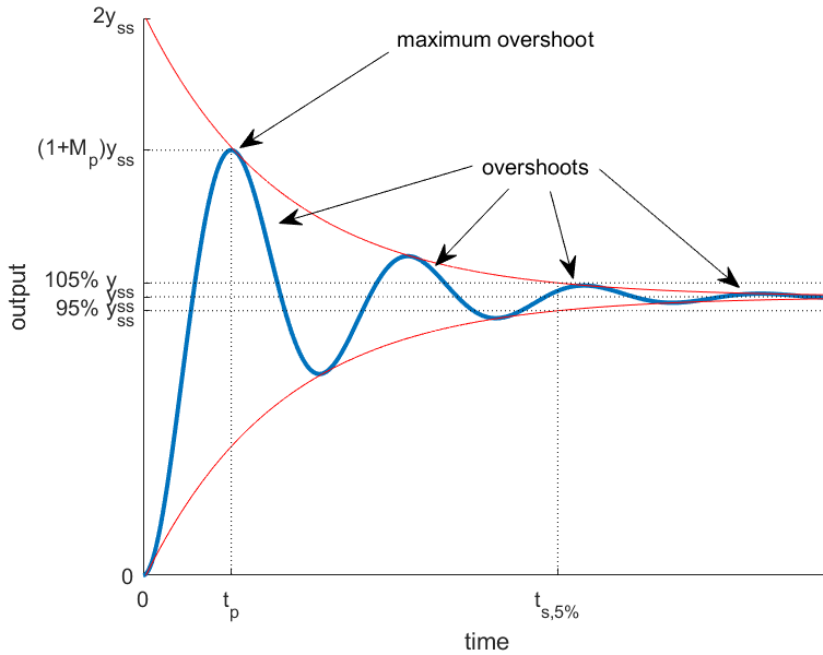


Figure 11.12: Overshoots, maximum overshoot M_p , peak time t_p , 5% settling time $t_{s,5\%}$ approximated from the envelope exponentials.

and thus, replacing x with different values, we find that (compare these values with those of a first order system without poles):

- the 10% settling time is $t_{s,10\%} = \frac{2.3}{\xi\omega_n}$;
- the 5% settling time is $t_{s,5\%} = \frac{3}{\xi\omega_n}$;
- the 2% settling time is $t_{s,2\%} = \frac{4}{\xi\omega_n}$;
- the 1% settling time is $t_{s,1\%} = \frac{4.6}{\xi\omega_n}$.

The frequency response in this case cannot be found from that of first order transfer functions using Theorem 11.1. We can write

Frequency response of underdamped system

$$\frac{b}{s^2 + 2\xi\omega_n s + \omega_n^2}$$

$$G(j\omega) = \frac{b}{-\omega^2 + 2\xi\omega_n j\omega + \omega_n^2} \quad (11.74)$$

$$|G(j\omega)| = \frac{b}{\sqrt{(\omega_n^2 - \omega^2)^2 + 4\xi^2\omega_n^2\omega^2}} \quad (11.75)$$

$$20 \log_{10} |G(j\omega)| = 20 \log_{10} b - 20 \log_{10} \sqrt{(\omega_n^2 - \omega^2)^2 + 4\xi^2\omega_n^2\omega^2} \quad (11.76)$$

$$\angle G(j\omega) = -\arctan \frac{2\xi\omega_n\omega}{\omega_n^2 - \omega^2} \quad (11.77)$$

and see that for high and low frequencies the frequency response is the same as in the overdamped and critically damped cases. But this time the gain (11.75) may not be monotonous. To find this, because numerator b is constant and the square root is monotonous, we only need to calculate the derivative

$$\frac{d}{d\omega} [(\omega_n^2 - \omega^2)^2 + 4\xi^2\omega_n^2\omega^2] = 2(\omega_n^2 - \omega^2)(-2\omega) + 8\xi^2\omega_n^2\omega \quad (11.78)$$

Equalling to zero we find that the gain has a maximum at a frequency called *Resonance frequency* **resonance frequency** ω_r given by

$$\begin{aligned} \omega_r^2 - \omega_n^2 + 2\xi^2\omega_n^2 &= 0 \\ \Leftrightarrow \omega_r^2 &= \omega_n^2 - 2\xi^2\omega_n^2 \\ \Rightarrow \omega_r &= \omega_n \sqrt{1 - 2\xi^2} \end{aligned} \quad (11.79)$$

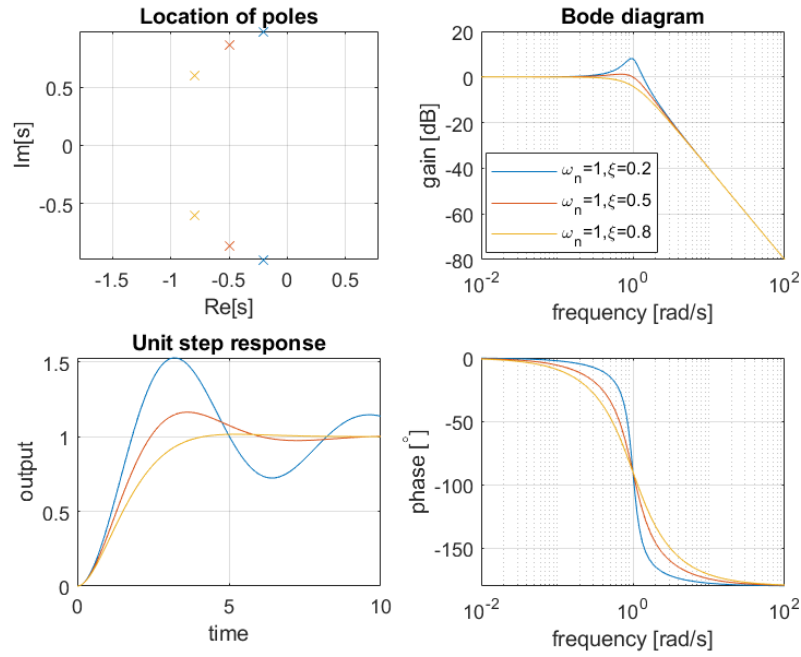


Figure 11.13: Pole location in the complex plane, unit step responses and Bode diagrams of three transfer functions given by $\frac{\omega_n^2}{s^2 + 2\xi\omega_n s + \omega_n^2}$, when ω_n is constant.

as long as the square root is real, i.e. $1 - 2\xi^2 > 0 \Leftrightarrow \xi < \frac{1}{\sqrt{2}} \approx 0.707$. In this case, the corresponding maximum value of the gain is found from (11.76):

$$\begin{aligned}
 20 \log_{10} |G(j\omega_r)| &= 20 \log_{10} b - 20 \log_{10} \sqrt{(\omega_n^2 - \omega_n^2(1 - 2\xi^2))^2 + 4\xi^2\omega_n^4(1 - 2\xi^2)} \\
 &= 20 \log_{10} b - 20 \log_{10} \sqrt{\omega_n^4(1 - 1 + 2\xi^2)^2 + \omega_n^4(4\xi^2 - 8\xi^4)} \\
 &= 20 \log_{10} b - 20 \log_{10} \omega_n^2 \sqrt{4\xi^4 + 4\xi^2 - 8\xi^4} \\
 &= \underbrace{20 \log_{10} b - 20 \log_{10} \omega_n^2}_{\text{gain at low frequencies}} - \underbrace{20 \log_{10} 2\xi \sqrt{1 - \xi^2}}_{>0} \quad (11.80)
 \end{aligned}$$

(Notice that since $0 < \xi < \frac{\sqrt{2}}{2}$ we have $2\xi\sqrt{1 - \xi^2} < \sqrt{2}\sqrt{1 - \frac{1}{2}} = 1$.) This frequency behaviour is shown in Figure 11.10, and the corresponding asymptotes in Figure 11.11, where the **resonant peak** of $-20 \log_{10} 2\xi\sqrt{1 - \xi^2}$ dB is marked.

Step response of undamped system $\frac{b}{s^2 + \omega_n^2}$

For $\xi = 0$, system (11.41) is marginally stable, and step response (11.65) simplifies to

$$y(t) = \frac{bK}{\omega_n^2} \left(1 - \sin \left(\omega_n t + \frac{\pi}{2} \right) \right) \quad (11.81)$$

Why ω_n is the natural frequency

Why ξ is the damping coefficient

and thus the steady-state response consists in oscillations with frequency ω_n that are not damped; that is why ω_n is called natural frequency: it is the frequency of the system's step response (or impulse response; check Table 2.1 again) oscillations when the damping coefficient is 0. The reason why ξ is the damping coefficient, by the way, is that it is proportional to the coefficient of the damper in (4.9); more generically, it is related to the coefficient of the energy dissipator, as can be seen e.g. in (5.24).

Frequency response of undamped system $\frac{b}{s^2 + \omega_n^2}$

The corresponding frequency response is given by (11.74) with $\xi = 0$:

$$G(j\omega) = \frac{b}{\omega_n^2 - \omega^2} \quad (11.82)$$

This is always a real number, and so the phase jumps from 0° (when $\omega < \omega_n$) to -180° (when $\omega > \omega_n$) as seen in Figure 11.10. At resonance frequency $\omega_r = \omega_n\sqrt{1 - 2\xi^2} = \omega_n$, the peak is

$$\lim_{\omega \rightarrow \omega_r} |G(j\omega_r)| = \lim_{\omega \rightarrow \omega_n} |G(j\omega_n)| = \lim_{\omega \rightarrow \omega_n} \frac{b}{\sqrt{(\omega_n^2 - \omega^2)^2}} = +\infty \quad (11.83)$$

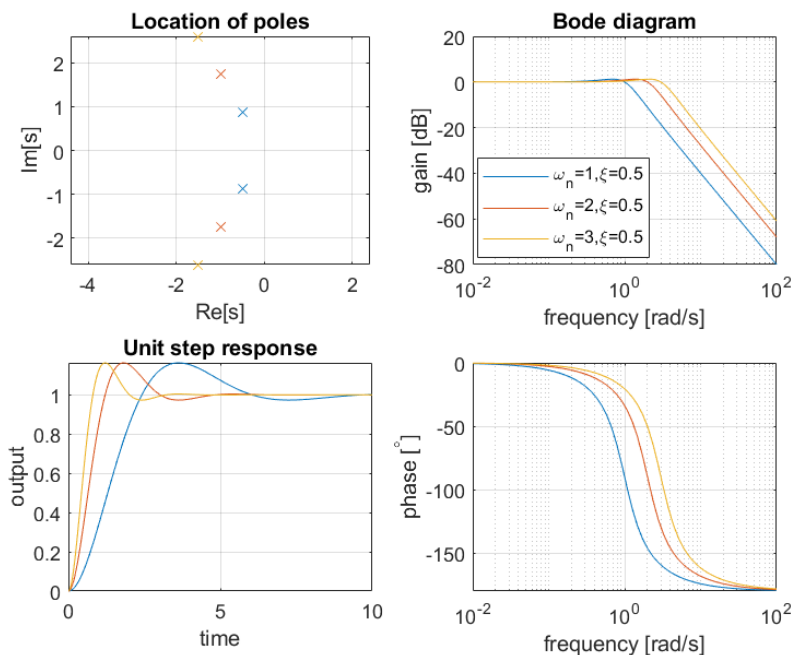


Figure 11.14: Pole location in the complex plane, unit step responses and Bode diagrams of three transfer functions given by $\frac{\omega_n^2}{s^2 + 2\xi\omega_n s + \omega_n^2}$, when ξ is constant.

Of course, in practice systems have some residual damping, and in any case sinusoidal outputs never have infinite amplitudes (reread Example 4.1 and Remark 10.7 if you need).

We can now sum up what happens for the four different cases of (11.41) we have studied (check Figure 11.6 again):

- step responses begin at 0 with a horizontal slope;
- step responses have a $\frac{bK}{\omega_n^2}$ steady-state (where K is the amplitude of the step);
- step responses have no overshoot if $\xi \geq 1$, and have a $M_p = e^{\frac{-\xi\pi}{\sqrt{1-\xi^2}}}$ overshoot if $0 \leq \xi < 1$;
- frequency responses have a phase that goes from 0° (at low frequencies) to -180° (at high frequencies) as the frequency increases;
- frequency responses have a gain that goes from $\frac{b}{\omega_n^2}$ (i.e. $20 \log_{10} \frac{b}{\omega_n^2}$ dB, at low frequencies) to 0 (i.e. $-\infty$ dB, at high frequencies) as the frequency increases;
- the slope of the gain is -40 dB per decade at high frequencies;
- the gain has no resonant peak if $\xi \geq \frac{\sqrt{2}}{2}$, and has a $-20 \log_{10} 2\xi\sqrt{1-\xi^2} > 0$ resonant peak if $0 \leq \xi < \frac{\sqrt{2}}{2}$ (so in this case, for a sinusoidal input with a frequency equal to or around ω_r , the system's sinusoidal steady-state output will have an amplitude larger than the input's). *Meaning of the resonant peak*

Figures 11.13–11.16 illustrate how step and frequency responses of (11.42) with complex conjugate poles change according to their position on the complex plane.

To conclude this section, the fifth and final case of a second order transfer function to be considered is that in which either pole is 0: (11.49) cannot be applied; (11.41) will integrate the output of the other pole. *Step response of $\frac{b}{s(s+a)}$*

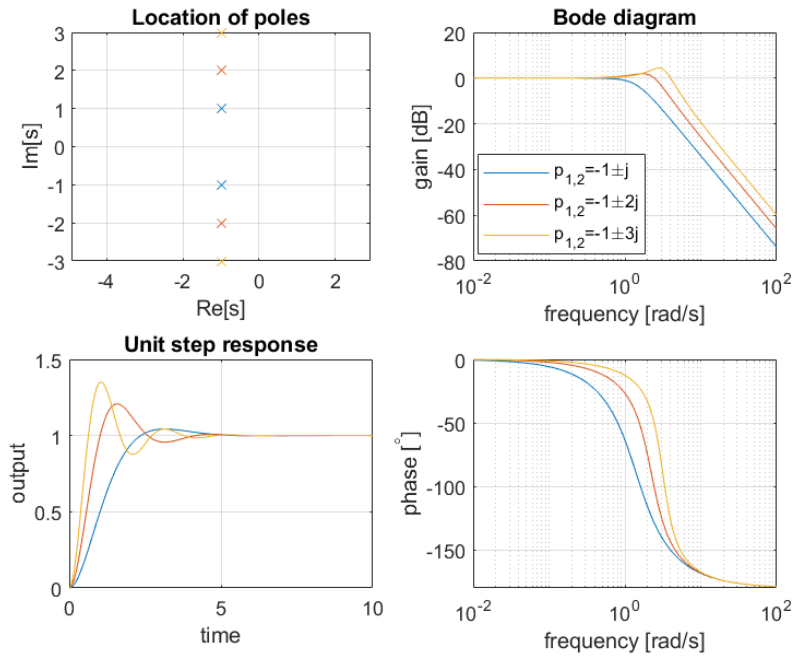


Figure 11.15: Pole location in the complex plane, unit step responses and Bode diagrams of three transfer functions given by $\frac{p_1 p_2}{(s+p_1)(s+p_2)}$, when the real part of the poles is constant.

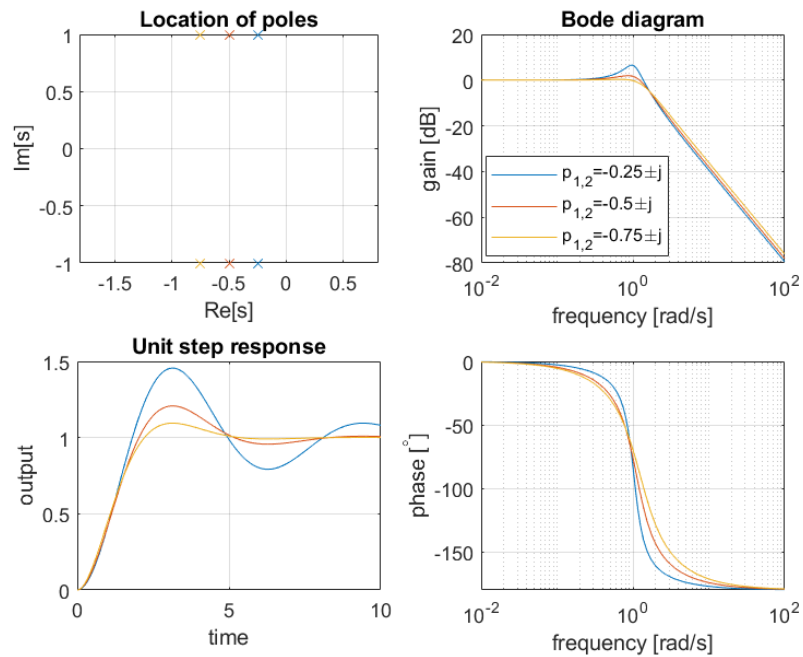


Figure 11.16: Pole location in the complex plane, unit step responses and Bode diagrams of three transfer functions given by $\frac{p_1 p_2}{(s+p_1)(s+p_2)}$, when the imaginary part of the poles is constant.

11.4 Systems with more zeros and poles: frequency responses

The frequency response of systems with more than two poles, or with zeros, or with a negative gain, can be found from the frequency response of

- each real pole and zero,
- each pair of complex conjugate poles and zeros, and
- its gain

found separately. To find the frequency response of any transfer function, write it as a product of smaller transfer functions found therein, and sum the corresponding gains and phases, according to Theorem 11.1.

Remark 11.3. It is often easier to write those smaller transfer functions so that they have a 0 dB gain at low frequencies (excepting poles or zeros at the origin, of course, which have no constant gain at low frequencies), and then add the effect of a gain as needed. This is how we will proceed in what follows. But it is possible to have different low-frequency gains for the different smaller transfer functions. \square

We already know the following:

- the frequency response of a pole at the origin $\frac{1}{s}$ was studied in section 11.1:

$$G(j\omega) = \frac{1}{j\omega} \quad (11.84)$$

$$20 \log_{10} |G(j\omega)| = -20 \log_{10} \omega \text{ dB} \quad (11.85)$$

$$\angle G(j\omega) = -90^\circ \quad (11.86)$$

- the frequency response of a real pole in the left complex half-plane $\frac{a}{s+a}$, $a > 0$ was studied in section 11.2:

$$G(j\omega) = \frac{a}{j\omega + a} \quad (11.87)$$

$$20 \log_{10} |G(j\omega)| \approx 0 \text{ dB}, \omega \ll a \quad (11.88)$$

$$\angle G(j\omega) \approx 0^\circ, \omega \ll a \quad (11.89)$$

$$20 \log_{10} |G(ja)| = 20 \log_{10} \frac{a}{\sqrt{a^2 + a^2}} = 20 \log_{10} \frac{1}{\sqrt{2}} = -3 \text{ dB} \quad (11.90)$$

$$\angle G(ja) = 0^\circ - \arctan \frac{a}{a} = -45^\circ \quad (11.91)$$

$$20 \log_{10} |G(j\omega)| \approx 20 \log_{10} a - 20 \log_{10} \omega \text{ dB}, \omega \gg a \quad (11.92)$$

$$\angle G(j\omega) \approx -90^\circ, \omega \gg a \quad (11.93)$$

- the frequency response of a real pole in the right complex half-plane $\frac{a}{s-a}$, $a > 0$ was also studied in section 11.2:

$$G(j\omega) = \frac{a}{j\omega - a} \quad (11.94)$$

$$20 \log_{10} |G(j\omega)| \approx 0 \text{ dB}, \omega \ll a \quad (11.95)$$

$$\angle G(j\omega) \approx -180^\circ, \omega \ll a \quad (11.96)$$

$$20 \log_{10} |G(ja)| = 20 \log_{10} \frac{a}{\sqrt{a^2 + a^2}} = 20 \log_{10} \frac{1}{\sqrt{2}} = -3 \text{ dB} \quad (11.97)$$

$$\angle G(ja) = 0^\circ - \arctan \frac{a}{-a} = -135^\circ \quad (11.98)$$

$$20 \log_{10} |G(j\omega)| \approx 20 \log_{10} a - 20 \log_{10} \omega \text{ dB}, \omega \gg a \quad (11.99)$$

$$\angle G(j\omega) \approx -90^\circ, \omega \gg a \quad (11.100)$$

- the frequency response of a pair of complex conjugate poles in the left complex half-plane $\frac{\omega_n^2}{s^2 + 2\xi\omega_n s + \omega_n^2}$, $\omega_n > 0, 0 \leq \xi < 1$ was studied in sec-

tion 11.3:

$$G(j\omega) = \frac{\omega_n^2}{\omega_n^2 - \omega^2 + j2\xi\omega_n\omega} \quad (11.101)$$

$$20 \log_{10} |G(j\omega)| \approx 0 \text{ dB}, \quad \omega \ll \omega_n \quad (11.102)$$

$$\angle G(j\omega) \approx 0^\circ, \quad \omega \ll \omega_n \quad (11.103)$$

$$20 \log_{10} |G(j\omega_n \sqrt{1 - 2\xi^2})| = -20 \log_{10} 2\xi \sqrt{1 - \xi^2} \text{ dB} > 0 \text{ dB} \quad (11.104)$$

$$20 \log_{10} |G(j\omega_n)| = -20 \log_{10} 2\xi \text{ dB} \quad (11.105)$$

$$\angle G(j\omega_n) = \angle \frac{1}{j2\xi} = -90^\circ \quad (11.106)$$

$$20 \log_{10} |G(j\omega)| \approx 20 \log_{10} \omega_n^2 - 40 \log_{10} \omega \text{ dB}, \quad \omega \gg \omega_n \quad (11.107)$$

$$\angle G(j\omega) \approx -180^\circ, \quad \omega \gg \omega_n \quad (11.108)$$

We will now find the following frequency responses:

- A positive gain $k > 0$:

$$G(j\omega) = k \quad (11.109)$$

$$20 \log_{10} |G(j\omega)| = 20 \log_{10} k \text{ dB} \quad (11.110)$$

$$\angle G(j\omega) = 0^\circ \quad (11.111)$$

- A negative gain $k > 0$:

$$G(j\omega) = -k \quad (11.112)$$

$$20 \log_{10} |G(j\omega)| = 20 \log_{10} |k| \text{ dB} \quad (11.113)$$

$$\angle G(j\omega) = -180^\circ \quad (11.114)$$

- A zero at the origin s :

$$G(j\omega) = j\omega \quad (11.115)$$

$$20 \log_{10} |G(j\omega)| = 20 \log_{10} \omega \text{ dB} \quad (11.116)$$

$$\angle G(j\omega) = 90^\circ \quad (11.117)$$

- A zero in the left complex half-plane $\frac{s+b}{b}$, $b > 0$:

$$G(j\omega) = \frac{j\omega + b}{b} \quad (11.118)$$

$$20 \log_{10} |G(j\omega)| \approx 0 \text{ dB}, \quad \omega \ll b \quad (11.119)$$

$$\angle G(j\omega) \approx 0^\circ, \quad \omega \ll b \quad (11.120)$$

$$20 \log_{10} |G(jb)| = 20 \log_{10} \frac{\sqrt{b^2 + b^2}}{b} = 20 \log_{10} \sqrt{2} = 3 \text{ dB} \quad (11.121)$$

$$\angle G(jb) = \arctan \frac{b}{b} = 45^\circ \quad (11.122)$$

$$20 \log_{10} |G(j\omega)| \approx 20 \log_{10} \omega - 20 \log_{10} b \text{ dB}, \quad \omega \gg b \quad (11.123)$$

$$\angle G(j\omega) \approx 90^\circ, \quad \omega \gg b \quad (11.124)$$

- A zero in the right complex half-plane $\frac{s-b}{b}$, $b > 0$:

$$G(j\omega) = \frac{j\omega - b}{b} \quad (11.125)$$

$$20 \log_{10} |G(j\omega)| \approx 0 \text{ dB}, \quad \omega \ll b \quad (11.126)$$

$$\angle G(j\omega) \approx 180^\circ, \quad \omega \ll b \quad (11.127)$$

$$20 \log_{10} |G(jb)| = 20 \log_{10} \frac{\sqrt{b^2 + b^2}}{b} = 20 \log_{10} \sqrt{2} = 3 \text{ dB} \quad (11.128)$$

$$\angle G(jb) = \arctan \frac{b}{-b} = 135^\circ \quad (11.129)$$

$$20 \log_{10} |G(j\omega)| \approx 20 \log_{10} \omega - 20 \log_{10} b \text{ dB}, \quad \omega \gg b \quad (11.130)$$

$$\angle G(j\omega) \approx 90^\circ, \quad \omega \gg b \quad (11.131)$$

- A pair of complex conjugate poles in the right complex half-plane $\frac{\omega_n^2}{s^2 - 2\xi\omega_n s + \omega_n^2}$, $\omega_n > 0$, $0 \leq \xi < 1$:

$$G(j\omega) = \frac{\omega_n^2}{\omega_n^2 - \omega^2 - j2\xi\omega_n\omega} \quad (11.132)$$

$$20 \log_{10} |G(j\omega)| \approx 0 \text{ dB}, \quad \omega \ll \omega_n \quad (11.133)$$

$$\angle G(j\omega) \approx 0^\circ \equiv -360^\circ, \quad \omega \ll \omega_n \quad (11.134)$$

$$20 \log_{10} |G(j\omega_n \sqrt{1 - 2\xi^2})| = -20 \log_{10} 2\xi \sqrt{1 - \xi^2} \text{ dB} > 0 \text{ dB} \quad (11.135)$$

$$20 \log_{10} |G(j\omega_n)| = -20 \log_{10} 2\xi \text{ dB} \quad (11.136)$$

$$\angle G(j\omega_n) = \angle \frac{1}{-j2\xi} = -270^\circ \quad (11.137)$$

$$20 \log_{10} |G(j\omega)| \approx 20 \log_{10} \omega_n^2 - 40 \log_{10} \omega \text{ dB}, \quad \omega \gg \omega_n \quad (11.138)$$

$$\angle G(j\omega) \approx -180^\circ \equiv -180^\circ, \quad \omega \gg \omega_n \quad (11.139)$$

(Notice that the phase goes up from -360° to -180° , or from 0° to 180° .)

- A pair of complex conjugate zeros in the left complex half-plane $\frac{s^2 + 2\xi\omega_n s + \omega_n^2}{\omega_n^2}$, $\omega_n > 0$, $0 \leq \xi < 1$:

$$G(j\omega) = \frac{\omega_n^2 - \omega^2 + j2\xi\omega_n\omega}{\omega_n^2} \quad (11.140)$$

$$20 \log_{10} |G(j\omega)| \approx 0 \text{ dB}, \quad \omega \ll \omega_n \quad (11.141)$$

$$\angle G(j\omega) \approx 0^\circ, \quad \omega \ll \omega_n \quad (11.142)$$

$$20 \log_{10} |G(j\omega_n \sqrt{1 - 2\xi^2})| = 20 \log_{10} 2\xi \sqrt{1 - \xi^2} \text{ dB} < 0 \text{ dB} \quad (11.143)$$

$$20 \log_{10} |G(j\omega_n)| = 20 \log_{10} 2\xi \text{ dB} \quad (11.144)$$

$$\angle G(j\omega_n) = \angle j2\xi = 90^\circ \quad (11.145)$$

$$20 \log_{10} |G(j\omega)| \approx 40 \log_{10} \omega - 20 \log_{10} \omega_n^2 \text{ dB}, \quad \omega \gg \omega_n \quad (11.146)$$

$$\angle G(j\omega) \approx 180^\circ, \quad \omega \gg \omega_n \quad (11.147)$$

- A pair of complex conjugate zeros in the right complex half-plane $\frac{s^2 - 2\xi\omega_n s + \omega_n^2}{\omega_n^2}$, $\omega_n > 0$, $\xi \geq 0$:

$$G(j\omega) = \frac{\omega_n^2 - \omega^2 - j2\xi\omega_n\omega}{\omega_n^2} \quad (11.148)$$

$$20 \log_{10} |G(j\omega)| \approx 0 \text{ dB}, \quad \omega \ll \omega_n \quad (11.149)$$

$$\angle G(j\omega) \approx 0^\circ, \quad \omega \ll \omega_n \quad (11.150)$$

$$20 \log_{10} |G(j\omega_n \sqrt{1 - 2\xi^2})| = 20 \log_{10} 2\xi \sqrt{1 - \xi^2} \text{ dB} < 0 \text{ dB} \quad (11.151)$$

$$20 \log_{10} |G(j\omega_n)| = 20 \log_{10} 2\xi \text{ dB} \quad (11.152)$$

$$\angle G(j\omega_n) = \angle -j2\xi = -90^\circ \quad (11.153)$$

$$20 \log_{10} |G(j\omega)| \approx 40 \log_{10} \omega - 20 \log_{10} \omega_n^2 \text{ dB}, \quad \omega \gg \omega_n \quad (11.154)$$

$$\angle G(j\omega) \approx -180^\circ, \quad \omega \gg \omega_n \quad (11.155)$$

All the corresponding frequency responses are summed up in Figures 11.17–11.19.

Example 11.1. Find the Bode diagram of

$$G(s) = \frac{s + 100}{s(s^2 + 0.5s + 1)} = \underbrace{\frac{100}{s}}_{G_1(s)} \times \underbrace{\frac{1}{s}}_{G_2(s)} \times \underbrace{\frac{s + 100}{100}}_{G_3(s)} \times \underbrace{\frac{1}{s^2 + 0.5s + 1}}_{G_4(s)} \quad (11.156)$$

First plot the asymptotes of the Bode diagrams of the four transfer function $G_1(s)$ to $G_4(s)$, in Figure 11.20. Add them to obtain the asymptotes of the Bode diagram of $G(s)$ in the same Figure, which also shows the actual Bode diagram. The asymptotes are seen to be a rather fair approximation, especially when the resonance peak given by $-20 \log_{10}(0.5\sqrt{1 - 0.25^2}) = 6.3 \text{ dB}$ is added. \square

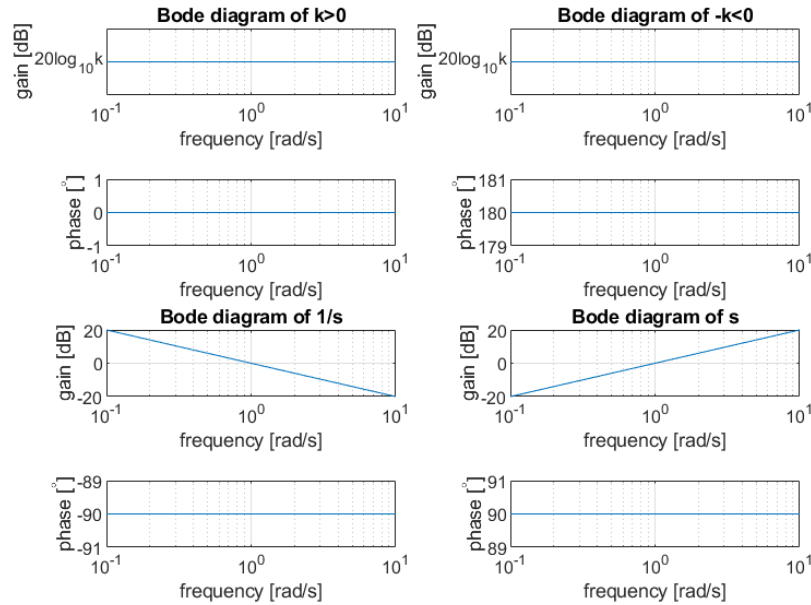


Figure 11.17: Bode diagrams of k , $-k$, $\frac{1}{s}$ and s , $k > 0$.

Remark 11.4. Since the phase is determined up to shifts of 360° , there are different ways of choosing which particular values are used. In all of them the phase is, of course, continuous with frequency ω whenever possible.

- Choose the low frequency phase value to be closest to zero. In this way a low frequency value of, say, -90° is preferred to 270° , no matter what. This criterion presents no reason to choose among 180° and -180° , when input and output are in phase opposition at low frequencies.
- Choose the low frequency phase value that better shows the number of zeros and poles at the origin. In this way, transfer function $\frac{s^3 N(s)}{D(s)}$, where polynomials $N(s)$ and $D(s)$ have no roots at the origin, will have a low frequency phase beginning at 270° and not at -90° , since a zero pushes the phase up and so this phase shows that the three zeros are responsible for the low frequency behaviour. The phase of transfer function $\frac{N(s)}{sD(s)}$ would begin at -90° for a similar reason. This criterion may or may not give the same result as the one above. It allows choosing among low frequency phases of 180° and -180° if there are two poles or two zeros at the origin, but not if there is a negative gain.
- Choose the closest values to zero in the entire frequency range of concern, as long as the phase is continuous with ω . \square

Example 11.2. Find the Bode diagram of

$$G(s) = \frac{-s + 5}{s^2(s + 10)} = \underbrace{-\frac{1}{2}}_{G_1(s)} \times \underbrace{\frac{1}{s^2}}_{G_2(s)} \times \underbrace{\frac{10}{s + 10}}_{G_3(s)} \times \underbrace{\frac{s - 5}{5}}_{G_4(s)} \quad (11.157)$$

First plot the asymptotes of the Bode diagrams of the four transfer function $G_1(s)$ to $G_4(s)$, in Figure 11.21. Add them to obtain the asymptotes of the Bode diagram of $G(s)$ in the same Figure, which also shows the actual Bode diagram. \square

Remark 11.5. Given the way Bode diagrams can be built from smaller transfer functions, this can be used to identify a model for a plant from its frequency response. The Bode diagram is split into the sum of several Bode diagrams corresponding to frequency responses in Figures 11.17–11.19; the model will be the product of the respective transfer functions. \square

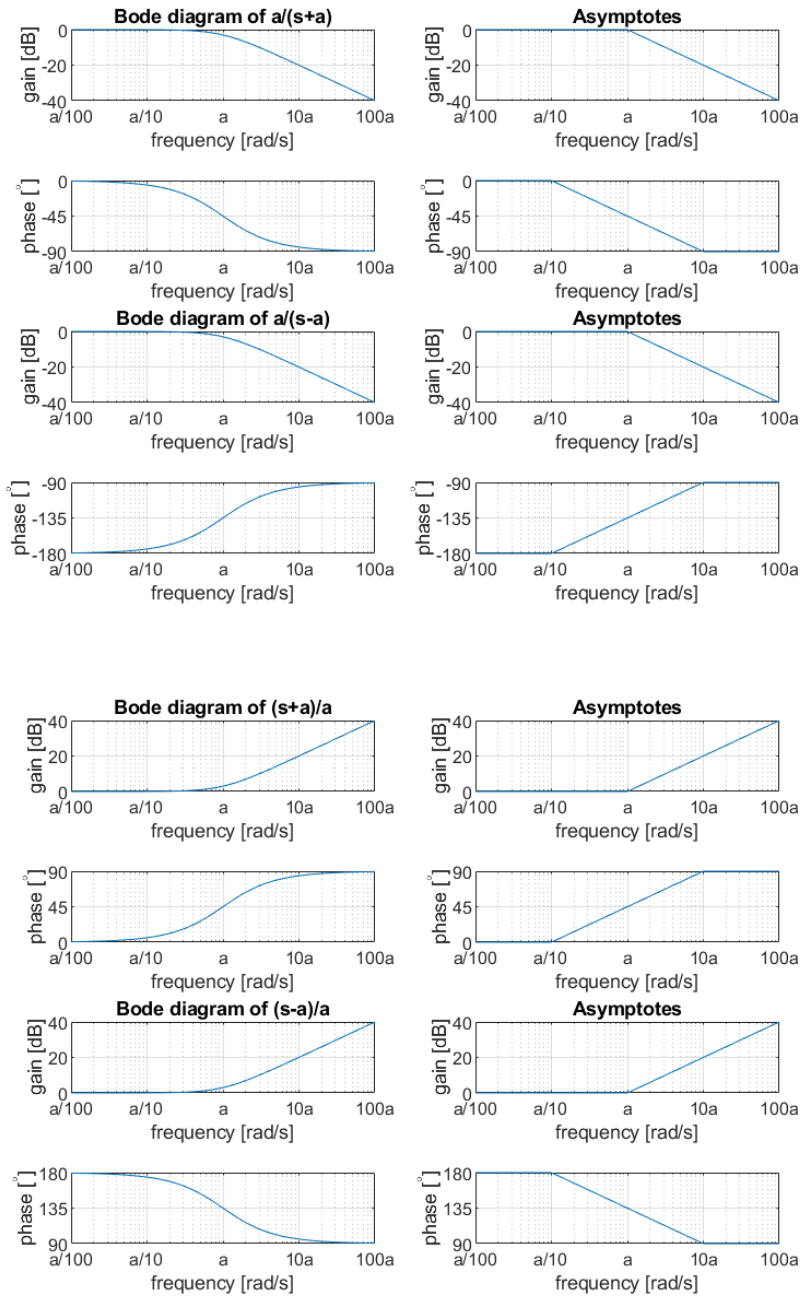


Figure 11.18: Bode diagrams (left) and corresponding asymptotes (right) of $\frac{a}{s+a}$, $\frac{a}{s-a}$, $\frac{s+a}{a}$ and $\frac{s-a}{a}$, $a > 0$.

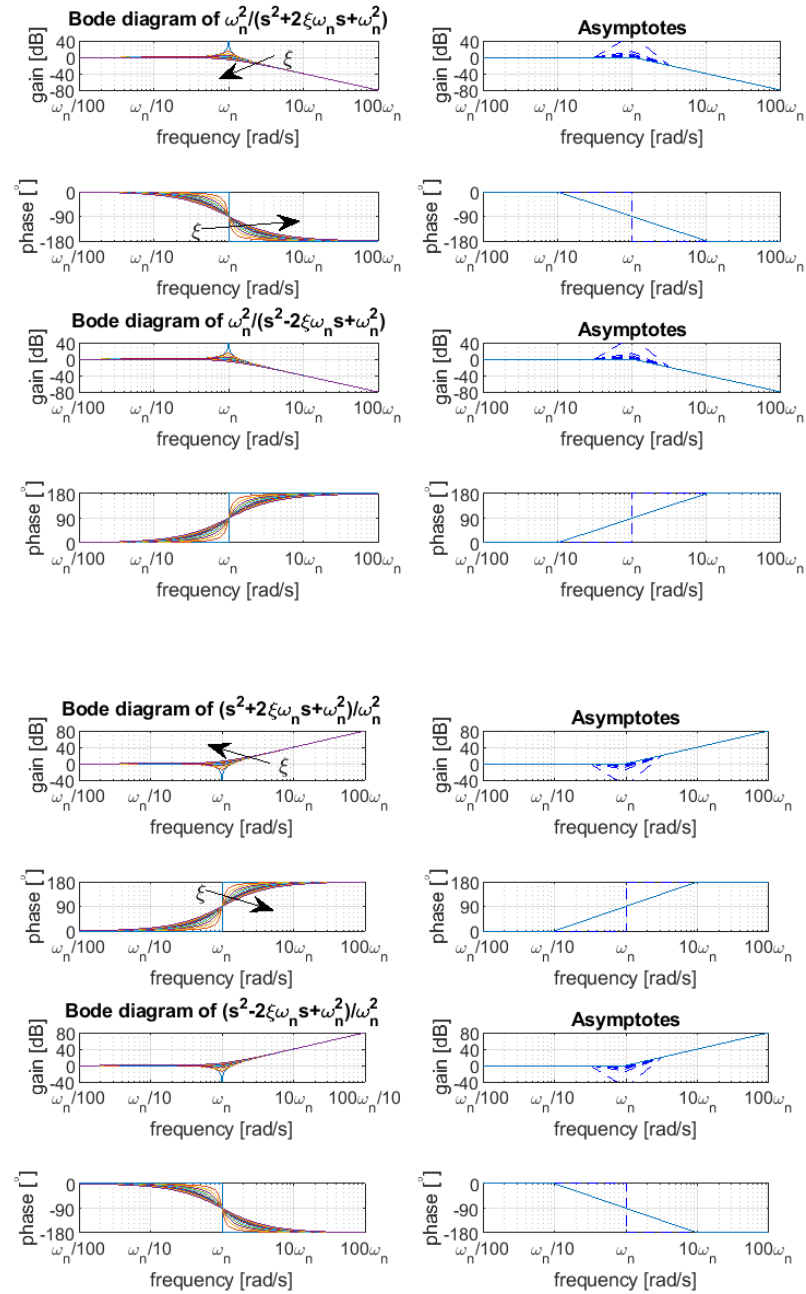


Figure 11.19: Bode diagrams (left) and corresponding asymptotes (right) of

$$\frac{\omega_n^2}{s^2 + 2\xi\omega_n s + \omega_n^2}, \frac{\omega_n^2}{s^2 - 2\xi\omega_n s + \omega_n^2}, \frac{s^2 + 2\xi\omega_n s + \omega_n^2}{\omega_n^2} \text{ and } \frac{s^2 - 2\xi\omega_n s + \omega_n^2}{\omega_n^2}, \omega_n > 0, \xi \geq 0.$$

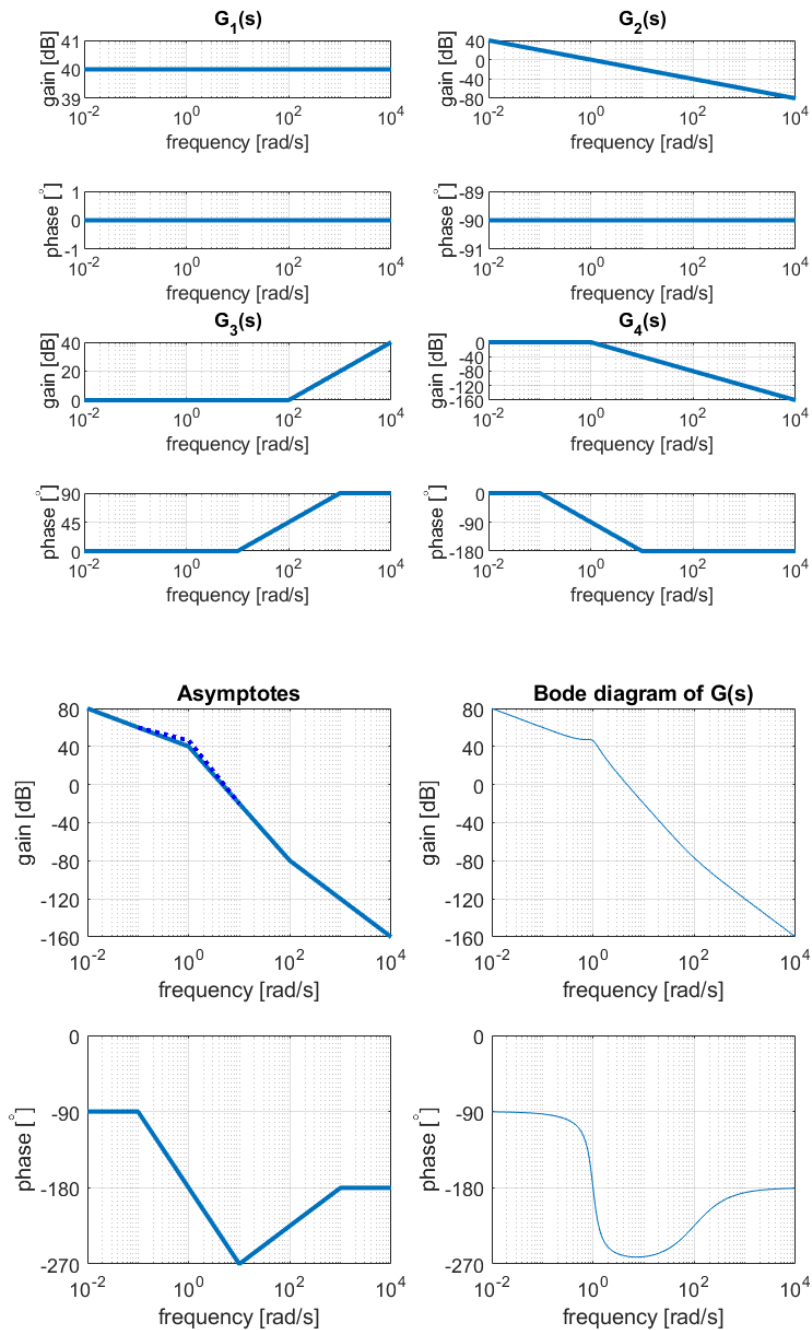


Figure 11.20: Building the Bode diagram of (11.156) from Example 11.1.

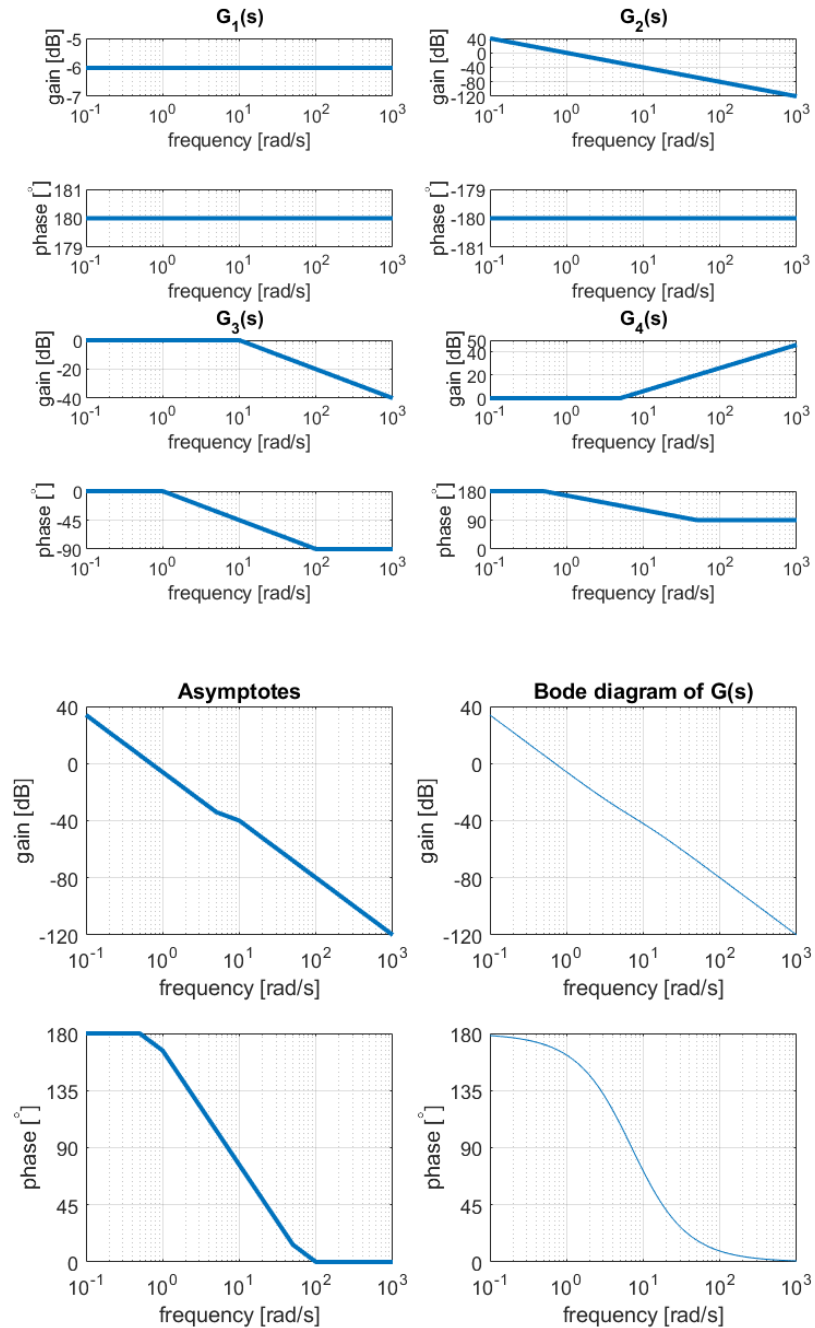


Figure 11.21: Building the Bode diagram of (11.157) from Example 11.2.

Example 11.3. Given either Bode diagram of Examples 11.1 or 11.2, first draw its asymptotes, as shown in Figures 11.20 or 11.21; then divide them into the separate asymptotes, also shown in the Figures, corresponding to transfer functions from Figures 11.17–11.19. The product of these transfer functions is the desired transfer function, either (11.156) or (11.157). \square

Definition 11.2. A zero on the left complex half-plane (i.e. a zero with a negative real part) is a **minimum phase zero**. A zero on the right complex half-plane (i.e. a zero with a positive real part) is a **non-minimum phase zero**. \square

Minimum and non-minimum phase zeros

Remark 11.6. The reason for these names can be seen in Figure 11.18: the phase of $\frac{s+a}{a}$, $a > 0$ is between 0° and 90° , while the phase of $\frac{s-a}{a}$, $a > 0$ is between 180° and 90° . The latter is thus, for low frequencies and in average, larger than the former. \square

A system's frequency response at high frequencies depends on the difference between the number of poles n and the number of zeros m . More precisely, it is clear from Figures 11.17–11.19 that, at high frequencies, each pole will contribute to the gain with a slope of -20 dB/decade, and each zero will contribute to the gain with a slope of 20 dB/decade. Consequently, the slope of the gain in a Bode diagram will be $20(m-n)$ dB per decade. So, as the frequency increases:

Gain slope at high frequencies

- the gain of strictly proper transfer functions goes to $-\infty$ dB, i.e. to 0 in absolute value;
- the gain of transfer functions with the same number of poles and zeros goes to a constant value (in both dB and absolute value);
- the gain of transfer functions which are not proper goes to $+\infty$ dB, i.e. to $+\infty$ in absolute value.

In the latter case, if the transfer function is stable, this means that inputs with oscillations of smaller and smaller periods correspond to steady-state output oscillations with larger and larger amplitudes (and with the same small period of the input). The velocity with which the output would be changing would become infinite. This is clearly impossible in practice, as it would require an arbitrarily large energy. (For instance, in the case of an output which is a position, there would a frequency above which this position would be changing faster than light speed.) The same happens even in the case of a constant gain for high frequencies, since the amplitude may not be increasing, but the period decreases. The only models which are physically possible are those with more poles than zeros, i.e. strictly proper models. That is why most models are strictly proper. Of course, all models are only valid for a range of parameters. So we can use models with as many zeros as poles, or more zeros than poles, being aware that for frequencies high enough they cannot be valid: there have to be unmodelled poles changing the behaviour of the plant so that it does not require an impossibly infinite energy.

Why transfer functions should be strictly proper

On the other hand, the slope of the gain at (very) low frequencies depends on the number of zeros or poles at the origin: it is clear from Figures 11.17–11.19 only they cause a slope at such frequencies:

Gain slope at low frequencies

- n poles at the origin cause a low frequency gain slope of $-20n$ dB/decade;
- m zeros at the origin cause a low frequency gain slope of $20m$ dB/decade.

Example 11.4. Consider the Bode diagram of Example 11.1 in Figure 11.20. At low frequencies, the gain has a -20 dB/decade slope. Consequently, there has to be a pole at the origin. At high frequencies, the gain has a -40 dB/decade slope. Consequently, the number of poles exceeds the number of zeros by 2 (we could have 0 zeros and 2 poles, or 1 zero and 3 poles, or 2 zeros and 4 poles, etc.). \square

Example 11.5. Consider the Bode diagram of Example 11.2 in Figure 11.21. At low frequencies, the gain has a -40 dB/decade slope. Consequently, there have to be 2 poles at the origin. At high frequencies, the gain has a -40 dB/decade slope. Consequently, the number of poles exceeds the number of zeros by 2. \square

Remark 11.7. Notice that the gain of a stable system at low frequencies is constant and equal to its static gain:

Stable system's gain at low frequencies

$$\lim_{\omega \rightarrow 0} G(j\omega) = \lim_{\omega \rightarrow 0} \frac{b_0 + b_1s + b_2s^2 + b_3s^3 + \dots}{\underbrace{a_0 + a_1s + a_2s^2 + a_3s^3 + \dots}_{G(s)}} \Bigg|_{s=j\omega} = \frac{b_0}{a_0} \quad (11.158)$$

Type of a transfer function

Definition 11.3. The **type** of a transfer function is its number of poles at the origin. \square

Example 11.6. Here are transfer functions of type 0, type 1, type 2 and type 3:

$$G_0(s) = \frac{1}{s + 1} \quad (11.159)$$

$$G_1(s) = \frac{1}{s(s + 1)} \quad (11.160)$$

$$G_2(s) = \frac{1}{s^2(s + 1)} \quad (11.161)$$

$$G_3(s) = \frac{1}{s^3(s + 1)} \quad (11.162)$$

11.5 Systems with more zeros and poles: stability

We know that a stable plant's poles have real negative parts. If there are only one or two poles, this is not difficult to verify. Things change if there are three or more poles, since finding the roots of polynomials of such orders is not trivial. There are of course numerical algorithms to do so efficiently. But there is a way of knowing if all the roots of a polynomial are on the left complex half-plane without having to compute them: the **Routh-Hurwitz criterion**. It even lets us know how many poles there are on the right complex half-plane. Demonstrating the validity of this criterion is not trivial and we will not do that. We will only present it together with examples of application.

Routh-Hurwitz criterion

Consider a polynomial in s given by

$$p(s) = \sum_{k=0}^n a_k s^k = a_n s^n + a_{n-1} s^{n-1} + \dots + a_1 s + a_0 \quad (11.163)$$

For each of its n roots, we want to know if it has positive real parts, negative real parts, or lies on the imaginary axis.

- If $a_0 = 0$, there is a root at the origin. Divide the polynomial by s and start again.
- If not all the a_k have the same sign, there is at least one root with positive real part. If all the a_k are negative, we can multiply $p(s)$ by -1 , which does not change its roots, and all coefficients become positive; so we can say instead that all the a_k must be positive if all the roots are to have negative real parts. This is a necessary condition, but it is not sufficient (it may happen that all the a_k are positive and still some roots have positive real parts).
- The number of roots with positive real parts is equal to the number of sign changes of the first column of the Routh-Hurwitz table, which has n lines, $\lceil \frac{n}{2} \rceil$ columns, and is built as follows:

Routh-Hurwitz table

s^n	a_n	a_{n-2}	a_{n-4}	\dots	0
s^{n-1}	a_{n-1}	a_{n-3}	a_{n-5}	\dots	0
s^{n-2}	b_1	b_2	\dots		
s^{n-3}	c_1	c_2	\dots		
\vdots	\vdots	\vdots			
s^2	d_1	d_2			
s^1	e_1				
s^0	f_1				

(11.164)

Here,

$$b_1 = \frac{a_{n-1}a_{n-2} - a_{n-3}a_n}{a_{n-1}} = a_{n-2} - \frac{a_{n-3}a_n}{a_{n-1}} \tag{11.165}$$

$$b_2 = \frac{a_{n-1}a_{n-4} - a_{n-5}a_n}{a_{n-1}} = a_{n-4} - \frac{a_{n-5}a_n}{a_{n-1}} \tag{11.166}$$

$$\vdots \tag{11.167}$$

$$c_1 = \frac{b_1a_{n-3} - b_2a_{n-1}}{b_1} = a_{n-3} - \frac{b_2a_{n-1}}{b_1} \tag{11.168}$$

$$c_2 = \frac{b_1a_{n-5} - b_3a_{n-1}}{b_1} = a_{n-5} - \frac{b_3a_{n-1}}{b_1} \tag{11.169}$$

$$\vdots \tag{11.170}$$

This pattern goes on in each line until all further elements are necessarily zero, and goes down until all n lines are filled in.

Example 11.7. Consider a transfer function given by $\frac{N(s)}{s^3+6s^2+11s+6}$. We build the Routh-Hurwitz table

s^3	1	11	(11.171)
s^2	6	6	
s	$\frac{6 \times 11 - 6 \times 1}{6} = 10$		
1	$\frac{10 \times 6 - 0 \times 6}{10} = 6$		

and verify that all elements in the first column are positive. So the transfer function is stable (whatever the numerator may be, since, remember, zeros have nothing to do with stability). □

If there is a zero in the left column, replace it with a vanishing ε . If when $\varepsilon \rightarrow 0^+$ all the coefficients below are positive, there is a pair of complex conjugate pure imaginary poles (which are marginally stable), and the other poles are stable. If some are negative, the transfer function is unstable.

Zero in the left column

Example 11.8. Consider a transfer function given by $\frac{N(s)}{s^3+s^2+4s+4}$. We build the Routh-Hurwitz table

s^3	1	4	(11.172)
s^2	1	4	
s	$\frac{1 \times 4 - 4 \times 1}{\varepsilon} = \emptyset$		
1	$\frac{\varepsilon \times 4 - 0 \times 1}{\varepsilon} = 4$		

and verify that when $\varepsilon \rightarrow 0^+$ the first column is entirely positive. So there are no unstable poles. The zero is caused by a pair of complex conjugate poles on the imaginary axis: $s^3 + s^2 + 4s + 4 = (s + 1)(s + 2j)(s - 2j)$. □

Example 11.9. Consider a transfer function given by $\frac{N(s)}{s^6+4s^5+9s^4+13s^3-17s-10}$. Since there are both positive and negative coefficients in the denominator, we can tell right away that the transfer function is not stable. (Remember that if they were all positive the transfer function might still be unstable.) To know how many of the 6 poles are unstable, we build the Routh-Hurwitz table

s^6	1	9	0	-10	(11.173)
s^5	4	13	-17		
s^4	$9 - \frac{13}{4} = \frac{23}{4}$		$\frac{17}{4}$	-10	
s^3	$13 - \frac{17}{\frac{23}{4}} = \frac{231}{23}$		$-17 + \frac{40}{\frac{23}{4}} = -\frac{231}{23}$		
s^2	$\frac{\frac{231}{23} \cdot \frac{17}{4} + \frac{231}{23} \cdot \frac{23}{4}}{\frac{231}{23}} = 10$		-10		
s	$\frac{10(-\frac{231}{23}) + 10 \cdot \frac{231}{23}}{10} = \emptyset$				
1	-10				

and verify that when $\varepsilon \rightarrow 0^+$ there is one sign change in the first column, from line s to line 1. So there are five stable poles, and one unstable pole (in fact $s^6 + 4s^5 + 9s^4 + 13s^3 - 17s - 10 = (s - 1)(s + 1)^2(s + 2)(s^2 + s + 5)$). □

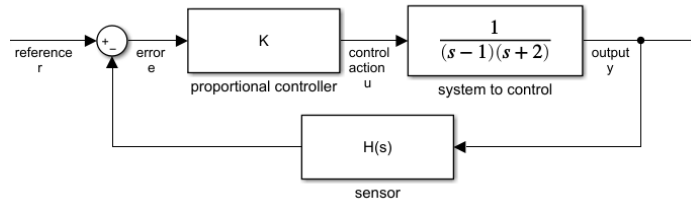


Figure 11.22: Control system from Examples 11.12 and 11.13.

Example 11.10. Consider a transfer function given by $\frac{N(s)}{s^3 - 12s + 16}$. We build the Routh-Hurwitz table

$$\begin{array}{c|cc}
 s^3 & 1 & -12 \\
 s^2 & \epsilon & 16 \\
 \hline
 s & -12 - \frac{16}{\epsilon} & \\
 1 & 16 &
 \end{array} \quad (11.174)$$

and verify that when $\epsilon \rightarrow 0^+$ the element below in the first column is negative. So there are two sign changes, and two unstable poles (in fact $s^3 - 12s + 16 = (s - 2)^2(s + 4)$). \square

Remark 11.8. Notice that coefficient a_0 always turns up in the last line of the table. If it does not, you got your calculations wrong. \square

Line of zeros

If there is a whole line of zeros, build a polynomial with the coefficients of the line above, differentiate it, and use the coefficients of the derivative to replace the zeros.

Example 11.11. Consider a transfer function given by $\frac{N(s)}{s^6 - 4s^4 + 4s^2 - 16}$. We begin the Routh-Hurwitz table

$$\begin{array}{c|cccc}
 s^6 & 1 & -4 & 4 & -16 \\
 s^5 & 0 & 0 & 0 &
 \end{array} \quad (11.175)$$

and verify that the second line only has zeros. So we look at the line above, which is line s^6 , and use the coefficients of that line with powers s^6 , s^4 , s^2 and s^0 to obtain $\frac{d}{ds}(s^6 - 4s^4 + 4s^2 - 16) = 6s^5 - 16s^3 + 8s$.

$$\begin{array}{c|cccc}
 s^6 & 1 & -4 & 4 & -16 \\
 s^5 & 6 & -16 & 8 & \\
 \hline
 s^4 & -\frac{4}{3} & \frac{8}{3} & -16 & \\
 s^3 & -4 & -64 & & \\
 s^2 & 24 & -16 & & \\
 s & -\frac{200}{3} & & & \\
 1 & -16 & & &
 \end{array} \quad (11.176)$$

There are three sign changes, and so three unstable poles (in fact $s^6 - 4s^4 + 4s^2 - 16 = (s - 1 + j)(s - 1 - j)(s + 1 + j)(s - 1 - j)(s + 2)(s - 2)$). \square

The Routh-Hurwitz criterion is important not only because it allows finding out by hand whether or not a system is stable, but above all because it lets us find analytical conditions for the stability of a closed-loop that depends on a parameter.

Example 11.12. Plant

$$\frac{y(s)}{u(s)} = \frac{1}{(s - 1)(s + 2)} \quad (11.177)$$

is controlled in closed loop with a proportional controller K , as seen in Figure 11.22. We first assume that the sensor is perfect, i.e. $H(s) = 1$. To know what values of K ensure that the closed loop is stable, we find the closed loop transfer function

$$\frac{y(s)}{r(s)} = \frac{\frac{K}{s^2 + s - 2}}{1 + \frac{K}{s^2 + s - 2}} = \frac{K}{s^2 + s + K - 2} \quad (11.178)$$

and build the corresponding Routh-Hurwitz table

$$\begin{array}{c|cc} s^2 & 1 & K-2 \\ s & 1 & \\ \hline 1 & K-2 & \end{array} \quad (11.179)$$

from which we see that all coefficients are positive in the left column if

$$K-2 > 0 \Leftrightarrow K > 2 \quad \square \quad (11.180)$$

Example 11.13. If in the last example we make $K = 10$ and then find out that the sensor has dynamics given by $H(s) = \frac{a}{s+a}$, we want to know what values of pole a still let the closed loop be stable. So the closed loop transfer function becomes

$$\frac{y(s)}{r(s)} = \frac{\frac{10}{s^2+s-2}}{1 + \frac{10a}{(s^2+s-2)(s+a)}} = \frac{10}{s^3 + (a+1)s^2 + (a-2)s + 8a} \quad (11.181)$$

From the Routh-Hurwitz table

$$\begin{array}{c|ccc} s^3 & 1 & a-2 & \\ s^2 & a+1 & 8a & \\ s & \frac{(a+1)(a-2)-8a}{a+1} = \frac{a^2-9a+2}{a+1} & & \\ \hline 1 & 8a & & \end{array} \quad (11.182)$$

we see that all coefficients are positive in the left column if

$$\begin{cases} a+1 > 0 \\ \frac{a^2-9a+2}{a+1} > 0 \\ 8a > 0 \end{cases} \Rightarrow \begin{cases} a > -1 \\ \frac{(a+0.22)(a-9.22)}{a+1} > 0 \\ a > 0 \end{cases} \quad (11.183)$$

For the second condition,

$a^2 - 9a + 2$		-0.22		-1		9.22	
$a + 1$	+	0	-	0	+	0	+
	-	0	+	∞	-	0	+

(11.184)

and so we conclude that the system is stable if $a > 9.22$. \square

Remark 11.9. In the example above, it would obviously have been wiser to choose the sensor first, and only then the proportional controller K . And there are surely other requirements for the control system other than ensuring stability; we will address them in Part IV. \square

Example 11.14. Plant

$$G(s) = \frac{1.5s^2 + 22.5s + 66}{s^2 - 8s - 9} \quad (11.185)$$

is controlled in closed loop with a proportional controller K and a perfect sensor. To know what values of K ensure that the closed loop is stable, we find the closed loop transfer function

$$\begin{aligned} \frac{y(s)}{r(s)} &= \frac{K \frac{1.5s^2 + 22.5s + 66}{s^2 - 8s - 9}}{1 + K \frac{1.5s^2 + 22.5s + 66}{s^2 - 8s - 9}} \\ &= \frac{K(1.5s^2 + 22.5s + 66)}{s^2 - 8s - 9 + K(1.5s^2 + 22.5s + 66)} \\ &= \frac{1.5Ks^2 + 22.5Ks + 66K}{(1.5K + 1)s^2 + (22.5K - 8)s + (66K - 9)} \end{aligned} \quad (11.186)$$

and build the Routh-Hurwitz table

$$\begin{array}{c|cc} s^2 & 1.5K + 1 & 66K - 9 \\ s & 22.5K - 8 & \\ \hline 1 & 66K - 9 & \end{array} \quad (11.187)$$

We are tempted to write

$$\begin{cases} 1 + 1.5K > 0 \\ 22.5K - 8 > 0 \\ 66K - 9 > 0 \end{cases} \Leftrightarrow \begin{cases} K > -\frac{2}{3} = -0.6667 \\ K > \frac{8}{22.5} = 0.3556 \\ K > \frac{9}{66} = 0.1364 \end{cases} \Rightarrow K > 0.3556 \quad (11.188)$$

and conclude that only these values of K make the closed loop stable. But if we make $K = -1$ in (11.186) we get

$$\frac{y(s)}{r(s)} = \frac{-1.5s^2 - 22.5s - 66}{-0.5s^2 - 30.5s - 75} = \frac{3s^2 + 45s + 132}{s^2 + 61s + 150} \quad (11.189)$$

which has poles in -58.4330 and -2.5670 and is thus stable. Why is this so? Because the number of unstable poles is the number of sign changes in the first column of the Routh-Hurwitz table; if they are all negative, there are no sign changes, and no unstable poles. Since $G(s)$ is not strictly proper, the entire first column depends on K , while in all previous examples there was always a positive number in that column, meaning that all others had to be positive too for all poles to be stable. In this case, we can let the entire column be negative, i.e.

$$\begin{cases} 1 + 1.5K < 0 \\ 22.5K - 8 < 0 \\ 66K - 9 < 0 \end{cases} \Leftrightarrow \begin{cases} K < -\frac{2}{3} = -0.6667 \\ K < \frac{8}{22.5} = 0.3556 \\ K < \frac{9}{66} = 0.1364 \end{cases} \Rightarrow K < -0.6667 \quad (11.190)$$

and thereby conclude that if $K \in]-\infty, -0.6667[\cup]0.3556, +\infty[$ the closed loop is stable. \square

11.6 Systems with more zeros and poles: time responses

We know from the proof of Theorem 10.2, or from

$$\mathcal{L}^{-1} \left[\frac{1}{s+a} \right] = e^{-at} \quad (11.191)$$

together with a reasoning similar to that of Remark 2.6, that each pole of a transfer function will originate an exponential in its time response:

Effects of the real part of a pole

- if the pole is stable, its (negative) real part corresponds to how fast this part of the time response is vanishing;

The imaginary parts of poles cause oscillations

- if there is an imaginary part, it will correspond (together with that of the pole's complex conjugate) to oscillations in time (which, the pole being stable, are vanishing).

Consequently,

Faster poles

- the more negative the real part of a stable pole is, the faster its effect in the time responses vanishes;

Slower poles

- the closer a stable pole is to the imaginary axis, the slower its effect in the time responses vanishes.

Meaning of the residues

When a transfer function is written as a partial fraction expansion, the residues show the weight that each pole will have in the time responses. Slower poles, even when weighted with a residue which is relatively small when compared with the others, have a lasting effect in time responses just for being slow, i.e. for the slowly vanishing effect of their contribution, and are thus called **dominant poles**.

Dominant poles

When a transfer function has only one dominant pole, or one pair of complex conjugate dominant poles, clearly far from the other ones, its time response will be mostly determined by that pole or poles. This is not so when there are several dominant poles, or, rather, when there are none, since no pole has an effect clearly dominating that of the others.

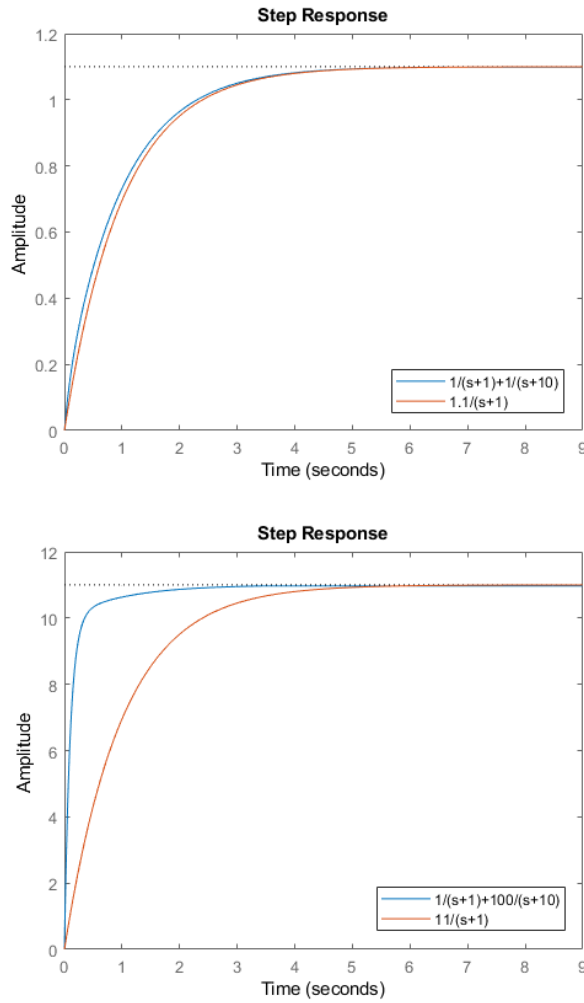


Figure 11.23: Unit step responses of the transfer functions of Example 11.15.

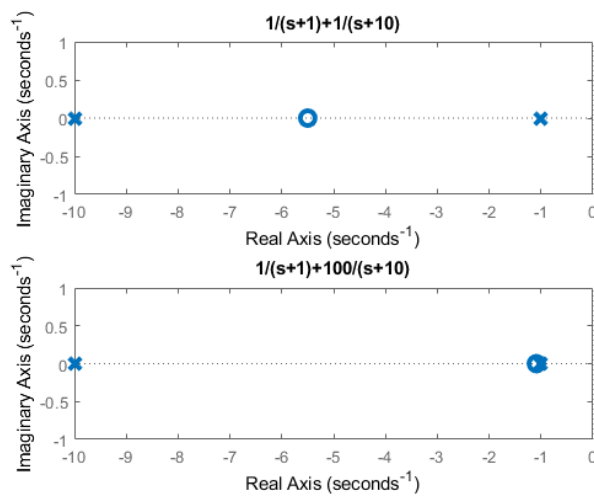


Figure 11.24: Poles and zeros of the transfer functions of Example 11.15.

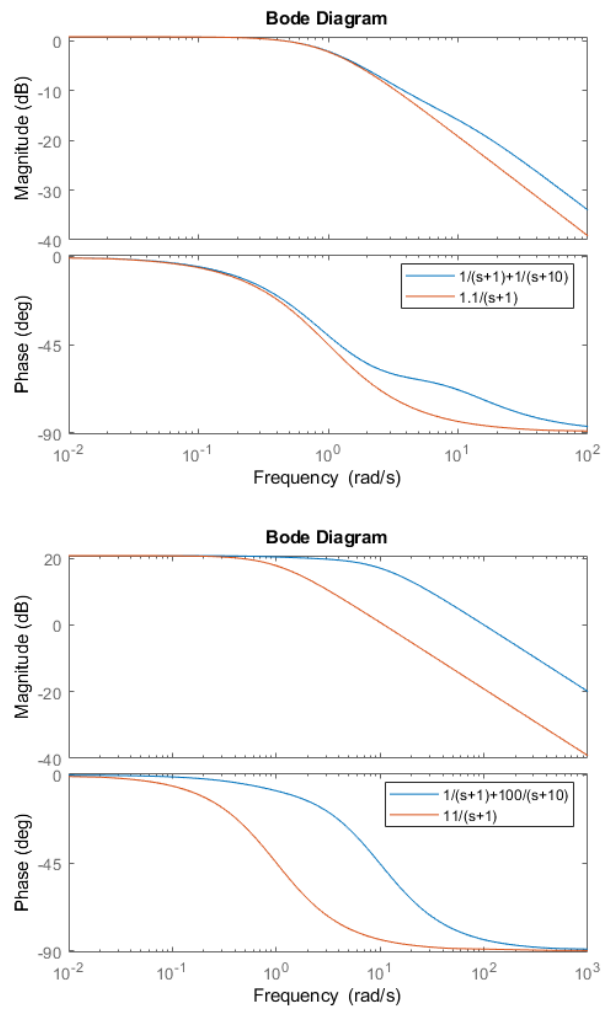


Figure 11.25: Bode plots of the transfer functions of Example 11.15.

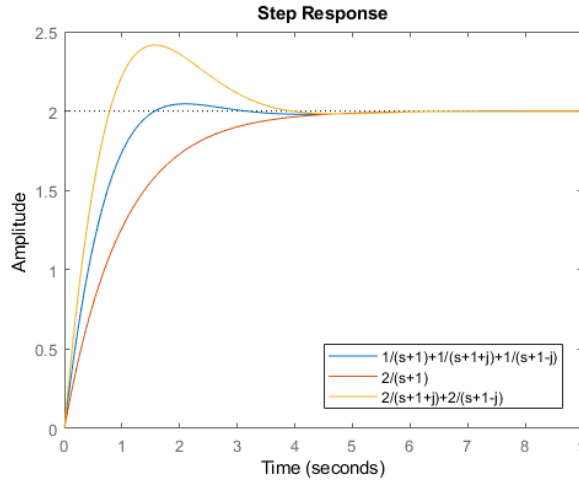


Figure 11.26: Unit step responses of the transfer functions of Example 11.16.

Example 11.15. Transfer function

$$G_1(s) = \frac{2s + 11}{s^2 + 11s + 10} = \frac{1}{s + 1} + \frac{1}{s + 10} \quad (11.192)$$

has a dominant pole, -1 , since pole -10 is faster. Residues are the same, so we can expect the step response of $G_1(s)$ to be rather similar to that of

$$\hat{G}_1(s) = \frac{\frac{11}{10}}{s + 1} \quad (11.193)$$

(which has the same static gain) as is indeed the case: see Figure 11.23. But transfer function

$$G_2(s) = \frac{101s + 110}{s^2 + 11s + 10} = \frac{1}{s + 1} + \frac{100}{s + 10} \quad (11.194)$$

which has the same poles has a much larger residue for the slower pole, and consequently its step response is not really that similar to that of

$$\hat{G}_2(s) = \frac{11}{s + 1} \quad (11.195)$$

(with the same static gain) as seen in Figure 11.23. This can also be seen from the position of the zero in both cases, given in Figure 11.24. $G_1(s)$ has the zero far from the dominant pole, in a location where its effect is faster, and so does not interfere significantly with its behaviour. $G_2(s)$ has the zero quite close to the dominant pole, thus interfering with its effect in time responses. \square

Remark 11.10. This can be more clearly understood looking at the Bode diagrams of $G_1(s)$ and $G_2(s)$ in Figure 11.25 in comparison with those of $\hat{G}_1(s)$ and $\hat{G}_2(s)$. The zero does not affect the frequency response of low frequency pole -1 for $G_1(s)$, but does so for $G_2(s)$. Remember that low frequencies correspond to larger time intervals, so in the case of $G_1(s)$ step responses will be, for large time intervals, similar to those of $\hat{G}_1(s)$. \square

Example 11.16. Transfer function

$$G(s) = \frac{3s^2 + 6s + 4}{s^3 + 3s^2 + 4s + 2} = \frac{1}{s + 1} + \frac{1}{s + 1 + j} + \frac{1}{s + 1 - j} \quad (11.196)$$

has three poles, none of which is dominant. So the step response of $G(s)$ is inbetween those of

$$\hat{G}_1(s) = \frac{2}{s + 1} \quad (11.197)$$

$$\hat{G}_2(s) = \frac{2}{s + 1 + j} + \frac{2}{s + 1 - j} \quad (11.198)$$

as seen in Figure 11.26. \square

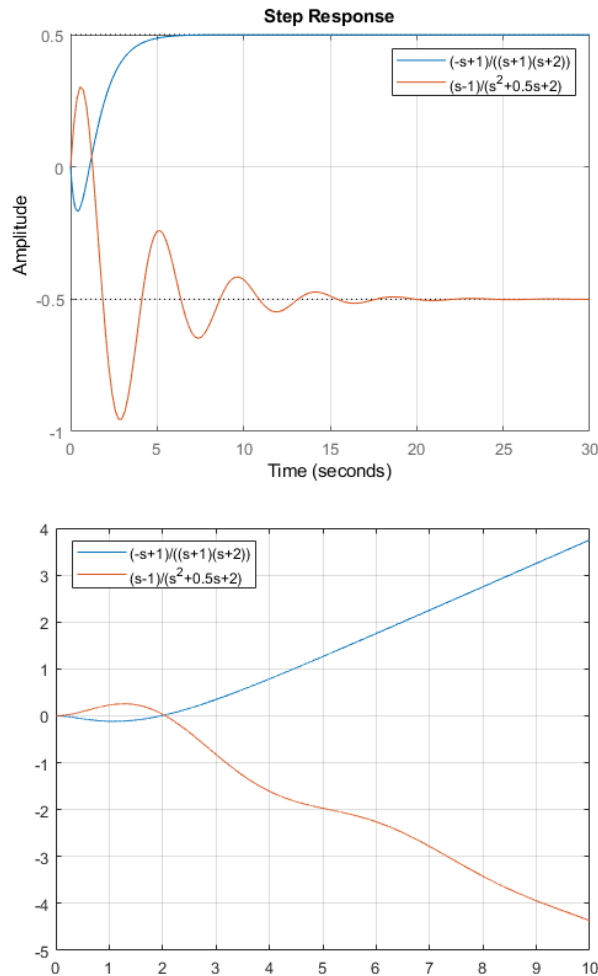


Figure 11.27: Unit step and unit ramp responses of the transfer functions of Example 11.17.

A time response can also be studied with respect to:

- its steady-state response, using the final value theorem, as seen above in Section 10.2;
- its transient response, from the difference between the number of zeros and poles, and the eventual presence of non-minimum phase zeros.

Undershoot

The latter cause an **undershoot** in time responses, i.e. the response begins with a negative derivative if the steady-state value is positive, or a positive derivative if the steady-state value is negative. In simpler terms, the response goes down before going up, or up before going down.

Example 11.17. Figure 11.27 shows the unit step and unit ramp responses of

$$G_1(s) = \frac{-s + 1}{(s + 1)(s + 2)} \quad (11.199)$$

$$G_2(s) = \frac{s - 1}{s^2 + 0.5s + 2} \quad (11.200)$$

The responses illustrate the following:

- undershoots exist in all time responses of the plant;
- there may or may not be further oscillations in the plant's response;
- undershoots exist whether the steady-state corresponds to a positive or a negative value. \square

Theorem 11.2. A non-minimum phase zero or pole causes an undershoot in time responses.

Proof. Let us consider the case of a transfer function with a non-minimum phase zero and a positive steady-state value; other cases are similar:

$$G(s) = \frac{-s + b}{\prod_{k=1}^n (s + p_k)} \quad (11.201)$$

The beginning of the transient response corresponds to very small time values, i.e. to very high frequencies. At high frequencies,

$$G(j\omega) = \frac{-j\omega + b}{\prod_{k=1}^n (j\omega + p_k)} \approx \frac{-j\omega}{(j\omega)^n} \quad (11.202)$$

which is also the frequency response of

$$\hat{G}(j\omega) = -\frac{1}{s^{n-1}} \quad (11.203)$$

So, for small values of t , $G(s)$ and $\hat{G}(s)$ behave similarly. $\hat{G}(s)$ has negative time responses to positive inputs (and positive time responses to negative inputs), and consequently so does $G(s)$ for $t \approx 0$. \square

Theorem 11.3. A transfer function with n poles and m zeros will have step responses $y(t)$ with $n - m - 1$ continuous derivatives, i.e.

Difference between the number of poles and the number of zeros

- if $n = m$, then $n - m - 1 = -1$: the step response $y(t)$ itself will be discontinuous;
- if $n = m + 1$, then $n - m - 1 = 0$: the step response $y(t)$ is continuous, but its first derivative $\frac{dy(t)}{dt}$ is discontinuous (and thus so are higher order derivatives);
- if $n = m + 2$, then $n - m - 1 = 1$: the step response $y(t)$ and its first derivative $\frac{dy(t)}{dt}$ are continuous, but its second derivative $\frac{d^2y(t)}{dt^2}$ is discontinuous (and thus so are higher order derivatives);
- if $n = m + 3$, then $n - m - 1 = 2$: the step response $y(t)$, its first derivative $\frac{dy(t)}{dt}$ and its second derivative $\frac{d^2y(t)}{dt^2}$ are continuous, but its third derivative $\frac{d^3y(t)}{dt^3}$ is discontinuous (and thus so are higher order derivatives);
- and so on.

Proof. The line of reasoning is similar to that of the theorem above, and because the system is linear we can consider the unit step response $y(t)$ of a transfer function with an arbitrary gain:

$$y(t) = \mathcal{L}^{-1}[Y(s)] = \mathcal{L}^{-1} \left[\frac{1}{s} \frac{\prod_{k=1}^m (s + z_k)}{\prod_{k=1}^n (s + p_k)} \right] \quad (11.204)$$

Again we are interested in the beginning of the transient response, which corresponds to very small time values, i.e. to very high frequencies, at which

$$Y(j\omega) = \frac{1}{j\omega} \frac{\prod_{k=1}^m (j\omega + z_k)}{\prod_{k=1}^n (j\omega + p_k)} \approx \frac{1}{j\omega} \frac{(j\omega)^m}{(j\omega)^n} = \frac{1}{(j\omega)^{n-m+1}} \quad (11.205)$$

This is the Fourier transform corresponding to $\frac{1}{s^{n-m+1}}$. It is easily proved by mathematical induction, from the Laplace transform of the Heaviside function and the Laplace transform of the integral, that

$$\mathcal{L}^{-1} \left[\frac{1}{s^k} \right] = \frac{t^{k-1}}{(k-1)!}, \quad t \geq 0, \quad k \in \mathbb{N} \quad (11.206)$$

Thus, for small values of t , $y(t)$ and $\frac{t^{n-m}}{(n-m)!}$ will behave similarly. The number of continuous derivatives of t^{n-m} , $n - m \in \mathbb{N}$, is as stated. \square

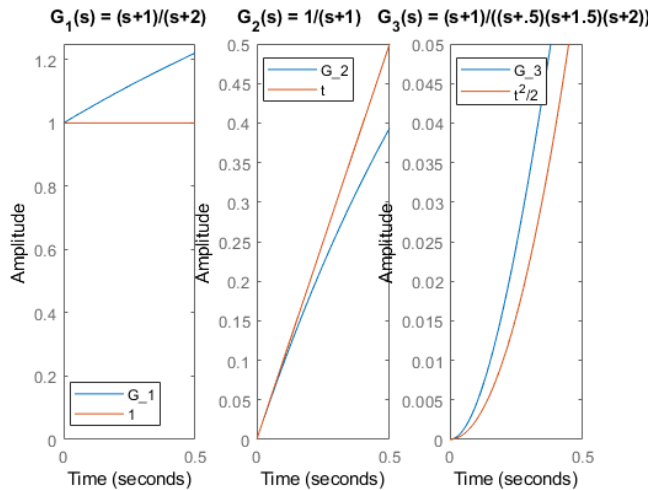


Figure 11.28: Step responses of three different transfer functions, for which the difference between the number of poles and zeros is 0, 1 and 2. Notice that, when difference is 0, the response is discontinuous, just as a step also is; when the difference is 1, the response is for a while close to a straight line; and when the difference is 2, the response is curved, and can be approximated by a parabola.

Remark 11.11. It is straightforward to find, in a similar manner, how many continuous derivatives will have the response of a system to a ramp, a parabola, or another input given by a polynomial. For inputs which are not polynomial, a polynomial approximation of the input may be used. \square

Example 11.18. Consider these transfer functions, with the following differences between the number of poles and zeros:

$$G_0(s) = \frac{s+1}{s+2}, \quad n-m=0 \quad (11.207)$$

$$G_1(s) = \frac{1}{s+1}, \quad n-m=1 \quad (11.208)$$

$$G_2(s) = \frac{s+1}{(s+0.5)(s+1.5)(s+2)}, \quad n-m=2 \quad (11.209)$$

According to Theorem 11.3 and its proof, at $t \approx 0$ their outputs will be similar to

$$y_0(t) = \mathcal{L}^{-1} \left[\frac{1}{s} \right] = H(t) \quad (11.210)$$

$$y_1(t) = \mathcal{L}^{-1} \left[\frac{1}{s} \frac{1}{s} \right] = t \quad (11.211)$$

$$y_2(t) = \mathcal{L}^{-1} \left[\frac{1}{s} \frac{1}{s^2} \right] = \frac{t^2}{2} \quad (11.212)$$

in what concerns the number of continuous derivatives. Figure 11.28 shows visually that it is so. \square

Glossary

‘To you I may seem a vulgar robber, but I bear on my shoulders the destiny of the human race. Your tribal life with its stone-age weapons and beehive huts, its primitive coracles and elementary social structure, has nothing to compare with our civilization — with our science, medicine and law, our armies, our architecture, our commerce, and our transport system which is rapidly annihilating space and time. Our right to supersede you is the right of the higher over the lower. Life —’

‘Half a moment,’ said Ransom in English. ‘That’s about as much as I can manage at one go.’ Then, turning to Oyarsa, he began

translating as well as he could. The process was difficult and the result — which he felt to be rather unsatisfactory — was something like this:

‘Among us, Oyarsa, there is a kind of *hnau* who will take other *hnaus*’ food and — and things, when they are not looking. He says he is not an ordinary one of that kind. He says what he does now will make very different things happen to those of our people who are not yet born. He says that, among you, *hnau* of one kindred all live together and the *hrossa* have spears like those we used a very long time ago and your huts are small and round and your boats small and light and like our old ones, and you have one ruler. He says it is different with us. He says we know much. There is a thing happens in our world when the body of a living creature feels pains and becomes weak, and he says we sometimes know how to stop it. He says we have many bent people and we kill them or shut them in huts and that we have people for settling quarrels between the bent *hnau* about their huts and mates and things. He says we have many ways for the *hnau* of one land to kill those of another and some are trained to do it. He says we build very big and strong huts of stones and other things — like the *pfiftriggi*. And he says we exchange many things among ourselves and can carry heavy weights very quickly a long way. Because of all this, he says it would not be the act of a bent *hnau* if our people killed all your people.’

C. S. LEWIS (1868 — †1963), *Out of the silent planet* (1938), 20

critically damped criticamente amortecido

damped frequency frequência amortecida

damping amortecimento

dominant pole polo dominante

maximum overshoot máximo sobreimpulso

minimum phase fase mínima

natural frequency frequência natural

non-minimum phase fase não-mínima

overdamped sobreamortecido

overshoot sobreimpulso

peak time tempo de pico

resonance frequency frequência de ressonância

resonant peak pico de ressonância

settling time tempo de estabelecimento

time constant constante de tempo

underdamped subamortecido

undamped sem amortecimento

undershoot subimpulso

Exercises

1. Sketch the following step responses, marking, whenever they exist,

- the settling time according to the 5% criterion,
- the settling time according to the 2% criterion,
- the steady-state value.

(a) $G(s) = \frac{15}{s+5}$, for input $u(t) = 4H(t)$

(b) $G(s) = \frac{10}{s-1}$, for input $u(t) = H(t)$

(c) $G(s) = \frac{1}{2s+1}$, for input $u(t) = -H(t)$

(d) $G(s) = \frac{-2}{4s+1}$, for input $u(t) = 10H(t)$

Table 11.1: Unit step response of Exercise 4.

time	output	time	output
0.0	0.0000	0.8	0.0950
0.1	0.0000	0.9	0.0982
0.2	0.0000	1.0	0.0993
0.3	0.0000	1.1	0.0998
0.4	0.0000	1.2	0.0999
0.5	0.0000	1.3	0.1000
0.6	0.0632	1.4	0.1000
0.7	0.0865	1.5	0.1000

(e) $G(s) = \frac{10}{s}$, for input $u(t) = H(t)$

2. Sketch the Bode diagrams of the following transfer functions, indicating
- the gain for low frequencies,
 - the frequency at which the gain is 3 dB below the gain for low frequencies,
 - the slope of the gain for high frequencies,
 - the phase for low frequencies,
 - the phase for high frequencies,
 - the frequency at which the phase is the average of those two values.

Hint: draw the asymptotes of the Bode diagram. Mark the frequency at which the gain has decreased 3 dB.

(a) $G(s) = \frac{15}{s+5}$

(b) $G(s) = \frac{1}{s+10}$

(c) $G(s) = \frac{1}{2s+1}$

(d) $G(s) = \frac{2}{4s+1}$

(e) $G(s) = \frac{10}{s}$

3. Let $G(s) = \frac{1}{s+1}$.

- (a) Consider the unit step response of $G(s)$. What is the settling time, according to the 5% criterion?
- (b) Find analytically the unit ramp response $y(t)$ of $G(s)$.
- (c) Find the analytical expression of the steady-state $y_{ss}(t)$ of that response $y(t)$.
- (d) How long does it take for $\left| \frac{y_{ss}(t) - y(t)}{y(t)} \right|$ to be less than 5%? In other words, find how long it takes for the unit ramp response to be within a 5% wide band around its steady state.
4. A first order system $\frac{K}{s+p}$ has the response tabulated in Table 11.1, when its input is a unit step applied at instant $t = 0.5$ s. Find the gain K and the pole p . *Hint:* subtract the response from the steady state value; you should now have an exponential with a negative power. Plot its logarithm and adjust a straight line.
5. For each of the transfer functions below, and for the corresponding step input, find, if they exist:
- the natural frequency ω_n and the damping factor ξ ,
 - the steady state value y_{ss} ,

- the peak time t_p and the maximum overshoot M_p (expressed in percentage),
- the 5% and the 2% settling times (use the expressions for the exponential envelope of the oscillations),
- the location of the poles,
- the time instant at which the response first reaches $\frac{y_{ss}}{2}$,

and sketch the step response.

- (a) $G(s) = \frac{7}{s^2 + 0.4s + 1}$, for input $u(t) = 0.1H(t)$
- (b) $G(s) = \frac{1}{s^2 + 5.1s + 9}$, for input $u(t) = 18H(t)$
- (c) $G(s) = \frac{1}{2s^2 + 8}$, for input $u(t) = H(t)$
- (d) $G(s) = \frac{10}{s^2 - s + 1}$, for input $u(t) = 2H(t)$
- (e) $G(s) = \frac{0.3}{s^2 + 4s - 1}$, for input $u(t) = 15H(t)$

6. Sketch the Bode diagrams of the following transfer functions, indicating

- the gain for low frequencies,
- the resonant peak value, and the frequency at which it is located, if indeed there is one,
- the slope of the gain for high frequencies,
- the phase for low frequencies,
- the phase for high frequencies,
- the frequency at which the phase is the average of those two values.

Hint: draw the asymptotes of the Bode diagram. If there is a resonant peak mark it in your plot.

- (a) $G(s) = \frac{1}{s^2 + 20s + 100}$
- (b) $G(s) = \frac{7}{s^2 + 0.4s + 1}$
- (c) $G(s) = \frac{1}{s^2 + 5.1s + 9}$
- (d) $G(s) = \frac{1}{2s^2 + 8}$

7. Find the second order transfer functions that, for a unit step input, have:

- (a) $t_p = 0.403$ s, $M_p = 16.3\%$, $y_{ss} = 0.8$
- (b) $t_p = 0.907$ s, $y(t_p) = 11.63$, $y_{ss} = 10$
- (c) $t_r = 0.132$ s, $t_{s2\%} = 2.0$ s, $y_{ss} = 0.5$

8. Consider the mechanical system in Figure 11.29. When $f(t) = 8.9$ N, $t \geq 0$, the output has $t_p = 1$ s, $M_p = 9.7\%$, and $y_{ss} = 3 \times 10^{-2}$ m.

- (a) Find the values of mass M , viscous damping coefficient B , and spring stiffness K .
- (b) Suppose we want the same steady-state regime and the same settling time, but a maximum overshoot of 0.15%. What should the new values of M , B and K be?

9. Plot the Bode diagrams of the following plants:

- (a) $G(s) = \frac{-4s + 20}{s^3 + 0.4s^2 + 4s}$
- (b) $\frac{d^3y(t)}{dt^3} + 16\frac{d^2y(t)}{dt^2} + 65\frac{dy(t)}{dt} + 50y(t) = 100\frac{du(t)}{dt} + 50u(t)$

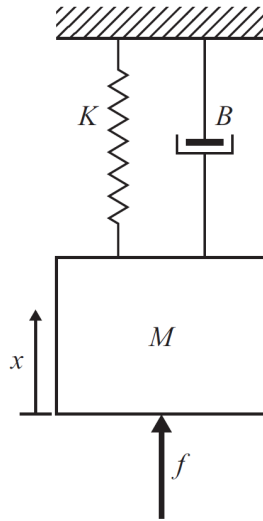


Figure 11.29: System of Exercise 8.

$$(c) G(s) = \frac{120(s+1)}{s(s+2)^2(s+3)}$$

$$(d) G(s) = \frac{s^2}{(s+0.5)(s+10)}$$

$$(e) G(s) = \frac{10s}{(s+10)(s^2+s+2)}$$

$$(f) G(s) = \frac{(s+4)(s+20)}{(s+1)(s+80)}$$

10. Establish a correspondence between the three Bode diagrams and the three unit step responses in Figure 11.30.
11. Establish a correspondence between the three Bode diagrams and the three unit step responses in Figure 11.31.
12. Find the transfer functions corresponding to the Bode diagrams in Figure 11.32.
13. Use the Routh-Hurwitz criterion to find how many unstable poles each of the following transfer functions has, and classify each system as stable, marginally stable, or unstable.

$$(a) \frac{s^2 + \frac{5}{7}s - 10}{s^4 - 2s^3 - 13s^2 + 14s + 24}$$

$$(b) \frac{s+2}{s^4 - 2s^3 - 13s^2 + 14s + 24}$$

$$(c) \frac{s+2}{s^6 - 2s^5 - 13s^4 + 14s^3 + 24s^2} \quad \text{Hint: can you put anything in evidence in the denominator?}$$

$$(d) \frac{s^3 + 2s^2 + s}{s^4 + 4s^3 + 4s + 5}$$

$$(e) \frac{s^3 + 2s^2 + s}{s^5 + 4s^4 + 4s^2 + 5s}$$

$$(f) \frac{s^3 + 2s^2 + s}{2s^3 - 6s + 4}$$

14. Find the ranges of values of $K_1, K_2 \in \mathbb{R}$ for which the systems having transfer functions with the following denominators are stable.
 - (a) $s^3 + 3s^2 + 10s + K_1$
 - (b) $s^3 + K_2s^2 + 10s + 5$
 - (c) $s^3 + 2s^2 + (K_1 + 1)s + K_2$
15. How many unstable, marginally stable, and stable poles do the following transfer functions have?

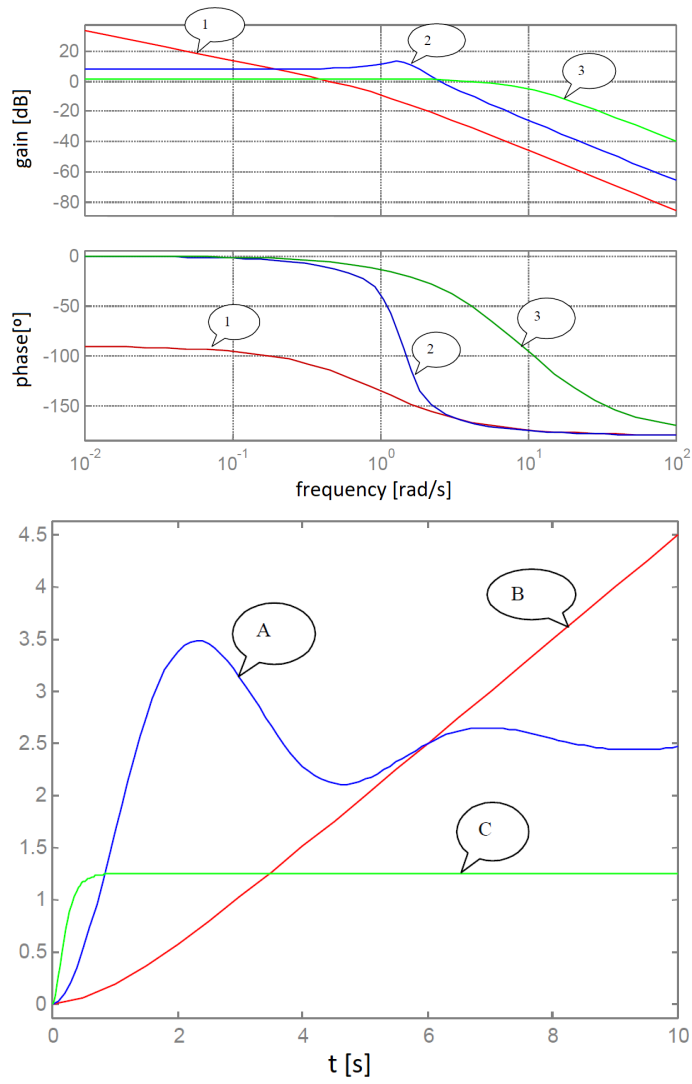


Figure 11.30: Bode diagrams and unit step responses of Exercise 10.

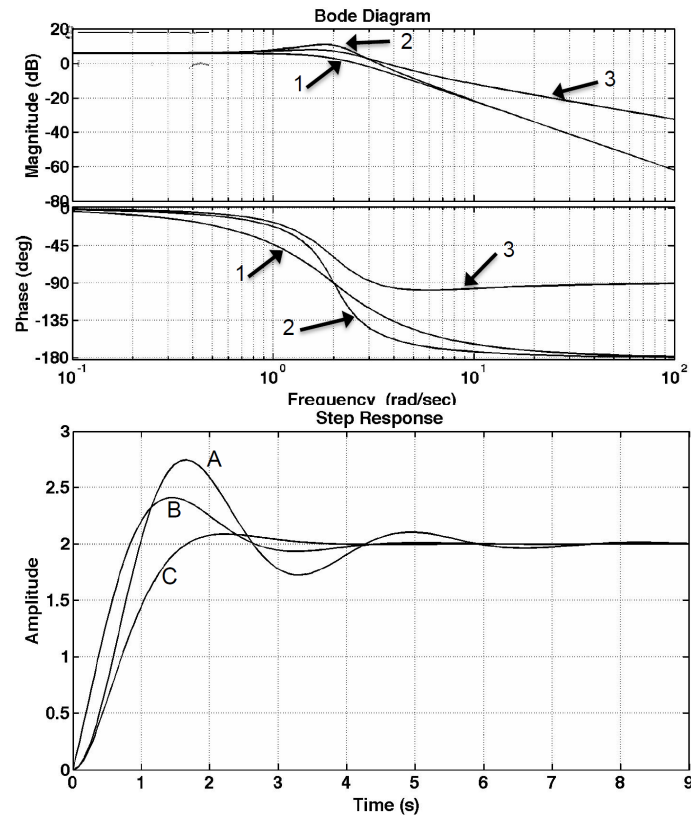


Figure 11.31: Bode diagrams and unit step responses of Exercise 11.

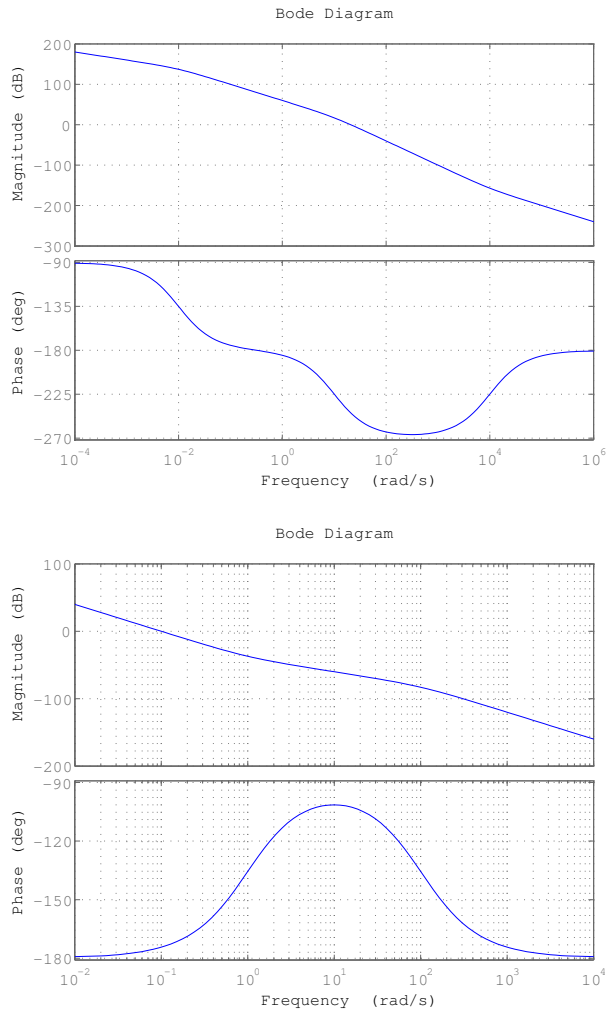


Figure 11.32: Bode diagrams of Exercise 12.

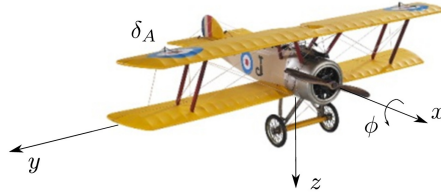


Figure 11.33: The Sopwith Camel, a fighter aircraft from World War I.

$$(a) \quad G_1(s) = \frac{s - 2}{s^4 + 2s^3 + 5s^2 + 8s + 4}$$

$$(b) \quad G_2(s) = \frac{s^4 + 20s^3 + 150s^2 + 500s + 625}{s^5 - 2s^4 + 5s^3 - 8s^2 + 4s}$$

$$(c) \quad G_3(s) = \frac{s^2 - 4}{s^3 + 5s^2 + 6s + 8}$$

16. Find analytically the unit step responses of $G_1(s) = \frac{100}{s + 10}$ and $G_2(s) = \frac{s + 100}{s + 10}$. Sketch them both in the same plot, marking the 5% settling time for each. Plot separately the difference between them. Then do the same for $G_3(s) = \frac{8}{s + 12}$ and $G_4(s) = \frac{s + 8}{s + 12}$.

17. Consider the following transfer functions:

$$G_1(s) = \frac{5050s + 10000}{s^2 + 101s + 100} \quad (11.213)$$

$$G_2(s) = \frac{101s + 10000}{s^2 + 101s + 100} \quad (11.214)$$

- (a) Find their poles.
 (b) Which pole is faster? Why?
 (c) Which of the two transfer functions will respond faster to a unit step? Why?
18. The roll angle ϕ (rotation around the x -axis) of the aircraft in Figure 11.33 is given by

$$\dot{p} = -2p + 12\delta_A \quad (11.215)$$

where $p = \dot{\phi}$ and δ_A is the aileron deflection. All variables are given in SI units.

- (a) Find transfer function $G_p(s) = \frac{p(s)}{\delta_A(s)}$.
 (b) Find transfer function $G_\phi(s) = \frac{\phi(s)}{\delta_A(s)}$.
 (c) Sketch $p(t)$ when the ailerons undergo a $1^\circ = \frac{\pi}{180}$ rad step.
 (d) What is $p(t)$ when $t = 0.5$ s?
 (e) Sketch $\phi(t)$ for the same input.
 (f) Suppose you want $\phi = 30^\circ$ after 20 s. Without calculating $\phi(t)$, find an approximate value for the necessary aileron deflection.
19. Plot the Bode diagram of transfer function $G(s) = \frac{120(s + 1)}{s(s + 2)^2(s + 3)}$.

Part III

Sensors and actuators

Then Dick gave a cry. “It’s just come into sight. Look, when we passed those tall trees on the bank yonder, Tall Stone came into view. It was behind them before that.”

“Good,” said Julian. “Now I’m going to stop paddling and keep Tall Stone in sight. If it goes out of sight I’ll tell you and you must back-paddle. Dick, can you possibly paddle and look out for something that could be Tock Hill on the opposite side? I daren’t take my eyes off Tall Stone in case it disappears.”

“Right,” said Dick, and paddled while he looked earnestly for Tock Hill.

“Got it!” he said suddenly. “It must be it! Look, over there — a funny little hill with a pointed top. Julian, can you still see Tall Stone?”

“Yes,” said Julian. “Keep your eyes on Tock Hill. Now it’s up to the girls. George, paddle away and see if you can spot Steeple.”

“I can see it now, already!” said George, and for one moment the boys took their eyes off Tall Stone and Tock Hill and looked where George pointed. They saw the steeple of a faraway church glinting in the morning sun.

“Good, good, good,” said Julian. “Now Anne — look for Chimney — look down towards the end of the lake — where the house is. Can you see its one chimney?”

“Not quite,” said Anne. “Paddle just a bit to the left — the left, I said, George! Yes — yes, I can see the one chimney. Stop paddling everyone. We’re here!”

They stopped paddling but the raft drifted on, and Anne lost the chimney again! They had to paddle back a bit until it came into sight. By that time George had lost her steeple!

At last all four things were in view at once, and the raft seemed to be still and unmoving on the quiet waters of the lake.

Enid BLYTON (1897 — †1968), *Five on a hike together* (1951), XVIII

In this part of the lecture notes:

Chapter 12 introduces the basic concepts of measuring chains and control loops.

Chapter 13 presents the technology of the most common types of sensors, and the criteria to choose them.

Chapter 14 presents the technology of the most common types of actuators, and the criteria to choose them.

Here is what you need to know beforehand to follow these chapters:

- The Laplace and Fourier transforms, from Chapter 2;
- Transfer functions, from Sections 4.1 and 4.2 of Chapter 4;
- System theory, from Part II.

Chapter 12

Measuring chains and actuation chains

‘Because, you see, I’m the only person who can open the door.’

‘But you have given me the word. Was that a lie?’

‘No — the word’s all right. But, you see, it’s one of these new-style electric doors. In fact, it’s really the very latest thing in doors. I’m rather proud of it. It opens to the words “Open Sesame” all right — *but to my voice only.*’

‘Your voice? I will choke your voice with my own hands. What do you mean — your voice only?’

‘Just what I say. Don’t clutch my throat like that, or you may alter my voice so that the door won’t recognise it. That’s better. It’s apt to be rather picky about voices. It got stuck up for a week once, when I had a cold and could only implore it in a hoarse whisper. Even in the ordinary way, I sometimes have to try several times before I hit on the exact right intonation.’

She turned and appealed to a short, thick-set man standing beside her.

‘Is this true? Is it possible?’

‘Perfectly, ma’am, I’m afraid,’ said the man civilly. From his voice Wimsey took him to be a superior workman of some kind — probably an engineer.

‘Is it an electrical device? Do you understand it?’

‘Yes, ma’am. It will have a microphone arrangement somewhere, which converts the sound into a series of vibrations controlling an electric needle. When the needle has traced the correct pattern, the circuit is completed and the door opens. The same thing can be done by light vibrations equally easily.’

‘Couldn’t you open it with tools?’

‘In time, yes, ma’am. But only by smashing the mechanism, which is probably well protected.’

‘You may take that for granted,’ interjected Wimsey reassuringly.

She put her hands to her head.

‘I’m afraid we’re done in,’ said the engineer, with a kind of respect in his tone for a good job of work.

Dorothy L. SAYERS (1893 — †1957), *Lord Peter Views the Body* (1928), The adventurous exploit of the cave of Ali Baba

In this chapter we take a closer look on sensors and actuators.

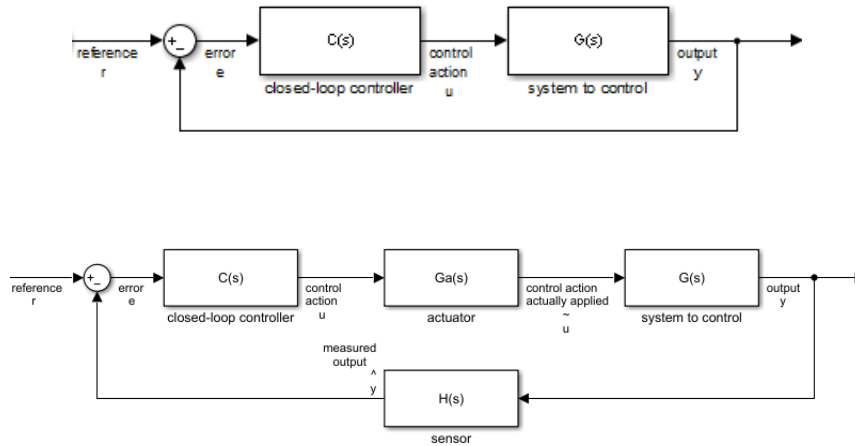


Figure 12.1: Closed loop control. Top: simplest representation (see Figure 9.13). Bottom: sensor and actuator explicitly shown.



Figure 12.2: Ammeter from the Tagus power plant (currently the Museum of Electricity), Lisbon. Notice that the scale is not linear.

12.1 What are measuring chains and actuation chains

Let us take another look at the simplest representation of a feedback control system, already known to us from Figure 9.13 and repeated in Figure 12.1. We saw in chapter 9 that, to feed back the output to the controller, we need a sensor that measures the output y . The sensor is a device that converts the variable to be read into another which can be read directly. This conversion is called **transduction**. While we can often assume a perfect sensor $H(s) = 1$, in which case the measured output \hat{y} is equal to the output itself (i.e. $\hat{y} = y$), the sensor must exist. Sensor technology will be addressed in more detail in chapter 13.

Example 12.1. The analogical ammeter in Figure 12.2 transduces current into an angle. The angle can be read directly by sight on a scale. This sensor does not record the reading. \square

Example 12.2. The seismograph in Figure 12.3 transduces a position (corresponding to the amplitude of the vibration of the ground) again into an angle. The reading is recorded on paper; time is transduced into a position thanks to a rotation of the drum obtained through clockwork. \square

Likewise, the control action u provided by the controller must somehow be implemented in the plant $G(s)$; this is done through an **actuator**: for instance, a motor in a mechanical plant, or a heater in a thermal plant. This is shown in Figure 12.1. Once more, we can have such a good actuator $G_a(s)$ that we can assume the control action actually applied \tilde{u} to be equal to the control

Transduction

Actuator

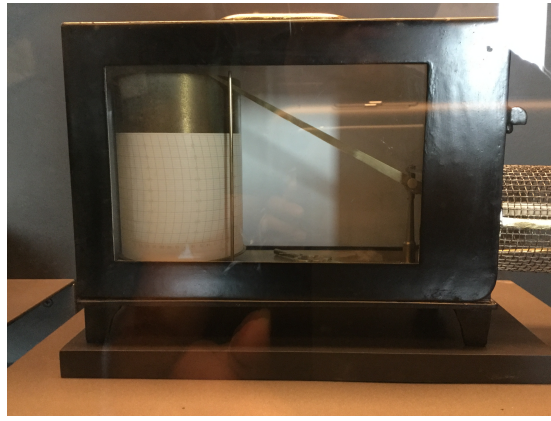


Figure 12.3: Seismograph from the Santa Marta lighthouse (currently a museum), Cascais.

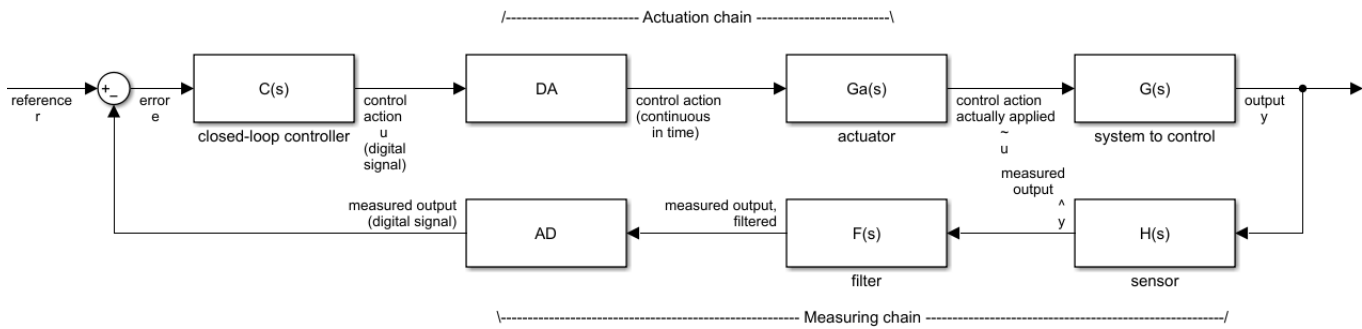


Figure 12.4: Closed loop control with sensor, actuator and AD/DA conversion explicitly shown.

action u that should be applied, the one given by the actuator — that is to say, $G_a(a) = 1$ or $\tilde{u} = u$. Even if this is not assumed, the (imperfect) behaviour of the actuator may be considered simply as part of the plant to be controlled $G(s)$. This makes sense if there is no way of changing the actuator, but if the actuator can be replaced by a different one (hopefully a better one, though poor replacement parts and actuator malfunctions are also part of life) it is easier to have separate models and change only the actuator model. Actuator technology will be addressed in more detail in chapter 14.

It is perfectly possible to have plants where no more elements need to be modelled: think of the flush tank from Figure 3.16, where the sensor is the floater, and the actuator the lever that rotates on a hinge and is connected to the floater. But nowadays most sensors and actuators are electronic, and perform a transduction of the measured variable to an electrical quantity (voltage or current); the controllers that use measurements to decide the control actions to apply are also electronic. Consequently, \hat{y} and u are digital signals (i.e. both discrete in time and quantised in amplitude, as we saw in section 3.2); while y is analogical, and \tilde{u} must be, if not analogical, at least continuous in time. That is why it is necessary to have a *digital to analogical converter*, or simply a *DA converter*, between the controller and the actuator, as seen in Figure 12.4. The function of the AD converter is to receive a digital signal as input and provide an analogical output (remember the example in Figure 3.13). Likewise, if after all the sensor provides a measurement \hat{y} which is not digital, an *analogical to digital converter*, or DA converter in short, is needed after the sensor. Even if the sensor itself already provides a digital output, it is because it incorporates the function of the DA converter. The function of the AD converter is to discretise (in time) and quantise (in amplitude) its input.

The AD and the DA conversions may be carried out by the same device, *AD/DA converter* which will have:

- analog inputs, and the corresponding digital outputs;

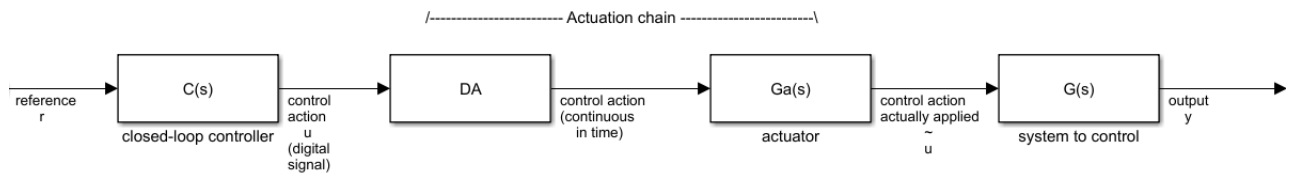


Figure 12.5: Open loop control with actuation chain explicitly shown.

- digital inputs, and the corresponding analog outputs.

Filter

Figure 12.4 shows yet another element after the sensor: a **filter**, that is, a system designed to eliminate noise from the measurement. This noise can be present in the output of the plant, or be a result of the sensor itself. The filter can be applied to the analog measurement before AD conversion, or be applied to the digital signal after AD conversion (this last option is the only one possible if the sensor provides a digital reading).

Figure 12.4 also indicates what, in a digital control system, constitutes

Measuring chain

- the **measuring chain**: sensor, filter, AD conversion;

Actuation chain

- the **actuation chain**: DA conversion, actuator.

If these chains are ideal, they can be modelled by $\frac{\hat{y}}{y} = 1$ and $\frac{\tilde{u}}{u} = 1$.

An open-loop control system will also have an actuation chain, as seen in Figure 12.5, but no measuring chain. A measuring chain is present in a system which is measured but not controlled.

12.2 Filters

We have just mentioned filters in connection with the measuring chain, as a means of eliminating noise from sensor measurements. In general, a filter is a system that eliminates some frequency components of a signal. These components are said to be *cut*. Those that are not eliminated are said to *pass* the filter. Figure 12.6 shows the ideal behaviour of four types of filters:

Types of filters

- **low-pass** filters eliminate high frequencies and pass low frequencies;
- **high-pass** filters eliminate low frequencies and pass high frequencies;
- **band-pass** filters eliminate both low and high frequencies and pass intermediate frequencies;
- **band-stop** or **band-reject** filters eliminate intermediate frequencies and pass both high and low frequencies.

In terms of periods of oscillations:

- low-pass filters eliminate small periods and pass large periods;
- high-pass filters eliminate large periods and pass small periods;
- band-pass filters eliminate both small and large periods and pass intermediate periods;
- band-stop filters eliminate intermediate periods and pass both small and large periods.

Cut-off frequency

What low and high frequencies (or large and small periods) are depends on the application. The cut-off frequency ω_c is the limit separating the pass-band (the interval of frequencies that pass the filter) from the stop-band (the interval of frequencies that the filter eliminates).

Perfect filters do not exist

No filter can be as good as what Figure 12.6 shows. Figure 12.7 pictures a more realistic behaviour. Notice in particular:

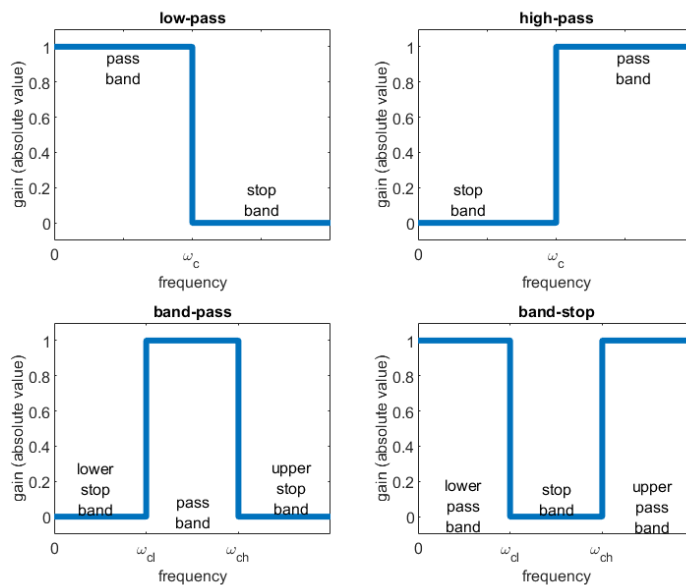


Figure 12.6: Ideal filters.

- that both the pass-band and the stop-band are defined within some tolerance (in Figure 12.7, δ_p is the tolerance of the pass-band and δ_s is the tolerance of the stop-band), which can be larger or narrower depending on what is needed for each particular application;
- that there may be ripples in each of those intervals of frequencies;
- that, because the gain of a filter within the stop-band will never reach 0 (i.e. $-\infty$ dB), no frequencies of the filtered signal are ever eliminated: only attenuated;
- that, because the gain within the pass-band will not always be 1 (i.e. 0 dB), neither will the phase of the filter in the interval be always 0° : and so, in practice, all filters always introduce some **distortion** in the signal that passes;
- that there is between the pass-band and the stop-band a transition band, i.e. a range of frequencies where the signal no longer passes (since the gain is below tolerance δ_p of the pass-band) and where it is not sufficiently attenuated (since the gain is above tolerance δ_s of the stop-band), lying between the cut-off frequency ω_c which delimits the pass-band and frequency ω_s which delimits the stop-band.

Example 12.3. A 1 Hz signal is measured with an electronic sensor, and the reading is corrupted with 50 Hz noise from the electric grid. To filter out this $2\pi \times 50 = 314$ rad/s noise without distorting the $2\pi \times 1 = 6.28$ rad/s signal, we will use a first order system, with the pole at the same distance from the frequencies of signal and noise in a Bode diagram. A Bode diagram has a logarithmic scale for frequencies, so the pole should be at the geometric mean of 6.28 rad/s and 314 rad/s, which is $\sqrt{6.28 \times 314} = 44.4$ rad/s:

$$F_1(s) = \frac{44.4}{s + 44.4} \quad (12.1)$$

Figure 12.8 shows the Bode diagram of $F_1(s)$, and the gain and the phase at the frequencies of both signal and noise:

- $|F_1(j 6.28)| = 10^{\frac{-0.0861}{20}} = 0.99$, which means that 99% of the signal passes: it undergoes a 1% attenuation;
- $\angle F_1(j 6.28) = -8^\circ$, which means that distortion due to a delay in the phase is not very significant;

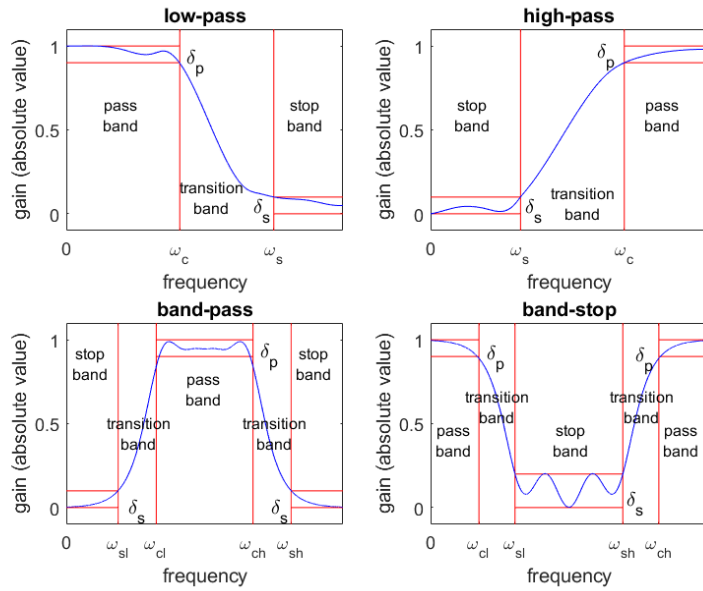


Figure 12.7: Behaviour of filters, closer to reality than the ideal behaviour in Figure 12.6. Notice that the band-stop filter shown has a larger value of δ_s , and larger ripples in the stop-band, than the other filters.

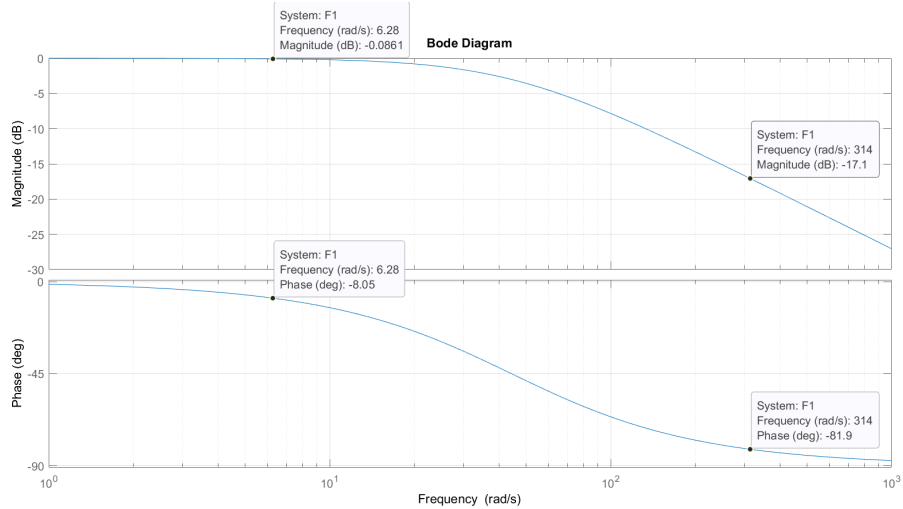


Figure 12.8: Bode diagram of filter $F_1(s)$ given by (12.1).

- $|F_1(j 314)| = 10^{-\frac{17.1}{20}} = 0.14$, which means that 14% of the noise passes: it undergoes a 86% attenuation;
- $\angle F_1(j 314) = -82^\circ$, which means that noise is not only attenuated: its phase is also significantly distorted. \square

Example 12.4. If the 14% attenuation is not enough, a second order filter can be used instead. Let us put two poles at 44.4 rad/s:

$$F_2(s) = \frac{44.4^2}{(s + 44.4)^2} \tag{12.2}$$

Figure 12.9 shows the Bode diagram of $F_2(s)$, and the gain and the phase at the frequencies of both signal and noise:

- $|F_2(j 6.28)| = 10^{-\frac{0.172}{20}} = 0.98$, which means that 98% of the signal passes: it undergoes a 2% attenuation;
- $\angle F_2(j 6.28) = -16^\circ$, which means that distortion due to a delay in the phase is now larger;

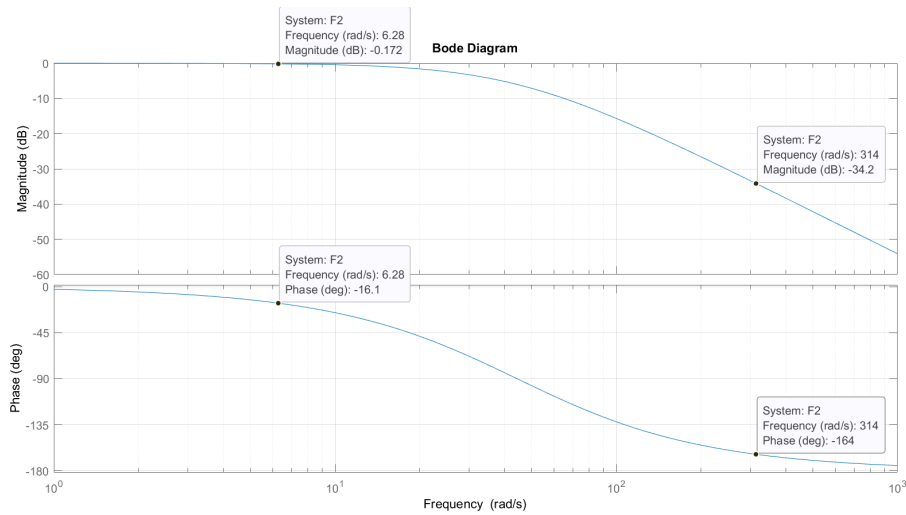


Figure 12.9: Bode diagram of filter $F_2(s)$ given by (12.2).

- $|F_2(j 314)| = 10^{-\frac{34.2}{20}} = 0.02$, which means that 2% of the noise passes: it undergoes a 98% attenuation;
- $\angle F_2(j 314) = -164^\circ$, which means that noise is now even more significantly distorted.

Notice how we got a better attenuation of the noise, paying the price of distorting more the signal. \square

Remark 12.1. In the previous example we used a second order system with damping coefficient $\xi = 1$ (double real pole). Notice that:

Second order filters should have $\frac{\sqrt{2}}{2} \leq \xi \leq 1$

- it does not make sense to have an overdamped system, $\xi > 1$, with two real poles which do not coincide: the transition band would be larger without any advantage (see Figures 11.8 and 11.9 again);
- it does not make sense to have an underdamped system with $\xi < \frac{\sqrt{2}}{2}$: there would be a resonance peak, meaning that the input would be amplified near the end of the pass band, which does not make sense for a filter (see Figures 11.10 and 11.11 again).

As a consequence, second order filters only make sense if critically damped or lightly underdamped. \square

Remark 12.2. In both examples above, filter design is rather easy, because signal and noise have well defined frequencies, which are significantly apart from each other. Whenever frequency ranges of signal or noise are not well known, or whenever they are close to each other, filter design becomes far more difficult. When they overlap, it is of course impossible to filter out noise without damaging the signal, or preserving the signal without letting some noise pass. \square

When filter design becomes hard

Remark 12.3. In both examples above, putting ω_c halfway through signal and noise frequencies is an attempt to balance good noise attenuation on the one hand and little signal distortion on the other. We could also have:

- fixed in advance the minimum acceptable attenuation of the noise, find the resulting ω_c , and then check if signal distortion is acceptable; or
- fixed in advance the maximum acceptable distortion of the signal, find the resulting ω_c , and then check if noise attenuation is acceptable. \square

To design filters of orders above 2, there are two criteria that can be followed:

- minimise the ripple of the pass-band (the price to pay being a larger transition band);
- minimise the width of the transition band (the price to pay being ripples in the pass-band).

Minimising ripples leads to Butterworth filters. We will not prove, but only state the following result:

Butterworth filters minimise pass-band ripples

- low-pass Butterworth filters of order n with cut-off frequency ω_c are given by

$$F(s) = \frac{1}{H_n\left(\frac{s}{\omega_c}\right)} \quad (12.3)$$

- high-pass Butterworth filters of order n with cut-off frequency ω_c are given by

$$F(s) = \frac{1}{H_n\left(\frac{\omega_c}{s}\right)} \quad (12.4)$$

where $H_n(x)$ are Butterworth polynomials, given by

$$H_1(x) = x + 1 \quad (12.5)$$

$$H_2(x) = x^2 + 1.4142x + 1 \quad (12.6)$$

$$H_3(x) = (x + 1)(x^2 + x + 1) \quad (12.7)$$

$$= x^3 + 2x^2 + 2x + 1 \quad (12.8)$$

$$H_4(x) = (x^2 + 0.7654x + 1)(x^2 + 1.8478x + 1) \quad (12.9)$$

$$= x^4 + 2.6131x^3 + 3.4142x^2 + 2.6131x + 1 \quad (12.10)$$

$$H_5(x) = (x + 1)(x^2 + 0.6180x + 1)(x^2 + 1.6180x + 1) \quad (12.11)$$

$$= x^5 + 3.2361x^4 + 5.2361x^3 + 5.2361x^2 + 3.2361x + 1 \quad (12.12)$$

⋮

$$H_n(x) = \sum_{k=0}^n a_{k,n} x^k \quad (12.13)$$

$$a_{0,n} = 1 \quad (12.14)$$

$$a_{k,n} = \prod_{m=1}^k \frac{\cos\left(\frac{(m-1)\pi}{2n}\right)}{\sin\left(\frac{m\pi}{2n}\right)}, \quad k = 1, \dots, n \quad (12.15)$$

Example 12.5. Figure 12.10 shows the Bode diagram of the first 10 Butterworth filters with $\omega_c = 44.4$ rad/s, as in Examples 12.3 and 12.4. Notice how steeper gain slopes, and consequently larger attenuations at high frequencies, are got at the cost of earlier distortions in the phase. A constant gain and a negative phase mean that the signal will be delayed; as we will see in Chapter 24, delays are a serious nuisance in control systems, and easily make them unstable.

The transfer functions of the first two filters are

$$F_1(s) = \frac{1}{\frac{s}{44.4} + 1} \quad (12.16)$$

$$F_2(s) = \frac{1}{\left(\frac{s}{44.4}\right)^2 + 2 \underbrace{\frac{\sqrt{2}}{2}}_{\xi} \frac{s}{44.4} + 1} = \frac{1971}{s^2 + 62.79s + 1971} \quad (12.17)$$

Notice that (12.16) is the same as (12.1). □

Chebyshev and elliptic filters minimise the transition band

Designing a filter by minimising the width of the transition band leads to Chebyshev and elliptic filters. We will not study them, just mention their existence. Butterworth, Chebyshev and elliptic filters of any order can be easily used in SIMULINK with block `Analog Filter Design` (from the `DSP System toolbox`). All you have to do is to specify what type of filter you want (low-pass, high-pass, band-pass, or band-stop), what design method you want (elliptic, Butterworth, Chebyshev...), the order of the filter, and the passband. Figure 12.11 compares filters of fifth order of these types.

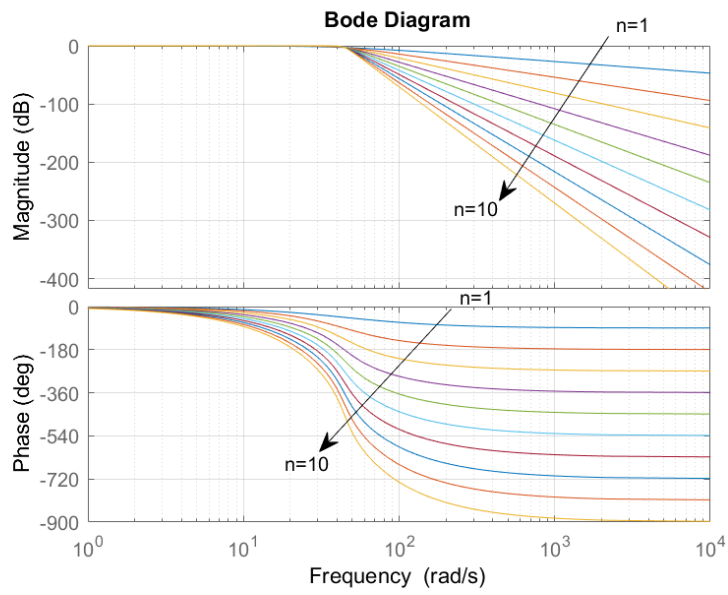


Figure 12.10: Bode diagram of Butterworth filters from Example 12.5.

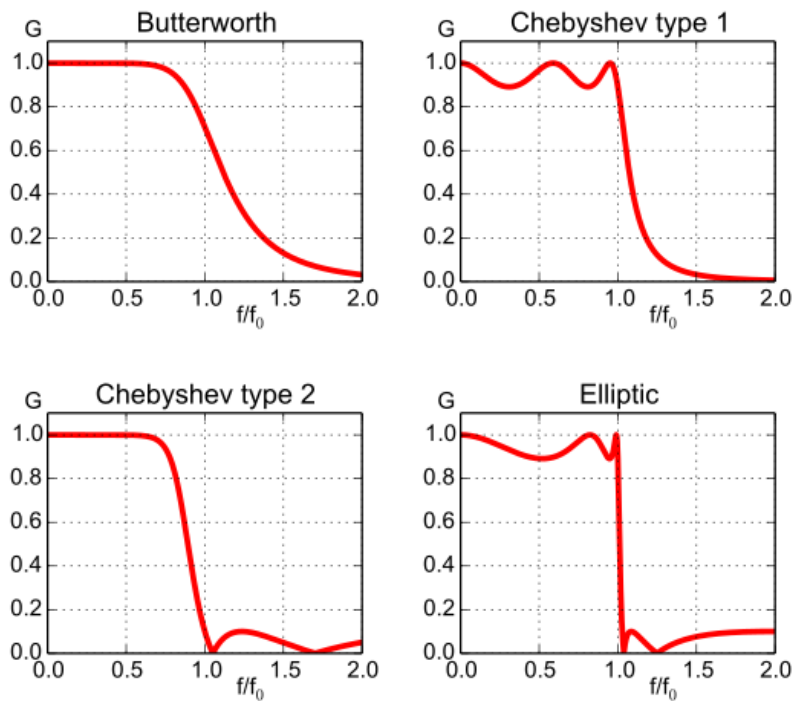


Figure 12.11: Absolute value of the gain of different fifth order filters, as a function of normalised frequency (source: Wikimedia).

12.3 Bandwidth

Bandwidth of a filter

Strictly speaking, the bandwidth ω_b of a filter is the width of its pass-band. Consequently:

- In the case of a low-pass filter, the lower limit of the pass-band is 0 rad/s, thus $\omega_b = \omega_c - 0 = \omega_c$.
- In the case of a high-pass filter, the upper limit of the pass-band is infinite, so strictly speaking $\omega_b = \infty - \omega_c = \infty$. But, since for a low-pass filter $\omega_b = \omega_c$, for high-pass filters ω_c is frequently given as the bandwidth as well.
- In the case of a band-stop filter, there are two pass-bands: the lower frequency pass-band has bandwidth $\omega_b = \omega_{cl} - 0 = \omega_{cl}$; the upper frequency pass-band has $\omega_b = \infty - \omega_{ch} = \infty$, and, as for a high-pass filter, ω_{ch} is often given instead.
- In the case of a band-pass filter, $\omega_b = \omega_{ch} - \omega_{cl}$ (see Figure 12.7). But the two limits of the pass-band ω_{cl} and ω_{ch} are often given instead, as for band-stop filters.

So in practice bandwidth ω_b is frequently identified with the cut-off frequency ω_c (or frequencies, if there are more than one). Consequently, ω_b will be the frequency (or frequencies) where the gain crosses tolerance δ_p of the pass-band (see Figure 12.7 again). The value of δ_p should be defined according to the specifications of the particular situation where the filter is employed. Very often, the -3 dB criterion is used, i.e. the bandwidth ω_b of a filter $F(s)$ is defined as the frequency where its gain $|F(j\omega)|$ crosses -3 dB, or, in absolute value,

-3 dB criterion for the bandwidth

$$20 \log_{10} |F(j\omega_b)| = -3 \text{ dB} \Leftrightarrow |F(j\omega_b)| = 10^{\frac{-3}{20}} = 0.708 \quad (12.18)$$

So, with this criterion, 70% of the signal passes, i.e. there is an attenuation of 30%. (Of course, if this is not acceptable, either because 30% of attenuation is too much or because larger attenuations of the signal can be tolerated, a different criterion should be used.) The advantage of this criterion is that

- a first-order filter $F(s) = \frac{a}{s+a}$ has a -3 dB gain at the frequency of the pole, as we see by (11.90), and so $\omega_b = a$;
- a second-order filter with $\xi = \frac{\sqrt{2}}{2}$ and natural frequency $\omega_n = a$, given by

$$F(s) = \frac{a^2}{s^2 + 2 \frac{\sqrt{2}}{2} as + a^2} \quad (12.19)$$

has a $-20 \log_{10} 2 \frac{\sqrt{2}}{2} = -3$ dB gain at frequency $\omega_n = a$, as we saw in (11.105), and so $\omega_b = a$;

- and indeed all Butterworth filters of any order, not only those of first-order and second-order just mentioned, will have $\omega_b = a$.

When a bandwidth is mentioned, and no indication is given about the criterion used to define it, you can safely assume that the -3 dB criterion was employed.

Bandwidth of a plant in general

The notion of bandwidth can be extended to any plant. Again, strictly speaking, the bandwidth is the width of a frequency range, but in practice the limits of the range are given instead, or rather only one limit if the other is 0 rad/s or infinity. The bandwidth will thus be once more a frequency or frequencies that satisfy one of several criteria, according to what better suits the plant.

- If the largest value of the gain of the plant is 0 dB, then the -3 dB criterion can be used just as it is for filters.
- If the largest value of the gain is larger than 0 dB, the -3 dB criterion can also be used just the same.

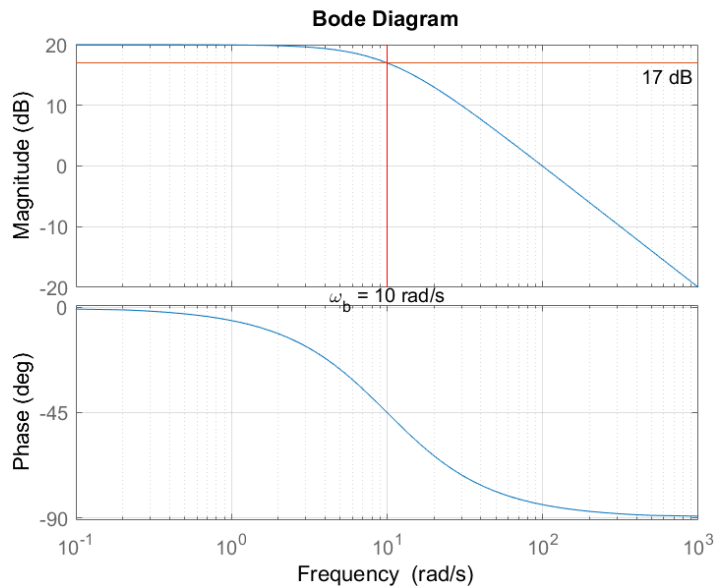


Figure 12.12: Bode diagram of a first-order plant with gain $g_{\max} = 20$ dB at low frequencies. Using the $g_{\max} - 3$ dB criterion for the bandwidth, the frequency of the pole is found.

- If the largest value of the gain is larger than 0 dB, the 0 dB criterion can be used instead. The bandwidth will then separate the frequencies where the input is amplified from those where the input is attenuated. In this case, the bandwidth will be just another name for the gain crossover frequency, according to Definition 10.9.
- If the largest value of the gain is some value g_{\max} dB, a criterion can be used by which the bandwidth is the frequency (or frequencies) where the gain crosses $g_{\max} - 3$ dB. See Figure 12.12 to see an example showing why this makes sense.
- The former criterion makes particular sense if $g_{\max} < 0$ dB, since the gain does not cross 0 dB and may not even cross -3 dB.
- The former criterion makes no sense if there are poles at the origin, since in that case $g_{\max} = +\infty$.
- Some other value that better suits the problem addressed can be used instead.

The 0 dB criterion is often useful, but since a plant's gain can be known with uncertainty, or undergo small changes, for safety the -3 dB criterion is used instead.

12.4 Signal characteristics

Having studied filters, let us now look again at AD/DA conversion:

- AD conversion consists in two changes:
 - the signal is sampled in time;
 - the amplitude of the signal is quantized.
- DA conversion consists in two changes:
 - the amplitude of the signal must be reconstructed;
 - the signal will become continuous in time.

Changes in time will be addressed below in Chapter 25. Here we will address changes in amplitude.

Digital signals are (almost always) binary, consisting in an integer number in base 2, i.e. they are a sequence of binary digits or **bits**, i.e. zeros or ones. A digital signal with n bits can assume 2^n values, ranging in decimal form from 0 to $2^n - 1$.

Values of an n-bit digital signal

Example 12.6. A 4-bit digital signal can assume $2^4 = 16$ values which are

binary	decimal
0000	0
0001	1
0010	2
0011	3
0100	4
0101	5
0110	6
0111	7
1000	8
1001	9
1010	10
1011	11
1100	12
1101	13
1110	14
1111	15

Notice that 16 in base 2 is 10000 and requires 5 bits to be represented. It is impossible to represent it with 4 bits. □

Remark 12.4. This is similar to what happens with an integer decimal number with n digits, which can assume 10^n values ranging from 0 to $10^n - 1$. With three decimal digits we can represent 999, but $10^3 = 1000$ would require 4 decimal digits. Of course, there are 1000 numbers from 0 (or 000, if you prefer) to 999. □

From base 2 to base 10

Converting integers from base 2 to base 10 can be done summing the powers of 2 corresponding to bits which are 1.

Remark 12.5. This is similar to the interpretation of a decimal integer number. We know that 60789, which has 5 decimal digits, is in fact

$$60789 = 6 \times 10^4 + 0 \times 10^3 + 7 \times 10^2 + 8 \times 10^1 + 9 \times 10^0 \tag{12.20}$$

The rightmost digit multiplies 10^0 and thus, if there are n digits, the leftmost digit multiplies 10^{n-1} . Binary numbers are even easier to handle, because bits are either 0 or 1. □

Example 12.7. Binary number 10011101, which has 8 bits, corresponds to decimal number

$$1 \times 2^7 + 0 \times 2^6 + 0 \times 2^5 + 1 \times 2^4 + 1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 = 128 + 0 + 0 + 16 + 8 + 4 + 0 + 1 = 157 \tag{12.21}$$

It is sometimes convenient to write the number like this:

order	7	6	5	4	3	2	1	0
bits	1	0	0	1	1	1	0	1
powers of 2	2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
weights	128	64	32	16	8	4	2	1

Thereafter, the weights corresponding to bits equal to 1 are summed. □

Example 12.8. Binary number 000111010110, which has 12 bits, corresponds to decimal number

order	11	10	9	8	7	6	5	4	3	2	1	0
bits	0	0	0	1	1	1	0	1	0	1	1	0
powers of 2	2^{11}	2^{10}	2^9	2^8	2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
weights	2048	1024	512	256	128	64	32	16	8	4	2	1

$$2^8 + 2^7 + 2^6 + 2^4 + 2^2 + 2^1 = 256 + 128 + 64 + 16 + 4 + 2 = 470 \quad \square \quad (12.22)$$

Remark 12.6. In the last example, there were useless zeros at the left of the last bit equal to 1. Of course, AD/DA converters work with a fixed number of bits, and thus only half of the values they can assume do not begin with a zero. \square

Converting integers from base 10 to base 2 can be done subtracting successively the largest possible power of 2. These powers will correspond to the bits that are equal to 1.

From base 10 to base 2

Example 12.9. Decimal number $m = 789$ can be converted to binary as follows. First we find all the powers of 2 until we exceed m .

2^{10}	2^9	2^8	2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
1024	512	256	128	64	32	16	8	4	2	1

Since $2^{10} > m$ we now know we will need 10 bits, corresponding to powers from 2^0 to 2^9 . Now we successively subtract the larger possible weight from the table above:

$$789 - 512 = 789 - 2^9 = 277 \quad (12.23)$$

$$277 - 256 = 277 - 2^8 = 21 \quad (12.24)$$

$$21 - 16 = 21 - 2^4 = 5 \quad (12.25)$$

$$5 - 4 = 5 - 2^2 = 1 \quad (12.26)$$

$$1 - 1 = 1 - 2^0 = 0 \quad (12.27)$$

The bits equal to 1 are those of the powers we used:

2^9	2^8	2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
1	1	0	0	0	1	0	1	0	1

Consequently $m = 1100010101_{(2)}$. If a larger number of bits is required, the number is padded with zeros at the left. For instance, if m is to be written using 16 bits, we will have 0000001100010101. \square

The number of bits in an AD/DA converter lets us know the **resolution** of its output, which is the smallest interval between two consecutive readings. *Resolution*

Example 12.10. An 8-bit AD converter receives signals ranging from 0 V to 5 V. Its resolution is $\frac{5-0}{2^8} = \frac{5}{256} = 0.0195 \text{ V} = 19.5 \text{ mV}$. \square

Example 12.11. A 12-bit DA converter outputs signals ranging from -10 V to 10 V . Its resolution is $\frac{10-(-10)}{2^{12}} = \frac{20}{4096} = 0.0049 \text{ V} = 4.9 \text{ mV}$. \square

The resolution can also be given referring to the variable which is being measured, in the case of AD conversion, or which is being output, in the case of DA conversion. This is fairly simple if transduction is linear. In the case of a sensor with linear transduction, the proportion between the variable of transduction and the measured variable is called **sensibility**. *Sensibility*

Example 12.12. Consider a sensor measuring distances from 0 cm to 50 cm and with an output ranging from 0 V to 5 V. Its sensibility is constant, and equal to $\frac{5-0}{50-0} = 0.1 \text{ V/cm}$. The sensor is connected to the 8-bit AD converter receiving signals from 0 V to 5 V of Example 12.10. The resolution of the distance measurement is $\frac{50-0}{2^8} = \frac{50}{256} = 0.195 \text{ cm}$. This can be graphically depicted as in Figure 12.13. \square

Example 12.13. Consider a sensor measuring temperatures from $20 \text{ }^\circ\text{C}$ to $120 \text{ }^\circ\text{C}$ and with an output ranging from 1.25 V to 7.5 V. Its sensibility is constant, and equal to $\frac{7.5-1.25}{120-20} = 0.0625 \text{ V/}^\circ\text{C}$. The sensor is connected to a 10-bit AD converter receiving signals from -10 V to 10 V . The resolution of the temperature measurement can be found in different ways:

- The range of the output of the sensor is $7.5 - 1.25 = 6.25 \text{ V}$. Consequently, we are using only $\frac{6.25}{20} = 0.3125$ of the $2^{10} = 1024$ values that the AD converter can output, i.e. we are using only $0.3125 \times 1024 = 320$ values. Thus the resolution is $\frac{120-20}{320} = 0.3125 \text{ }^\circ\text{C}$.

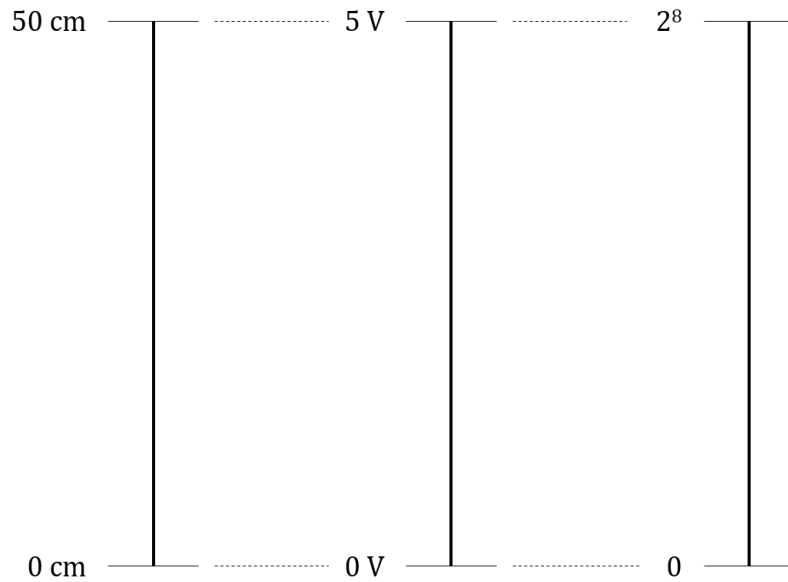


Figure 12.13: Relations in the measuring chain of Example 12.12.

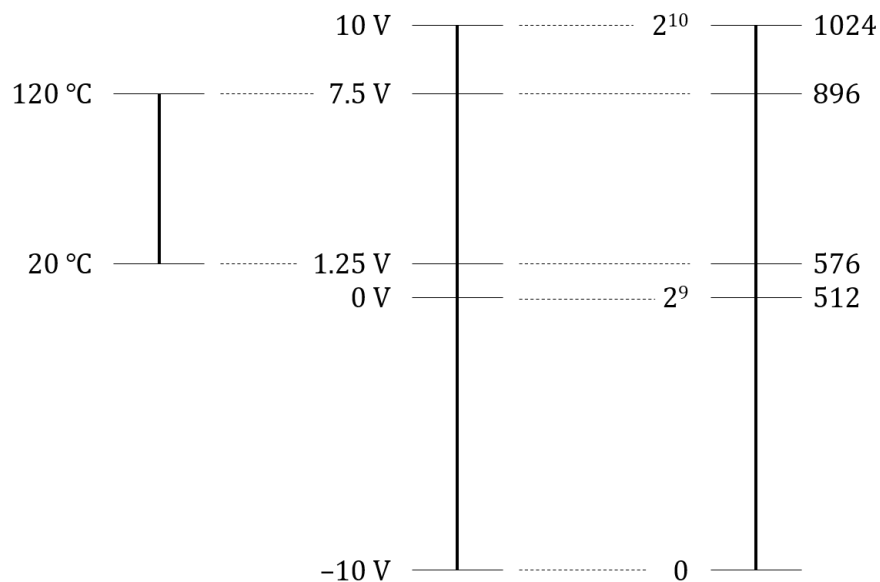


Figure 12.14: Relations in the measuring chain of Example 12.13.

- The resolution of the AD converter in Volt is $\frac{10 - (-10)}{2^{10}} = \frac{20}{1024} = 0.0195$ V. Hence the resolution in degree Celsius is $\frac{0.0195}{0.0625} = 0.3125$ °C.
- 7.5 V corresponds to output $\frac{7.5 - (-10)}{10 - (-10)} \times 1024 = 896$. 1.25 V corresponds to output $\frac{1.25 - (-10)}{10 - (-10)} \times 1024 = 576$. Thus the resolution is $\frac{120 - 20}{896 - 576} = 0.3125$ °C.

Of course, all these ways of reasoning consist in fact in the same calculation and give the same result. See Figure 12.14. □

*AD/DA
rounds down*

conversion In the previous example, outputs were found with rules of three. When the result is not integer, it must be rounded down, because both AD and DA converters work as shown in Figure 12.15 (in the next section we will see why). Notice that, as a result:

- in AD conversion, when the input is the maximum voltage admissible V_{\max} , the output should be 2^n , where n is the number of bits, but, since it is impossible to represent this number with n bits, the output is $2^n - 1$ instead;

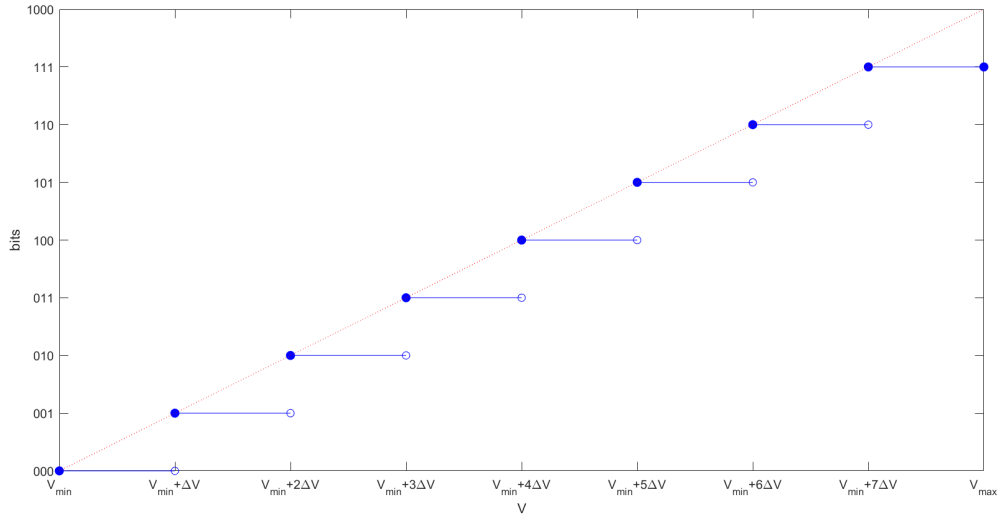


Figure 12.15: Blue line: how AD and DA conversions take place, for a 3-bit converter working with voltages in the range from V_{\min} to V_{\max} ; the resolution is $\Delta V = \frac{V_{\max} - V_{\min}}{8}$. The extension to any other number of bits is immediate.

- in DA conversion, since the largest input is $2^n - 1$, the output never reaches V_{\max} , but only $V_{\max} - \Delta V = V_{\max} - \frac{V_{\max} - V_{\min}}{2^n}$.

Example 12.14. A 16-bit AD converter working in the ± 10 V range receives 2 V as input. The output is

$$\left\lfloor \frac{2 - (-10)}{10 - (-10)} 2^{16} \right\rfloor = \lfloor 39321.6 \rfloor = 39321 \quad (12.28)$$

Converting this to binary form,

$$39321 - 2^{15} = 6553 \quad (12.29)$$

$$6553 - 2^{12} = 2457 \quad (12.30)$$

$$2457 - 2^{11} = 409 \quad (12.31)$$

$$409 - 2^8 = 153 \quad (12.32)$$

$$153 - 2^6 = 25 \quad (12.33)$$

$$25 - 2^4 = 9 \quad (12.34)$$

$$9 - 2^3 = 1 \quad (12.35)$$

$$1 - 2^0 = 0 \quad (12.36)$$

we get 1001100101011001. \square

Example 12.15. A 6-bit DA converter has outputs in the -5 V to 5 V range and receives 110001 as input. Since $110001_{(2)} = 2^5 + 2^4 + 2^0 = 32 + 16 + 1 = 49$, the output will be

$$-5 + \frac{2^5 + 2^4 + 2^0}{2^6} (5 - (-5)) = -5 + \frac{32 + 16 + 1}{64} 10 = \frac{490}{64} - 5 = 2.656 \text{ V} \quad \square \quad (12.37)$$

Knowing the resolution is important to calculate the **precision**:

Precision

- the precision ε_{\max} of a measuring chain in the maximum absolute value of the error ε between the measured value \hat{x} and the exact value of the variable that is being measured x , that is,

$$\varepsilon_{\max} = \max_{\hat{x}} |\varepsilon| = \max_{\hat{x}} |\hat{x} - x| \quad (12.38)$$

- the precision ε_{\max} of an actuation chain is the maximum absolute value of the error ε between the control action applied \tilde{u} and the desired control action u found by the controller, that is,

$$\varepsilon_{\max} = \max_{\tilde{u}} |\varepsilon| = \max_{\tilde{u}} |\tilde{u} - u| \quad (12.39)$$

The precision of the chain, being a maximum value of an absolute error, is the sum of the precision of its components.

Systematic vs. stochastic errors

Indeed, sensor and actuator accuracy is affected by errors, which can be classified as **systematic** when their value remains the same for identical conditions, or **stochastic** when they are a random variable. Systematic errors lead to a **bias** in measurements. Stochastic errors corresponding to random variables that do not have a zero mean also lead to bias in measurements. We will study stochastic systems in detail in Part VIII of these Lecture Notes.

Accuracy as % of full-scale reading

Example 12.16. Suppose that the sensor from Example 12.12 has an **accuracy** of 5 mm. Since the range is 50 cm, this is often indicated as $\frac{5 \text{ mm}}{50 \text{ cm}} = 1\%$ of full-scale reading. This means that the accuracy is $1\% \times 50 \text{ cm} = 5 \text{ mm}$.

Since the resolution of the AD conversion is $0.39 \text{ cm} = 3.9 \text{ mm}$, if there are no other sources of noise present, the precision of the measuring chain shall be $5 + 3.9 = 8.9 \text{ mm}$. \square

Example 12.17. Analog measurements have resolution and precision too. Consider the weighing scale in Figure 12.16. Mass can be measured in the 5 kg to 100 kg range, with a resolution of 0.1 kg, and a precision of 0.1 kg. There is no finer resolution because the graduation has ten marks per kilogram. The precision is a result of the characteristics of the device. \square

Example 12.18. In Example 12.17 the resolution and the precision have the same value, but this is often not the case. Figure 12.17 shows luggage scales with a resolution of 1 g, but with a precision of 5 g or 10 g depending on the range where the measurement falls. (Varying precisions are found for some types of sensors, especially because of non-linearities, as we shall see in Chapter 13.) \square

Remark 12.7. It makes sense to have a resolution equal to the precision, as in Example 12.17, in which case all figures of the measurement are certain; and it makes sense to have a resolution higher than the precision, as in Example 12.18, in which case the last figure of the measurement is uncertain. It would make no sense to have a resolution more than 10 times larger than the precision, since in that case at least the last figure of the measurement would have no significance. It would also make no sense to have a resolution coarser than the precision, since the capacities of the sensor would be wasted. \square

In Examples 12.12 and 12.16 the range of outputs of the sensor S is precisely the range of inputs of the AD converter C . If $S \neq C$ there are three possible situations:

- If $S \not\subset C$, then it is necessary to modify S , adding (or subtracting) some value, or multiplying it by some gain, or both, until $S \subset C$. We will see in the next section that this can be easily done with op-amps.
- If $S \subset C$, but S is clearly smaller than C , and the resolution of the AD conversion has a significant weight in the precision of the measurement chain, then S should be modified as well.
- If $S \subset C$, and the resolution of the AD conversion is fine, there is no need to do anything.

A similar reasoning applies to DA conversion in the actuation chain.

Example 12.19. A sensor has an output in the 0 V to 10 V range, and will be connected to a 24 bit AD converter that receives values in the 0 V to 5 V range. The output of the sensor must be halved. Otherwise, the AD would overflow and its output would saturate. See Figure 12.18. \square

Example 12.20. A sensor has an output in the 0 V to 10 V range, and will be connected to a 24 bit AD converter that receives values in the -5 V to 5 V range. 5 V must be subtracted to the output of the sensor. See Figure 12.19. \square

Example 12.21. An electric motor receives an input in the 0 V to 10 V range, corresponding to rotation speeds between 1 rpm and 100 rpm. It will be connected to an 8 bit DA converter that has an output in the 0 V to 5 V range.



Figure 12.16: Weighing scales once used in the Bełchatów coal mine, Poland.



Figure 12.17: Luggage scales.

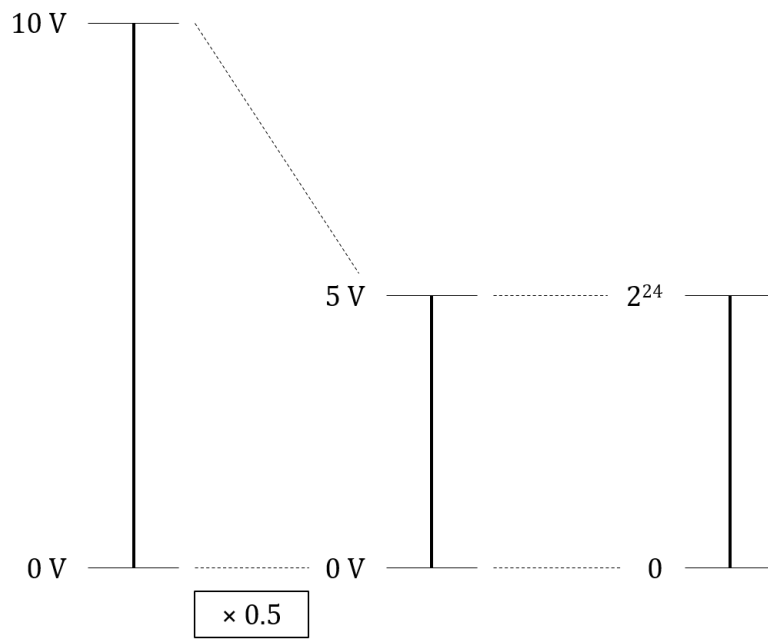


Figure 12.18: Relations in the measuring chain of Example 12.19.

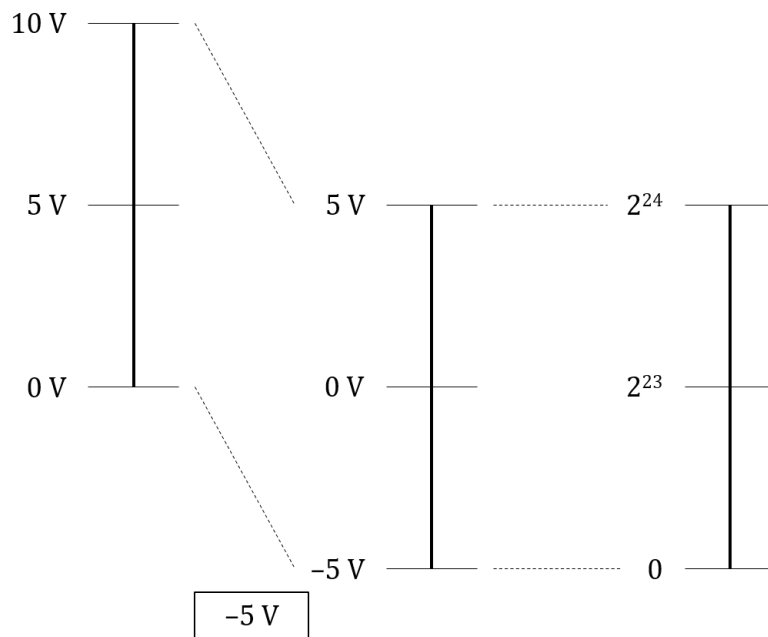


Figure 12.19: Relations in the measuring chain of Example 12.20.

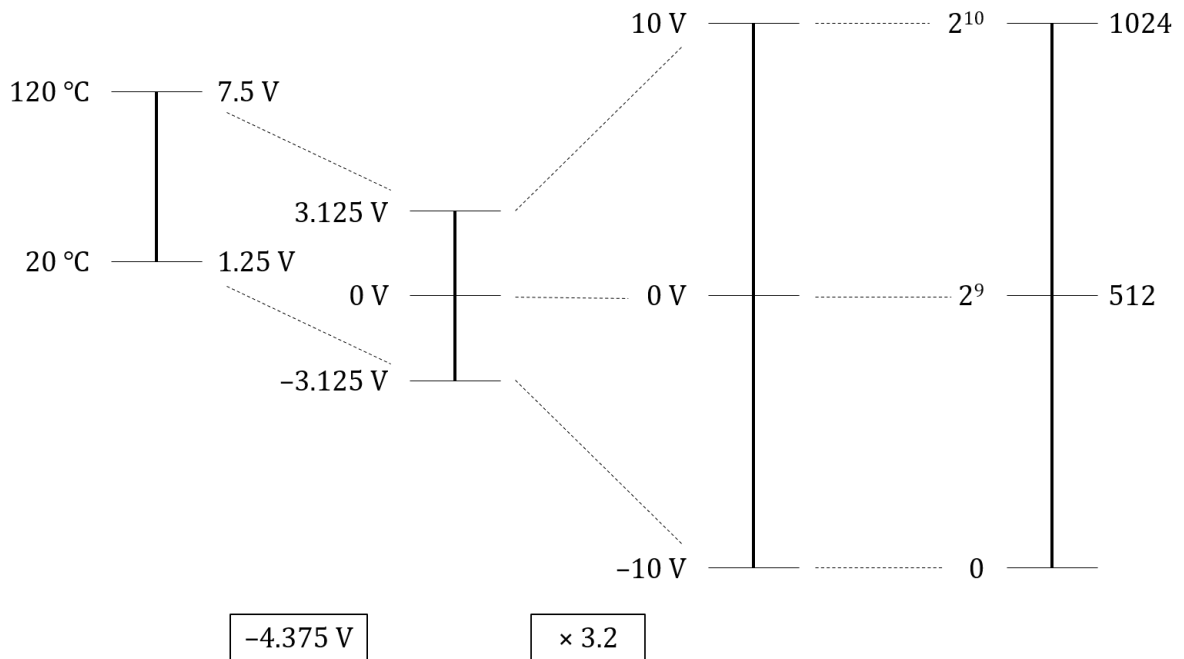


Figure 12.20: Relations in the measuring chain of Example 12.23. Compare this with the original situation in Figure 12.14.

We do not want the motor to run faster than 45 rpm. Thus control actions larger than 4.5 V are not needed, and the output of the DA converter can be connected directly to the motor (likely through a power amplifier, as we will see in Chapter 14). \square

Example 12.22. If the temperature sensor from Example 12.13 has an accuracy of 3 °C, then the precision of the measurement is 3.3125 °C. The contribution of the AD converter for this is 0.3125 °C and is not very relevant when compared to sensor accuracy. So the sensor can be directly connected to the AD converter. \square

Example 12.23. If the temperature sensor from Example 12.13 has an accuracy of 0.1 °C, then the precision of the measurement is 0.4125 °C. The contribution of the AD conversion is significant, and if we need a precision of 0.2 °C it is clear that it is the resolution of the AD conversion that is preventing this specification from being fulfilled. The resolution is poor because only 320 of the 1024 values that the AD can output are being used. If all values were used, the precision of the measurement would be $0.1 + \frac{120-20}{1024} = 0.1977 < 0.2$ °C as required. So the output of the sensor must be corrected:

- its range is from 1.25 V to 7.5 V, with an amplitude of 6.25 V and a mean value of 4.375 V;
- we need it to range from -10 V to 10 V, with an amplitude of 20 V and a mean value of 0 V;
- so first its mean should be corrected by subtracting 4.375 V;
- then its range must be corrected through an amplification of $\frac{20}{6.25} = 3.2$ times.

See Figure 12.20. \square

12.5 Op-amp implementation of signal conditioning

Signal conditioning consists in the several operations underwent by a signal in the measuring chain or in the actuation chain — filtering, correction of the mean, amplification, etc.. They can be implemented using op-amps in the several configurations we studied in Chapter 5.

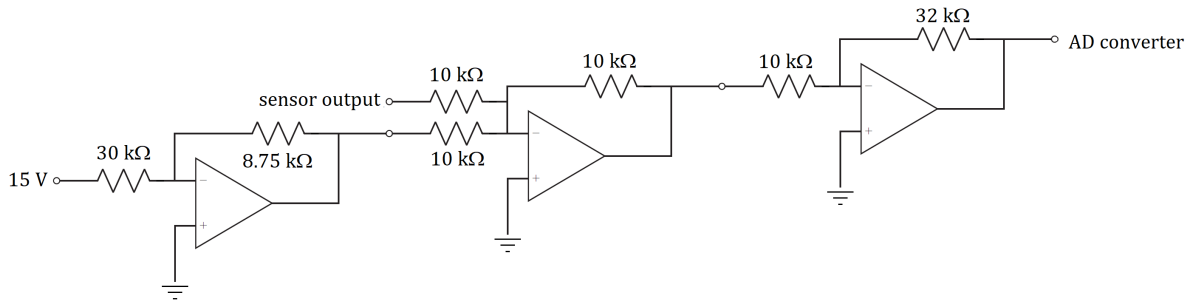


Figure 12.21: Circuit of Example 12.24, implementing the relations of Example 12.23.

Example 12.24. The operations in Example 12.23 can be implemented with op-amps with ± 15 V supply tensions as shown in Figure 12.21. Notice that:

- Because a -4.375 V is unlikely to be directly available, it was obtained from the 15 V tension which must be available to supply the op-amps.
- Care must be taken not to exceed supply tensions anywhere. That is why it is not possible to multiply the sensor output $S \in [1.25 \text{ V}, 7.5 \text{ V}]$ by 3.2 first, and only then correct the mean value subtracting 14 V. Apparently the final result would be the same, since of course $3.2(x - 4.375) = 3.2x - 14$. But the first operation would require voltages in the range

$$3.2 \times [1.25 \text{ V}, 7.5 \text{ V}] = [4 \text{ V}, 24 \text{ V}] \not\subset [-15 \text{ V}, 15 \text{ V}] \quad (12.40)$$

and so the op-amps would saturate.

- The circuit in Figure 12.21 is not the only possible way of implementing the desired operations. A weighted subtracting op-amp could have been used instead. (Try to draw the corresponding circuit.)
- In particular, resistor values could change as long as the proportions for each op-amp are kept. However, resistor values should not be too small or too large.
- Resistor values that cannot be found off the shelf can be obtained combining resistors in series (or in parallel), or with a potentiometer. The later option allows calibrating the measuring (or actuation) chain, i.e. changing the parameters of the components (in this case, resistance values) so that the value read is exact in situations where the accurate value to be read is known. *Calibration*
- Small resistor values mean high currents, since from Ohm's law $I = \frac{U}{R}$. And high currents mean significant energy dissipation by Joule effect, since the power dissipated is given by $P = RI^2 = UI$; this also means that the resistances would get very hot.
- Another reason to avoid small resistor values is that jump wires, cables, and other components have resistance too; such resistances are small, but cannot be neglected if resistors have small values. As a consequence, proportions of resistances around the op-amps would not be the desired ones, and signal conditioning would be inaccurate.
- Large resistor values are also undesirable, since currents would be small. In that case, noise affecting currents would hardly be neglectable, and measurements would be unreliable. \square

Filters can be implemented analogically with op-amps. Several configurations that can be used for filters have appeared in Exercises 1 and 5 of Chapter 5; see also Exercises 6 and 7 in this chapter. These filters filter signals before AD conversion. We will see in Part V how to implement (numerically) filters after AD conversion, i.e. filters that receive the signal already discretised in time.

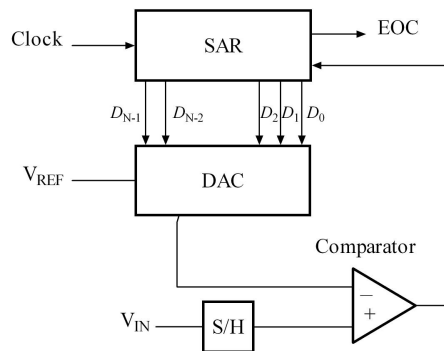


Figure 12.22: Successive approximation AD converter with N bits (source: Wikimedia). S/H (sample and hold) denotes discretisation in time; DAC is the digital to analog converter; and SAR (successive approximation register) is the microprocessor's generator of binary numbers according to the binary search algorithm.

AD/DA conversion is always done using commercially available converters, but it is important to know how it is done using op-amps, since this makes clearer why it behaves how it does:

- DA conversion is obtained with a sequence of summing amplifiers, as seen in Exercise 9 below;
- AD conversion is obtained with one or more comparator op-amps. The input is compared with the values of the voltages of possible binary outputs (the full points in Figure 12.15). The output is the largest binary number that is not larger than the input. That is why AD conversion results in a step-like behaviour as seen in Figure 12.15, rounding the output down. While there are different possible configurations, the most common ones are:
 - There are as many comparators as possible outputs. This is fast, since all comparisons are done simultaneously. On the other hand, if the number of bits n is high, 2^n op-amps take lots of space and energy.
 - There is only one comparator, that sequentially compares the input with a variable voltage. This voltage is obtained, using a DA converter, from a binary number sent by a microprocessor. The comparison could be done beginning with 0 and then going up one bit at a time, but this takes too much time; in practice, the microprocessor usually runs a binary search algorithm (which you may know from computer programming) instead:
 1. If the conversion has n bits, the first comparison is always with 2^{n-1} , which corresponds to the middle of interval $[V_{\min}, V_{\max}]$.
 2. If the input is larger, the output must be in the $[2^{n-1}, 2^n - 1]$ range. If the input is smaller, the output must be in the $[0, 2^{n-1} - 1]$ range.
 3. The midpoint of this interval where we now know that the output must be is used for a new comparison.
 4. This goes on until it is clear that the interval where the output must be has only one point; i.e. until it is clear that input V verifies $V \geq m$ and $V < m + 1$. Then m is the output.

These are called successive approximation AD converters. See Figure 12.22.

Glossary

Ora menganava tanto
 que cuydey que ereis vos santo
 & vos falais castellano

Gil VICENTE (1465? — †1536?), *Floresta deganos* (1536)

accuracy precisão
active filter filtro ativo
actuation chain cadeia de atuação
AD converter conversor AD
band-pass (filter) (filtro) passa-banda
band-stop, band-reject (filter) (filtro) corta-banda
bandwidth largura de banda
bias enviesamento
calibration calibração
cut-off frequency frequência de corte
DA converter conversor DA
distortion distorção
filter filtro
high-pass (filter) (filtro) passa-alto
low-pass (filter) (filtro) passa-baixo
measuring chain cadeia de medida
measuring range gama de medida
pass-band banda (de frequências) de passagem
passive filter filtro passivo
precision precisão
repeatability repetibilidade
resolution resolução
signal conditioning condicionamento de sinal
sensitivity sensibilidade
stochastic error erro estocástico
stop-band banda (de frequências) de corte, banda de rejeição
successive approximation aproximação sucessiva
systematic error erro sistemático
transduction transdução
transition band banda (de frequências) de transição

Exercises

- The speedometer of a car's dashboard receives a signal in the $[0, 2]$ V range and shows speed values in the $[0, 200]$ km/h range. It will be connected to a sensor that measures speeds in the $[0, 50]$ m/s range, providing a reading in the $[0, 50]$ mV range. Design the signal conditioning needed to connect the sensor to the dashboard's speedometer.
- A submarine has a sonar that creates a sound and detects its reflection by an object. The sound travels from the submarine to the object and back, and then is detected. The speed of sound in water is 1500 m/s.
 - How is the distance d to the object related to the time t between the emission of the sound and the detection of its reflection?
 - The sound detector that measures t returns a tension given by 100 mV per second. It will be connected directly to a display that receives signals in the $[0, 5]$ V range. What is the largest distance d that can be shown, before the display saturates?
- An accelerometer outputs 14 mV per g , where $g = 9.8 \text{ m/s}^2$. Design a signal conditioning to convert this into 0.25 V per m/s^2 .
- A quadcopter has a sensor to measure its height. The maximum frequency of this signal is 10 Hz. There is noise at frequencies of 1 kHz and higher. Design a filter with at least 37 dB attenuation for noise, while letting at least 99% of the signal pass.
- A sensor reads a signal given by $y(t) = \cos(10t)$. A power amplifier introduces noise estimated as $d(t) = 0.5 \cos(10^4 t)$. In order to reduce the amplitude of the noise to 0.01, the filter in Figure 12.23 was employed.

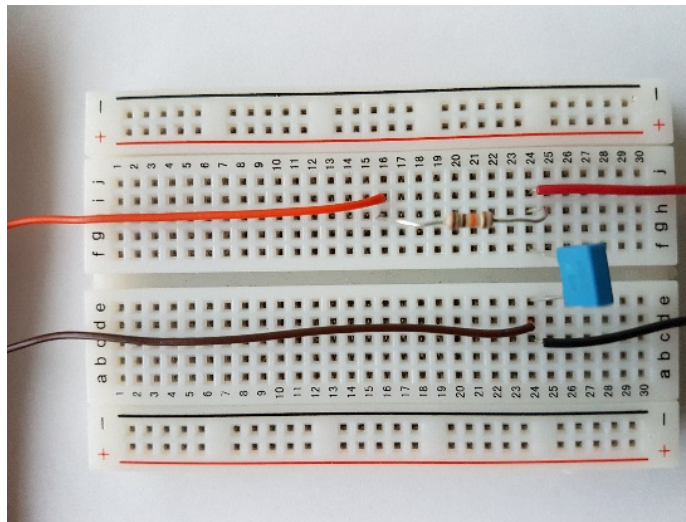


Figure 12.23: Circuit with the filter of Exercise 5. The sensor is connected to the left of the breadboard, and the filtered signal is read to the right (source: Professor Carlos Carneira).

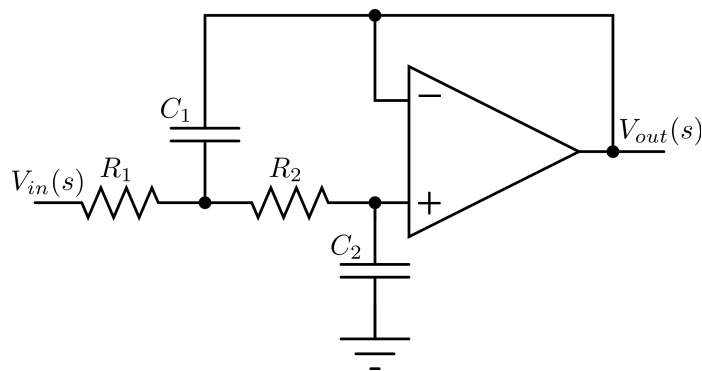


Figure 12.24: Circuit with the filter of Exercise 6.

- (a) The resistance is $R = 10 \text{ k}\Omega$ and the capacitor is $C = 6 \mu\text{F}$. Explain why the filter is not suitable for its purpose.
 - (b) What values of R and C would be suitable?
6. Consider the filter in Figure 12.24.
 - (a) Find transfer function $\frac{V_{out}(s)}{V_{in}(s)}$.
 - (b) What kind of filter is this?
 - (c) What values of R_1 , R_2 , C_1 and C_2 would you use to obtain a cut-off frequency of 100 Hz?
 - (d) Plot the resulting Bode diagram.
 - (e) Replace resistors by capacitors and vice-versa, and answer again all the questions above.
 7. Repeat the last exercise, using the filters with the circuits in Figure 12.25.
 8. An 8-bit AD converter receives values in the in the $[0, 10]$ V range.
 - (a) What is the converter's input when its output is 10101001?
 - (b) What is the converter's input when its output is 01010111?
 - (c) What is the converter's output when its input is 3 V?
 - (d) What is the converter's resolution?
 9. Consider the DA converter in Figure 12.26.

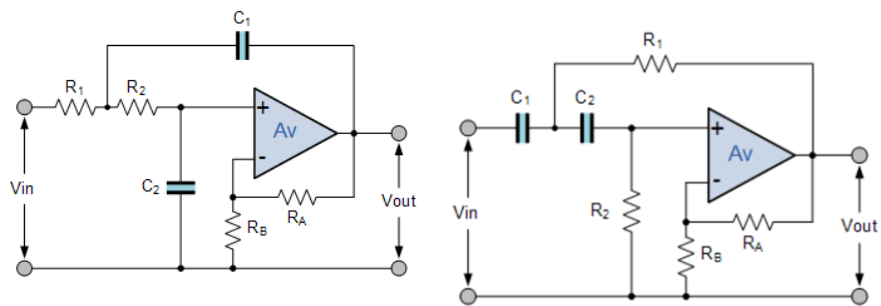


Figure 12.25: Circuit with the filters of Exercise 7.

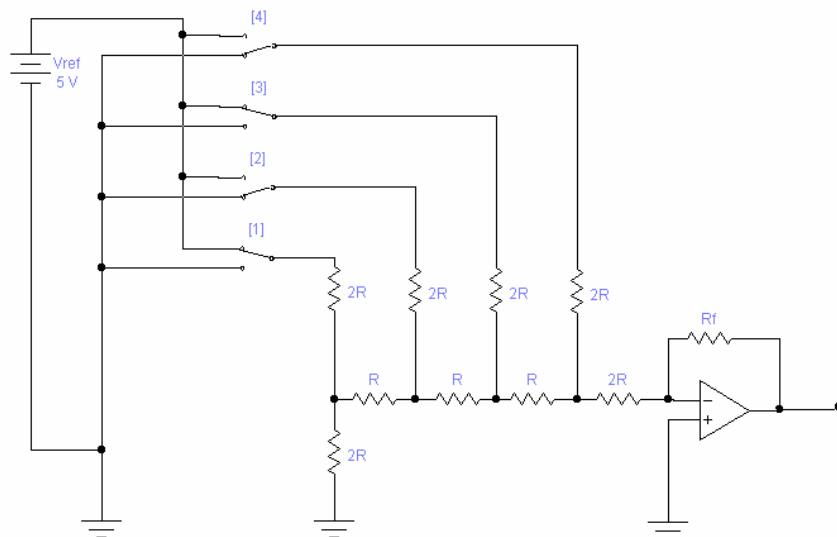


Figure 12.26: DA converter of Exercise 9.

- (a) What is the converter's output when its input is 1001?
 - (b) What is the converter's output when its input is 0101?
10. A distance sensor with a sensibility of 0.1 V/mm measures distances in the range from 20 mm to 120 mm. The reading will be read by an AD converter that receives signals in the $[-5, +5] \text{ V}$ range. Project a signal conditioning circuit with only one OpAmp, supplied with $\pm 15 \text{ V}$ tensions; assume that you have a DC $+5 \text{ V}$ tension available.

Chapter 13

Sensors

Once outside the town, Stomper led them into a thick sedge and bade them to be small and quiet lest they be seen by Sorhed's agents, who would soon revive and mount the hunt.

The party was still panting when sharp-eared Arrowroot adjusted the volume on his hearing aid and laid his head to the ground.

"Hark and lo!" he whispered, "I do hear the sound of Nine Riders galloping nigh the road in full battle array." A few minutes later a dispirited brace of steers ambled awkwardly past, but to give Stomper his due, they did carry some rather lethal-looking antlerettes.

"The foul Nozdrul have bewitched my ears," mumbled Stomper as he apologetically replaced his batteries, "but it is safe to proceed, for the nonce." It was at that moment that the thundering hooves of the dreaded pig riders echoed along the road.

Henry N. BEARD (1945 — ...), Douglas C. KENNEY (1946 — †1980), *Bored of the Rings* (1969), III

In this chapter we will study how sensors for different types of physical quantities work. Most sensors nowadays transduce what they are measuring into a voltage or a current, which will then be acquired electronically. Even if the signal is not recorded, but merely shown to a human user, such sensors are employed with a digital display. As a consequence, measuring chains need, almost always, to be supplied energy to function.

Sensors used to display a reading on a scale, as in Figure 12.2, are not found today as often as they once were. In any case,

- if a sensor transduces the quantity measured into an angular (or linear) position, easily shown on a scale, a further transduction of the angle into a voltage (or a current) allows acquiring the measurement electronically;
- if a sensor transduces the quantity measured into a voltage (or a current), a further transduction into an angle (or a linear position) allows showing the reading on a scale.

In either case, the measurement will be based on more than one transduction. This is usual: as we will see, there are sensors with working principles based upon more than two transductions.

The output of some sensors is already digital; they do not require AD conversion.

You are strongly advised to search the Internet for websites of manufacturers and sellers of the different types of sensors, and check the technical documentation available. This is the best way to get acquainted with the characteristics of the different sensors: range, resolution, precision, bandwidth, type of output, type of supply, repeatability, linearity, temperature or humidity limitations — and if possible check yet another very important non-technical characteristic which is the price.

The working principle of some sensors uses two or more transductions

Important characteristics of sensors

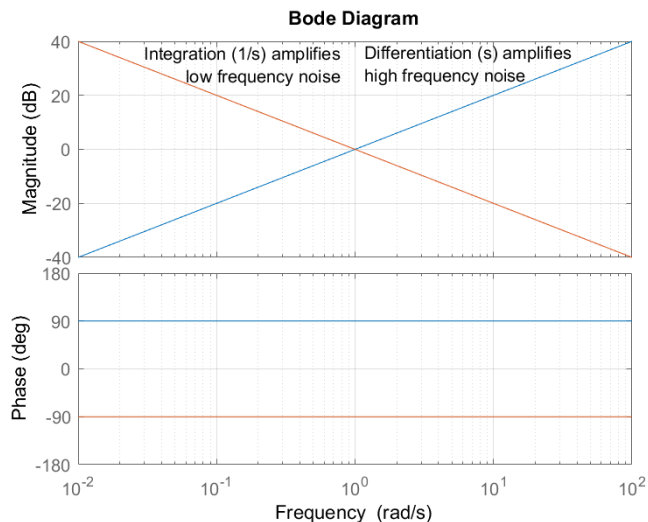


Figure 13.1: Bode diagrams of s (differentiation) and of $\frac{1}{s}$ (integration).

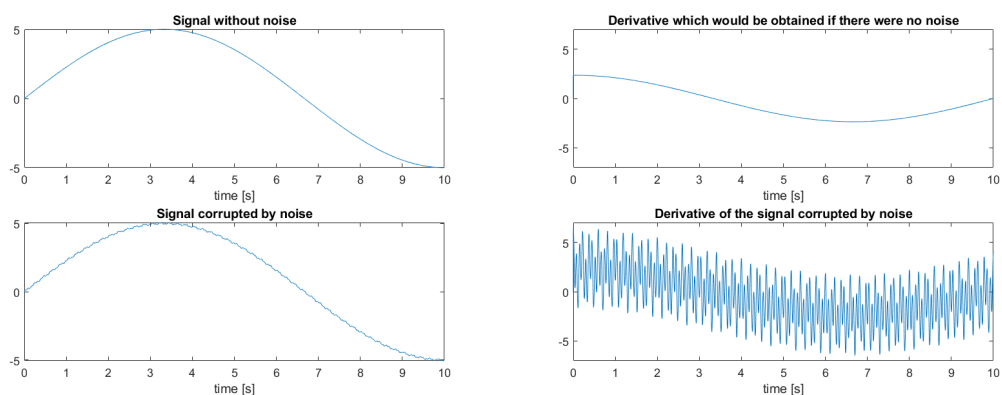


Figure 13.2: Left: a signal corrupted by high frequency noise. Right: the derivative we would like to have, and what we really get when we differentiate without bothering to filter.

13.1 Position, proximity, velocity and acceleration sensors

Finding x , \dot{x} and \ddot{x} from one another

Position, velocity and acceleration can be found from one another. This means that a sensor for one of these variables can be used to find the others. This is done in practice, but is not without problems:

- Velocity can be found differentiating position, and acceleration can be found differentiating velocity or differentiating position twice. But differentiation amplifies high-frequency noise: see the Bode diagram in Figure 13.1, and an example of the effect in Figure 13.2. Amplified high-frequency (i.e. short period) noise is felt at once.
- Velocity can be found integrating acceleration, and position can be found integrating velocity or integrating acceleration twice. But integration amplifies low-frequency noise: see the Bode diagram in Figure 13.1, and an example of the effect in Figure 13.3. Amplified low-frequency (i.e. large period) noise is felt only after a while.

Relation between proximity and position sensors

Proximity sensors return a binary variable related to position x : the sensor detects if $x < x_{\max}$, where x_{\max} is some threshold below which presence is detected. Position sensors can be used as proximity sensors: it is merely a question of comparing the measured distance x with the desired threshold. Several proximity sensors can be used to measure a distance, as seen in Figure 13.4; the resolution will be the distance between two consecutive sensors.

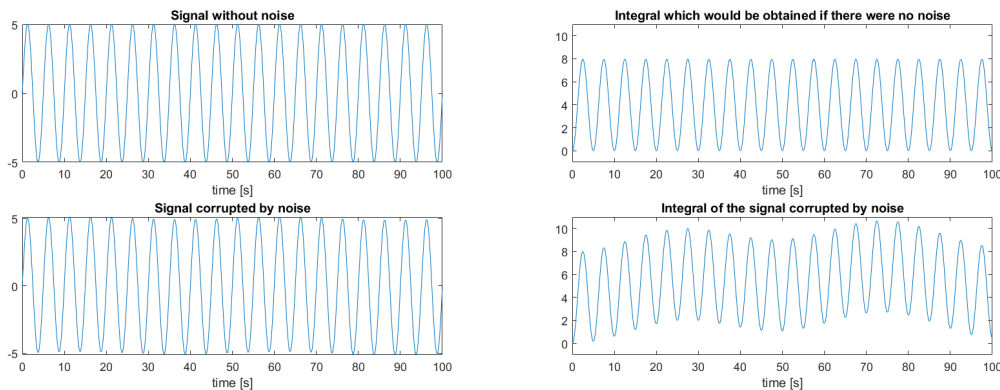


Figure 13.3: Left: a signal corrupted by low frequency noise. Right: the integral we would like to have, and what we really get when we integrate without bothering to filter.

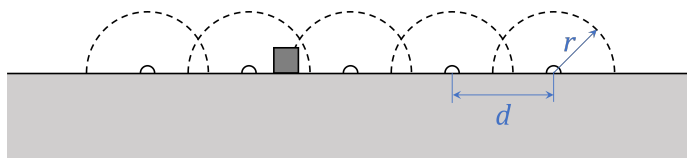


Figure 13.4: Several proximity sensors with range r , placed at a distance d from each other, can be used to measure position (in this case, the position of the dark square), if $d < 2r$.

This distance can never be larger than x_{\max} , and in practice should probably be inferior; otherwise it is possible that an object will not be detected by any sensor.

Proximity sensors can also be easily used to measure the angular velocity of non-homogeneous bodies. The example of a cog is shown in Figure 13.5: the sensor detects the proximity of a tooth; the angular velocity is found differentiating with respect to time the value of a counter that accumulates the number of teeth.

Example 13.1. The cog of Figure 13.5 has 28 teeth. A proximity sensor has the output shown in that Figure; i.e., 14 teeth could be counted during 1 s. Thus, the cog rotates at $\frac{14}{28} = 0.5$ Hz, i.e. $0.5 \times 2\pi = \pi$ rad/s, or still $0.5 \times 60 = 30$ rpm.

Usually the values of a counter would be known instead of the sensor output. For instance, the counter would have 753 at $t = 31$ s, and 767 at $t = 32$ s; thus, $\omega = \frac{767-753}{32-31} = 0.5$ Hz = π rad/s = 30 rpm. \square

As we saw in Chapter 4, position, velocity and acceleration can be linear or angular. Sensors for linear quantities can be used for angular quantities and vice-versa, using e.g. a rack and pinion (remember Figure 4.15) or a roller wheel. We will now study the more frequently found working principles of these types

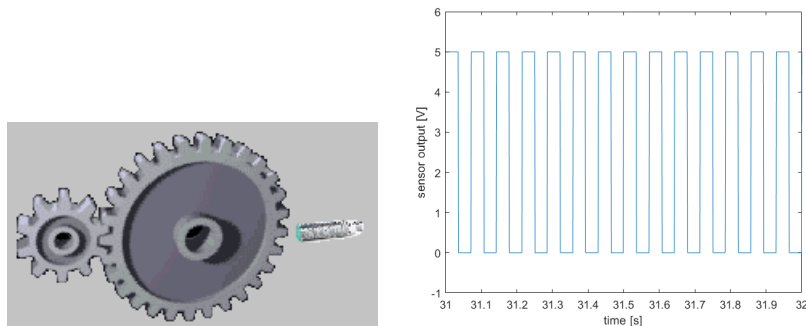


Figure 13.5: Left: a proximity sensor detects the teeth of a cog (source: adapted from Wikimedia). Right: output of the proximity sensor for Example 13.1.

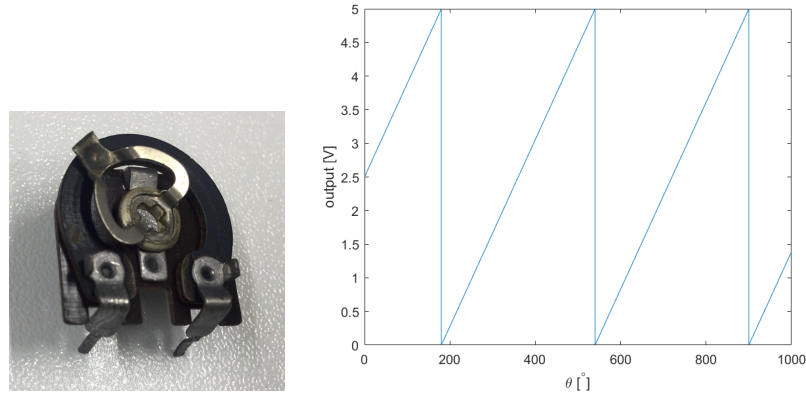


Figure 13.6: Left: rotary potentiometer. Right: output of a rotary potentiometer with one terminal grounded and another at 5 V.

of sensors. Some only work for linear quantities; others only work for angular quantities; others work for both.

13.2 Working principles of position sensors

The following sensors measure both linear and angular positions.

Potentiometric sensors

- **Potentiometric sensors** consist in a potentiometer with the terminals at fixed voltages. The resistance is uniform, and so the position of the slider is proportional to the resistance between the slider and one of the terminals; see (5.2) and Figure 5.3 again. These sensors use to be relatively cheap, but, because the slider must move and overcome friction, they are subject to wear and there are limitations to the velocity of the slider. In other words, *the bandwidth of the sensor is limited* (i.e. the frequency of the changes in the measured position cannot be very high).

Rotary potentiometers

Rotary potentiometers have the resistance shaped in a round form, as in Figure 13.6. In this case, for a multi-turn measurement, the output will have a discontinuity when the cursor passes from one terminal to the other.

- **Encoders** are more often used to measure angular position, though linear encoders exist as well. A rotary encoder is a sensor consisting in a wheel with perforations (shaped as slits), disposed in such a way that the light of a LED on one side of the wheel may be detected by a light sensor (which we will address below in Section 13.6) on the other side, if the hole is aligned with both the LED and the sensor. Instead of a perforated wheel, it is possible to have a wheel with a surface which either reflects light or not; in this case, both the LED and the light sensor are on the same side of the wheel. Because the encoder output depends on light sensors detecting light or not, it is digital by construction. There are two main types of encoders:

Incremental encoders

- **Incremental encoders**, or **relative encoders**, have a row of n slits along the rim (or non-reflective dark spots, though incremental encoders more often have slits). As the wheel turns, the output of a light sensor mounted as seen in Figure 13.7 will be a square wave. (More precisely, if the rotation speed is constant, the output will be a square wave. Otherwise, the up and down steps will not be evenly distributed in time.) A counter counts the number m of peaks in this wave, and thus, since the counter was reset, the angle θ that the wheel has described is

$$\theta = \frac{m}{n} 360^\circ \quad (13.1)$$

If the wheel always turns in the same sense of rotation, this is enough; otherwise, the counter just keeps incrementing its value even when

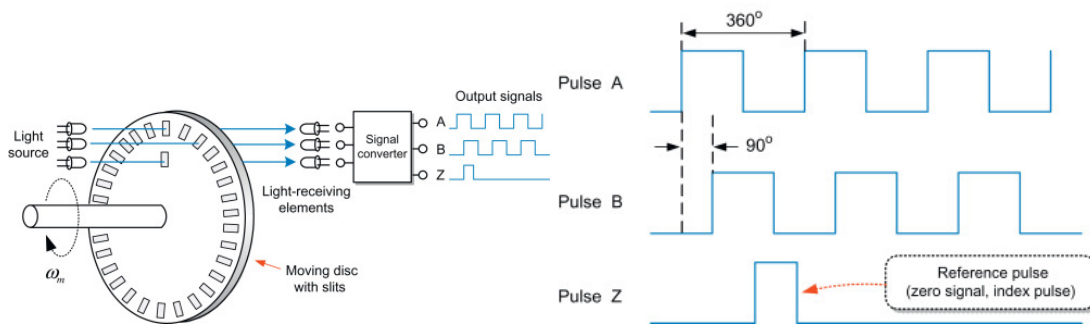


Figure 13.7: Left: incremental encoder. Right: output of the light sensors. (Source: ISBN 9780128123195.)

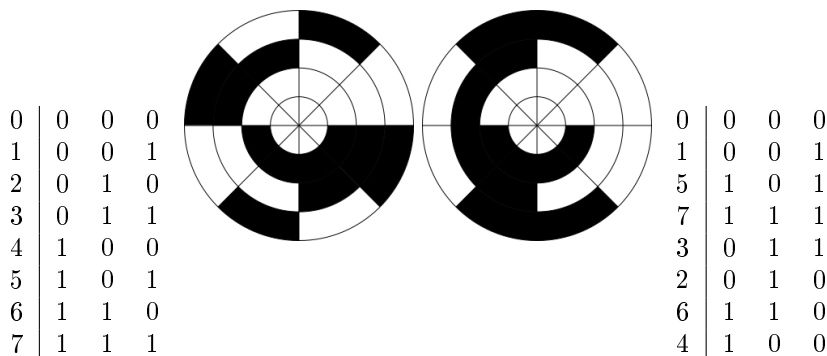


Figure 13.8: Left: 3-bit encoder wheel with binary numbers. Right: 3-bit encoder wheel with Gray code. (Source: Wikimedia.) Notice how on the left all three bits may change at one time, while on the right, by construction, only one changes at a time.

the shaft rotates *back*. To know the sense of rotation, two different light sensors are needed. They will either be slightly misaligned so that they will not detect a slit at the same time, or correspond to two different rows of misaligned slits along the rim. In this way, there will be two square waves, and knowing which of them goes up first it is possible to know if the wheel turns clockwise or counter-clockwise, i.e. if the count should be incremented or decremented. In any case, the rotation is measured from the position of the shaft when the counter was reset. Some encoders include a third sensor, to detect a single slit that will correspond to 0° ; but this will only work if that sensor is ever activated (small oscillating shaft rotations may prevent that). And, in any case, the position is found integrating — or rather, since a digital value is used, summing — a measurement.

- **Absolute encoders** give the absolute value of the angular position of the shaft. The wheel is divided into n concentric annuli, and into 2^n sectors, as seen in Figure 13.8. Each sector has its n sectors of annuli painted so as to correspond to a different number in base 2. (Slits would prevent a smooth transition of the reading from one sector to the next.) There are n sensors, and the binary number formed by the readings of the sensors gives the angular position of the shaft, with resolution $\frac{360^\circ}{2^n}$. Numbers may follow in numerical order, but in this way it is possible that, when the reading changes from one sector to the next, different light sensors will change its output at slightly different instants. Consequently, for a short period of time, the output will be wrong. Thus, it is usual to arrange the binary numbers in such a way that only one bit changes at a time. This arrangement is known as *Gray code* and shown in Figure 13.8.

Absolute encoders

Gray code

The encoders described above are *optical encoders*, by far the most common. Magnetic encoders are similar but replace the LED and the light sensor by magnetic pulses. Instead of slits in the wheel, or zones that alternately reflect and do not reflect light, the wheel has permanent magnets,

Optical vs. magnetic encoders

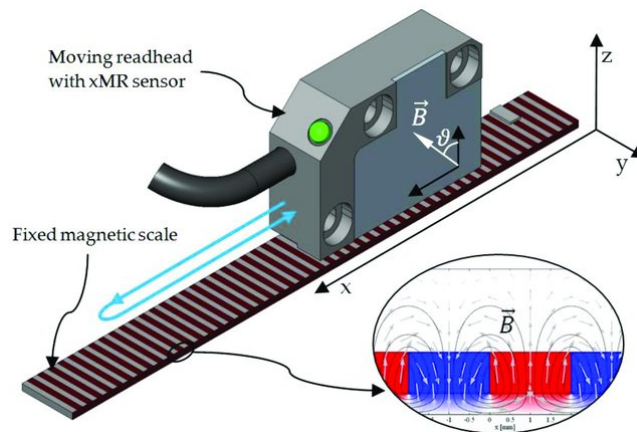


Figure 13.9: Magnetic linear incremental encoder (source: DOI 10.3390/s18072281).

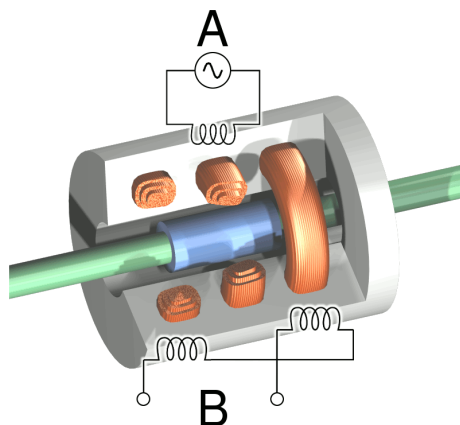


Figure 13.10: LVDT distance sensor (source: Wikimedia).

and their presence is detected by a circuit where a current is induced.

Measuring a linear position with an encoder is normally done using a rack and pinion or a roller wheel, but there are linear encoders (though not as common as rotary encoders) such as the one in Figure 13.9.

Linear encoders

LVDT and RVDT sensors

- **Linear variable differential transformers (LVDT sensors) and rotary variable differential transformers (RVDT sensors)** have three coils and a ferromagnetic core connected by a shaft to the position to be measured. The shaft moves back and forward, as seen in Figure 13.10, in a LVDT; in a RVDT, shown in Figure 13.11, it rotates. A fixed voltage is applied to coil *A*, the *primary coil*, and the core becomes an electromagnet. If it is equidistant from both the other coils, the *secondary coils*, the core induces no current in them, and there is no voltage *B*; this is called the *null point*. If the core moves aside, there will be a current in one of the secondary coils, and $B \neq 0$. LVDT and RVDT sensors have good resolution and bandwidth, little wear, and few temperature restrictions.

The following sensors measure linear position only.

Capacitive sensors

- **Capacitive sensors**, or capacitance sensors, have the position to be measured connected to one of the plates of a capacitor. If the moving target is conductive, it may be used as one of the plates. The capacitance C of a capacitor with plates of area A separated by distance x is given by

$$C = \frac{A\epsilon}{x} \quad (13.2)$$

where ϵ the dielectric constant of the medium between the plates. Capacitance is then measured applying a known current, so that the resulting

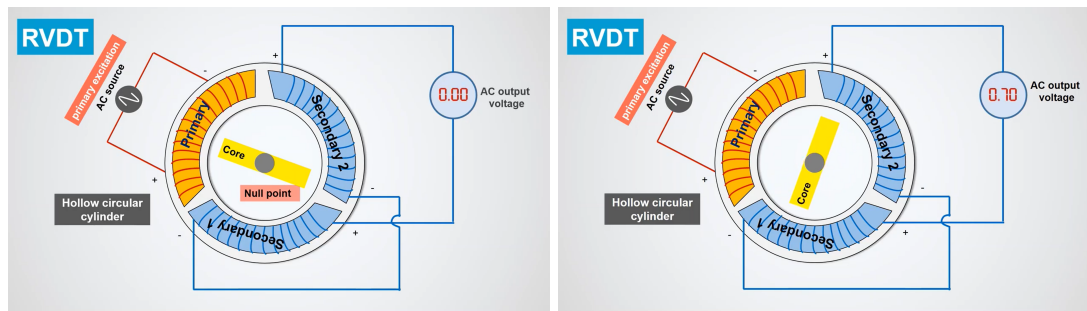


Figure 13.11: Left: RVDT with its core at the null point, and thus equal voltages in the secondary windings. Right: different voltages in the two secondary windings. (Source: https://www.youtube.com/watch?v=1NKxISM_W3M.)

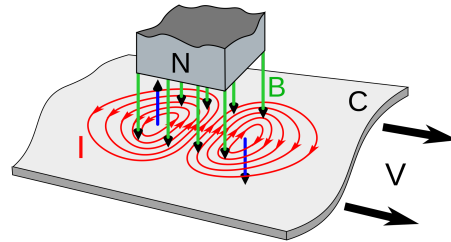


Figure 13.12: Eddy currents with intensity I induced by the magnetic field B of a magnet on a moving conducting material C with velocity V (source: Wikimedia).

voltage, after the capacitor is charged, is inversely proportional to the capacity C (remember (5.9)), and consequently linear with x .

- **Eddy current sensors** are based on eddy currents, or Foucault currents, and require that the target be conductive. As Figure 13.12 shows, a moving magnetic field induces an electric current in a conductive material. Since the target may not be moving, sensors use an electromagnet with a high frequency alternating current. In this way, the magnetic field is always changing, just as if a permanent magnet were moving. The closer the target, the more energy these currents have. They generate a magnetic field that counteracts the one that originates them. The result is that the sensor and the target can be modelled as a transformer with a variable coupling coefficient (remember Figure 5.22). The sensor itself will be similar to an LR system with both impedance L and resistance R depending on the distance x , as in Figure 13.13. These sensors operate in a wide amplitude of temperatures and have a high bandwidth. *Eddy current sensors*
- **Hall sensors**, or Hall effect sensors, are based on the Hall effect, explained in Figure 13.14, and again require a conductive target. The presence of the conductive target in the magnetic field changes the dielectric constant and consequently the Hall voltage. The variation of this voltage is used to find the distance of the target. *Hall sensors*
- **Inductive sensors** are similar, and are also based on an electromagnet. Unlike Hall sensors, only variations in the magnetic field caused by a metallic target are detected. Consequently, these sensors are proximity sensors. *Inductive sensors*
- **Strain gauges** are resistances glued to a surface, in such a way that will be stretched or compressed with it. They are used to measure small *Strain gauges*



Figure 13.13: Model of an eddy currents sensor.

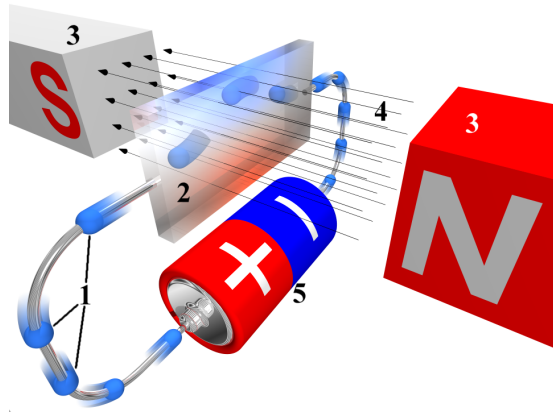


Figure 13.14: Hall effect: magnet (3) generates a magnetic field (4) that deviates the electrons (1) passing through plate (2) which is part of the circuit fed by battery (5). The force acting on the electrons is called Lorentz force. Because electrons are deflected, there is a difference of electric potential between the upper and the lower sides of plate (2), called Hall voltage. In practice, the plate is usually a semiconductor. (Source: adapted from Wikimedia.)

displacements (as well as several other quantities, as we will see in the following sections).

- When the surface is stretched, so is the resistance: its length increases and its cross-section decreases. (5.2) shows that the resistance will consequently increase.
- When the surface is compressed, so is the resistance: its length decreases and its cross-section increases. (5.2) shows that the resistance will consequently increase.

See Figure 13.15. The resistance of a strain gauge has several windings to maximise the changes in length and cross-section; see Figure 13.16. It should be glued so that these windings are parallel, or oblique, to the direction of the stretch or compression; it should never be perpendicular thereto.

Theorem 13.1. The variation of the resistance ΔR of a strain gauge is linear with the variation of its length ΔL , provided that it be small.

Proof. As the strain gauge is solid, its volume before and after the change in length is the same. Let the initial length be L and the initial cross-section be A . The final length will be $L + \Delta L$, and the final cross-section will be $A - \Delta A$; the variations ΔL and ΔA will be either both positive or both negative. Neglecting second order terms (that is why we need a small ΔL),

$$\begin{aligned}
 LA &= (L + \Delta L)(A - \Delta A) = LA - L\Delta A + A\Delta L - \overbrace{\Delta L\Delta A}^{\approx 0} \\
 \Rightarrow \Delta A &= \frac{A}{L}\Delta L
 \end{aligned}
 \tag{13.3}$$

From (5.2), we know that the initial resistance R is $R = \rho \frac{L}{A}$, and that the

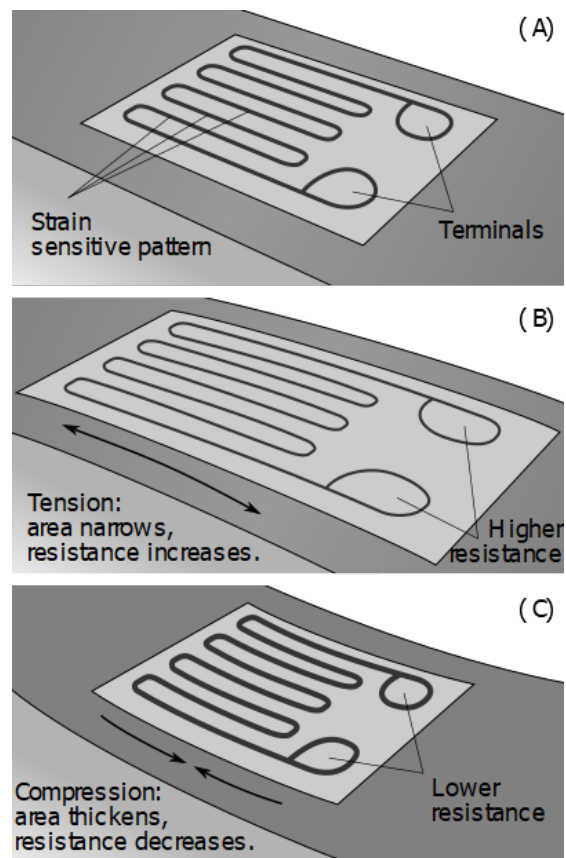


Figure 13.15: How a strain gauge's length and cross section change when it is stretched and compressed (source: Wikimedia).



Figure 13.16: A strain gauge (source: Wikimedia).

final resistance is

$$\begin{aligned}
 R + \Delta R &= \rho \frac{L + \Delta L}{A - \Delta A} \\
 \Leftrightarrow \rho \frac{L}{A} + \Delta R &= \rho \frac{L + \Delta L}{A - \Delta A} \\
 \Rightarrow \Delta R &= \rho \frac{L + \Delta L}{A - \frac{A}{L} \Delta L} - \rho \frac{L}{A} \\
 &= \rho \left(\underbrace{\frac{L^2 + L\Delta L}{LA - A\Delta L}}_{\approx LA} - \frac{L^2}{LA} \right) \\
 &= \rho \frac{L\Delta L}{LA} = \frac{\rho}{A} \Delta L \square
 \end{aligned} \tag{13.4}$$

Since a strain gauge transduces the variation in length into a resistance, another transduction is needed to obtain a voltage. The simplest way to do this would be a **tension divider**, as in Figure 13.17. A constant reference voltage V_i is provided, the output is V_o , and we let the additional resistor of the divider be equal to the nominal value of the resistance of the gauge:

$$\begin{aligned}
 \begin{cases} R + \Delta R = \frac{V_i - V_o}{i} \\ R = \frac{V_o}{i} \end{cases} &\Rightarrow i = \frac{V_i - V_o}{R + \Delta R} = \frac{V_o}{R} \\
 \Rightarrow V_i R - V_o R &= V_o R + V_o \Delta R \Leftrightarrow V_o (2R + \Delta R) = V_i R \\
 \Leftrightarrow \frac{V_o}{V_i} &= \frac{R}{2R + \Delta R}
 \end{aligned} \tag{13.5}$$

Because, as we saw, $\Delta R \ll R$, this configuration seldom, if ever, leads to acceptable results, as $\frac{V_o}{V_i}$ will be nearly constant. That is why an alternative configuration, the **Wheatstone bridge**, is used instead. There are in fact three different variations, shown in Figure 13.17, depending on the number of gauges used:

Wheatstone bridge: quarter bridge

- In a **quarter bridge**, there is one gauge, and

$$\begin{aligned}
 \begin{cases} V_A = V_i \frac{R}{R+R+\Delta R} \\ V_B = V_i \frac{R}{R+R} \end{cases} &\Rightarrow V_o = V_B - V_A = V_i \left(\frac{1}{2} - \frac{1}{2 + \frac{\Delta R}{R}} \right) \\
 \Rightarrow \frac{V_o}{V_i} &= \frac{2 + \frac{\Delta R}{R}}{2 \left(2 + \frac{\Delta R}{R} \right)} - \frac{2}{2 \left(2 + \frac{\Delta R}{R} \right)} \approx \frac{1}{4} \frac{\Delta R}{R}
 \end{aligned} \tag{13.6}$$

Wheatstone bridge: half bridge

- In a **half bridge**, there are two gauges, placed in such a way that one will be stretched and the other compressed; i.e. the variations of resistance will be *symmetrical*. If the distance measured is the result of bending a beam, that means that a gauge is placed on top and another at the bottom. Then

$$\begin{aligned}
 \begin{cases} V_A = V_i \frac{R - \Delta R}{R - \Delta R + R + \Delta R} \\ V_B = V_i \frac{R}{R + R} \end{cases} &\Rightarrow V_o = V_B - V_A = V_i \left(\frac{1}{2} - \frac{1 - \frac{\Delta R}{R}}{2} \right) \\
 \Rightarrow \frac{V_o}{V_i} &= \frac{1}{2} \frac{\Delta R}{R}
 \end{aligned} \tag{13.7}$$

Wheatstone bridge: full bridge

- In a **full bridge**, there are four gauges, placed in such a way that two will be stretched and two compressed:

$$\begin{aligned}
 \begin{cases} V_A = V_i \frac{R - \Delta R}{R - \Delta R + R + \Delta R} \\ V_B = V_i \frac{R + \Delta R}{R - \Delta R + R + \Delta R} \end{cases} &\Rightarrow V_o = V_B - V_A = V_i \left(\frac{1 + \frac{\Delta R}{R}}{2} - \frac{1 - \frac{\Delta R}{R}}{2} \right) \\
 \Rightarrow \frac{V_o}{V_i} &= \frac{\Delta R}{R}
 \end{aligned} \tag{13.8}$$

Even with a Wheatstone bridge, the output voltage usually needs to be significantly amplified to be read by an AD converter.

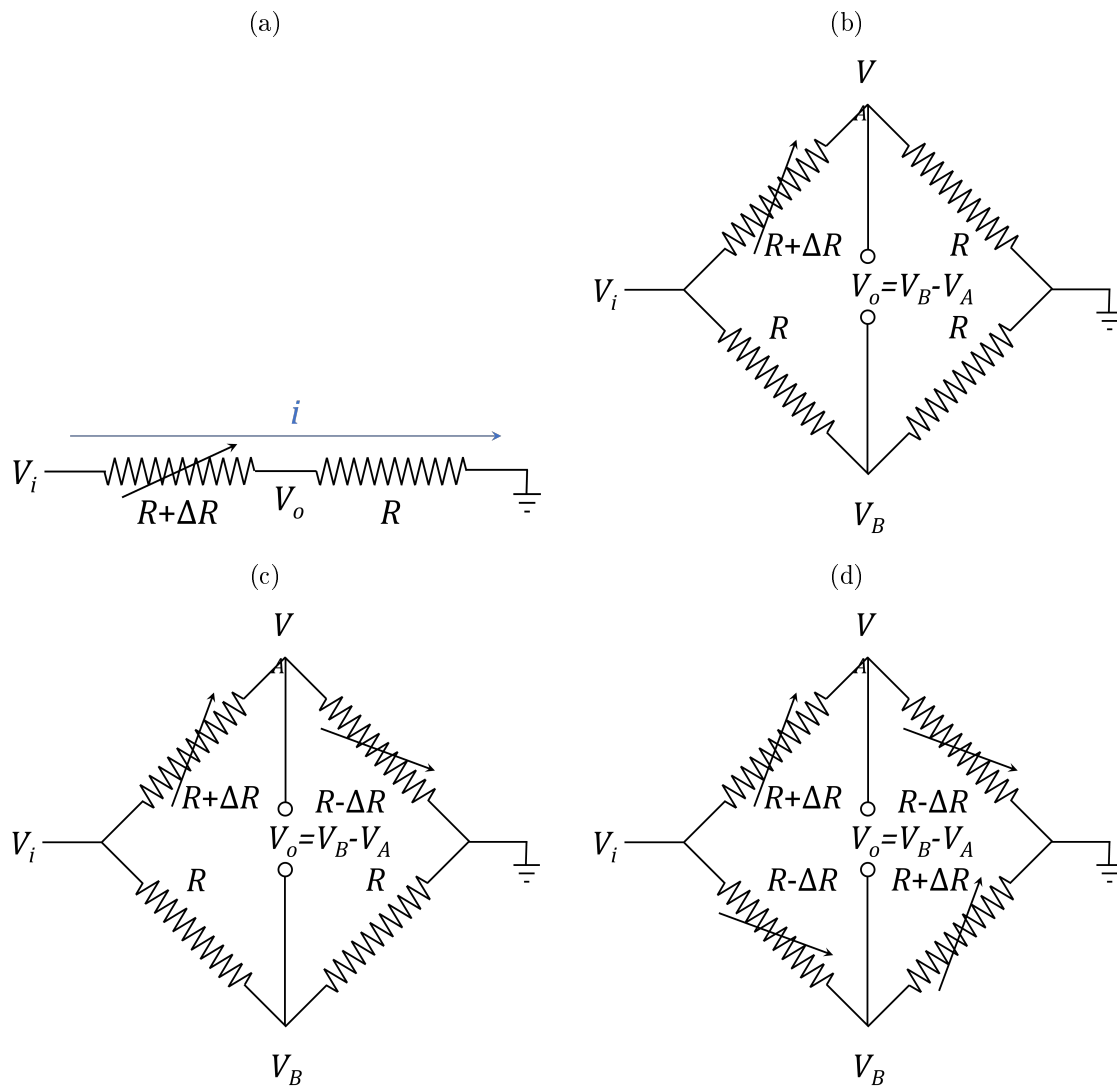


Figure 13.17: A strain gauge needs to be used with other resistors. (a) Voltage divider, a very poor configuration because the output hardly changes. (b) Wheatstone quarter bridge, with one gauge. (c) Wheatstone half bridge, with two symmetrical gauges. (d) Wheatstone full bridge, with two pairs of symmetrical gauges.

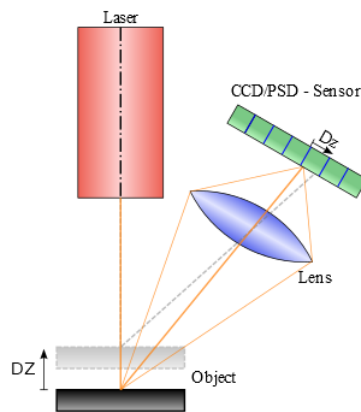


Figure 13.18: Measuring a distance using a laser triangulation sensor (source: Wikimedia). When the object comes closer by a distance DZ , its image in the light sensor moves by a distance Dz .

- **Laser sensors** use a laser (Light Amplification by Stimulated Emission of Radiation) beam to measure distance in one of two ways:

- Pulses of light are emitted and reflected in an object. The reflected light is called *return*, and detected; the time t between the emission of the pulse and the detection of the return is given by $c = \frac{d}{t}$, where c is the speed of light and d the desired distance. This is the way a lidar (LIght Detection And Ranging) works. Lidar sensors are often used in autonomous vehicles, because they are fast (they have a very large bandwidth) and can return distances for a wide range of angles, allowing a 3D image of the environment to be built. They are also used in cartography. In this case, the location and altitude of an airplane carrying a lidar are found using a GPS (which we will address below); the lidar gives the distance of the airplane to the ground, once more sweeping a wide range of angles. Since the GPS gives the altitude not in relation to the ground, but as a distance to the centre of the Earth, the lidar readings can be used to build a 3D image of what the airplane is flying over. The range of distances a lidar can measure varies significantly; it can be of tens, hundreds, or even thousands of meters. Larger ranges result in lower accuracies.

Lidar

- The laser is emitted continuously, is reflected, and detected. The position where it is detected is related with the distance to the surface as shown in Figure 13.18. This is called **triangulation**. The range of lengths that can be measured in this way is smaller than that of the lidar, but on the other hand an accuracy of almost $1 \mu\text{m}$ can be obtained.

Laser triangulation

- **Ultrasound sensors** emit ultrasounds and measure the time it takes for the echo to return. This may be done with a single transducer, that both emits the ultrasound and measures the echo. In a way, they are similar to a lidar, but using ultrasounds instead of laser pulses; this means that the sensor works even when there is smoke or fog (environments that pose problems to laser sensors), or when the material fails to reflect a laser beam because of being translucent. On the other hand, the range is usually of at most a few meters only, and care must be taken because the speed of sound is not constant (it depends on temperature and on other usually less relevant factors).

Ultrasound sensors

- A similar principle can be found in a **sonar** (SOund Navigation And Ranging), that uses audible sound to measure underwater distances.

Sonar

- **GNSS** (Global Navigation Satellite Systems) deserve to be mentioned here, though they are significantly different from the precedent technologies and are used for different purposes. The better known GNSS is the GPS (Global Positioning System) of the government of the United States

GPS

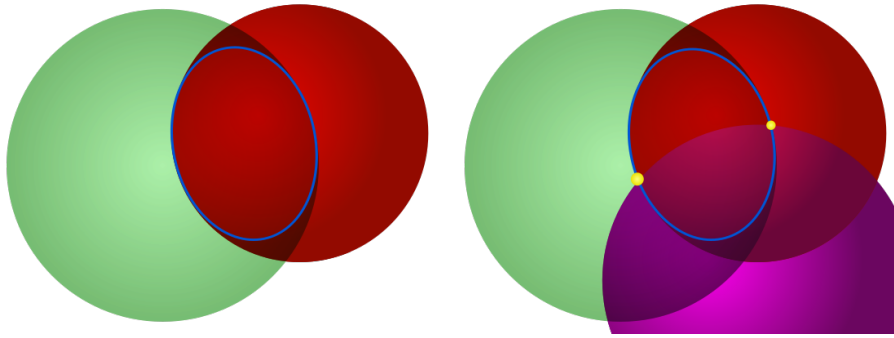


Figure 13.19: Knowing one's distance to a point of known location (such as a GPS satellite), we know we are on the surface of a sphere. Knowing one's distance to two points of known location, we know we are on the intersection of two spheres, i.e. on a circle (source: Wikimedia). Knowing one's distance to three points of known location, the current location can be known precisely if there are no errors in the distances or the positions of the spheres' centres (source: Wikimedia). (In fact the intersection of three spheres leads to two points, but the GPS receiver is presumed to be at the point closer to the Earth, and not in outer space.) Knowing one's distance to more than three points of known location allows a better reckoning of the current location using least squares.

of America, though there are others, such as the European Union's Galileo, and many are compatible with each other; in any case, GPS is often used by synecdoche to design all GNSSs. These systems are based on satellites orbiting the Earth, broadcasting radio signals that include the satellite's location and the time, given by an atomic clock (this is a very simplified description, but suffices for our purpose). GPS receivers read these signals and find how much time the signal took to arrive; from this time t , the distance d to the satellite is found as $c = \frac{d}{t}$, where c is the speed of light. Knowing the distance to several such satellites, and their position, the current location is found as the point which is at the correct distance from all the satellites (see Figure 13.19). Receiving four signals, it is possible to correct the GPS receiver's clock (which is not as precise as the satellites' clocks).

As to angular position sensors:

- There is an additional sensor technology that measures angular position only and deserves a mention. A **resolver** is similar to a transformer (which we studied in Section 5.4), or, rather, to two transformers, with the primary winding in a rotor, and two secondary windings in a stator within which the rotor rotates. (Notice the difference to a RVDT.) The number of turns in the two secondary windings must be the same; the number of turns in the primary winding can be different, but to simplify we will assume that all windings have the same number of turns, and that there are no losses.

Resolver

The primary voltage V_P is an AC voltage with a frequency ω_I which must be clearly above the frequency of rotation of the rotor:

$$V_P(t) = A \sin(\omega_I t) \quad (13.9)$$

The two secondary windings make an angle of 90° with each other; see Figure 13.20. When the primary is aligned with one of them, a voltage equal to $V_P(t)$ will be induced in that winding, while the other one, being perpendicular, will have no current whatsoever. When the rotor makes an angle θ with one of the windings, the voltages will be

$$V_{S1}(t) = A \cos \theta \sin(\omega_I t) \quad (13.10)$$

$$V_{S2}(t) = A \sin \theta \sin(\omega_I t) \quad (13.11)$$

For this reason, the two stator windings are known as \cos and \sin windings. If the rotor rotates with a constant angular velocity $\dot{\theta}$, their voltages

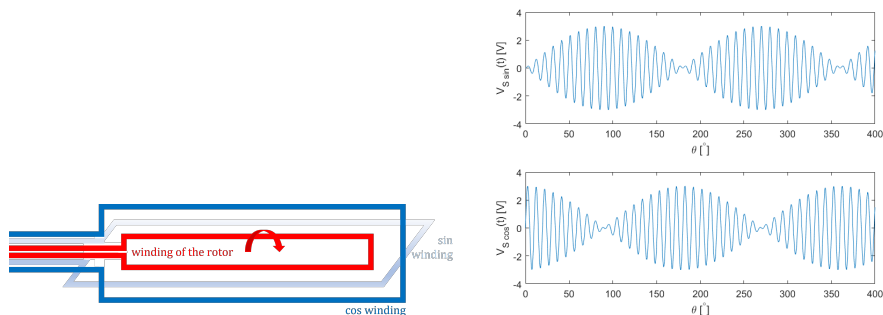


Figure 13.20: Left: the three windings of a resolver. Right: how the voltages of the sin and cos windings change with time when the rotor rotates with a constant angular velocity.

change as shown in Figure 13.20, and $\theta(t)$ is found as

$$\theta(t) = \arctan \frac{V_{S2}(t)}{V_{S1}(t)} \quad (13.12)$$

Because the signs of both the numerator and the denominator are known, $\theta(t)$ can be found in any quadrant.

Clinometer

- **Inclinometer** or **clinometer** is a generic name for a sensor that measures the inclination or tilt of a probe, either from the vertical or from the horizontal. Any working principle for an angular position sensor will do, as long as it can measure small angles with a reasonable precision; frequently, servo accelerometers, addressed below in Section 13.3, are used.

Compass

- A **compass** is often employed with a GPS to know not only one's position, but one's orientation as well. Most compasses make use of the Hall effect (remember Figure 13.14).

13.3 Working principles of velocity and acceleration sensors

Tachometers

Only angular velocity sensors, known as **tachometers**, use some working principles not yet addressed:

- Some are DC generators, i.e. they are DC motors that, instead of transforming electrical energy into movement, transform movement into a DC current. We will address DC motors in Chapter 14.

Stroboscope

- A **stroboscope** is a flashing light. It is used to find an angular velocity in the following way:
 - The rotating body has one distinctive mark.
 - The frequency of the flashes ω is varied until the distinctive mark seems to be stopped.
 - This means that, in period $\frac{2\pi}{\omega}$, the disc completed an integer number of rotations.
 - The frequency of the flashes is increased until the highest value for which the mark seems to be stopped. In this case, the disc will complete only one rotation in period $\frac{2\pi}{\omega}$, and its angular velocity is ω .
 - It is usual to check that this is the real period increasing the frequency of the flashes to 2ω . The mark should then be seen twice, on opposite sides of the rotating body. Notice that this test is not completely conclusive. If during the suspected period $\frac{2\pi}{\omega}$ the body completes in fact an even number of rotations n , doubling the frequency we will see in the new period $\frac{\pi}{\omega}$ a number of rotations $\frac{n}{2}$ which is still integer, and we will know that ω was not yet the angular velocity of

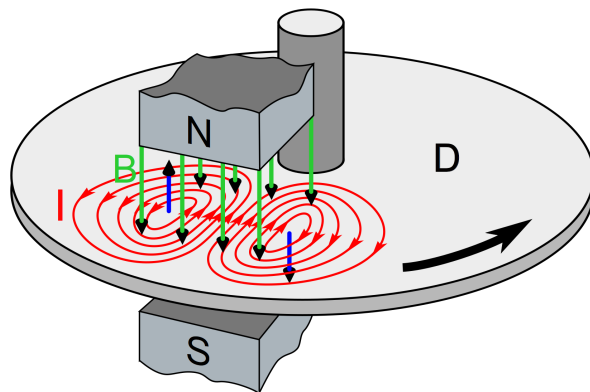


Figure 13.21: Eddy currents in a disk induced by a magnet (source: Wikimedia). If the disk rotates, and the magnet does not, it functions as a brake. If the magnet rotates, it drags the disk along.

the rotating body. But if n is odd, say, equal to 3, in the new period $\frac{\pi}{\omega}$ there will be 1.5 rotations, and the mark will be seen on opposite sides of the body just as if ω were the actual angular velocity of the rotating body. Thus, doubling ω should only be done when it is quite certain that the body is rotating at most twice during period $\frac{2\pi}{\omega}$.

- **Drag cup tachometers** are based on eddy currents, already mentioned *Drag cup* about eddy current sensors. While eddy current sensors use an AC electromagnet, so as to be able to measure the position of a target which did not move, these tachometers use a permanent magnet that rotates with the body whose angular velocity we want to measure. It is this rotation that induces eddy currents in a metallic (usually aluminium) cup, within which the magnet rotates. As mentioned, these currents themselves generate a magnetic field; thus, there is a torque between the magnet and the disk, proportional to the velocity of the magnet. In other words, the rotation of the magnet drags the cup; it would cause it to rotate as well, but the cup is connected to a rotational spring, that exerts a force proportional to the angular position, as in (4.24). Thus, the cup will stop at an angular position proportional to the angular velocity of the magnet. See Figure 13.21.

The cup may be connected to a pointer (to show the velocity on a graduated dial) or connected to a sensor of angular position. These sensors were often used in cars, though nowadays less often. In a car, they can be used to show the velocity of the vehicle (with the magnet connected to the output of the gearing box, and thus rotating with the wheels) or the velocity of the engine (with the magnet connected to the input of the gearing box, and thus rotating with the engine).

Accelerometers, or acceleration sensors, use the following working principles:

- This chapter is still being written.

13.4 Sensors for force, binary, pressure and level



13.5 Sensors for flow and pressure in flows

In the picture: National Pantheon, or Church of Saint Engratia, Lisbon (source: <http://www.panteaonacional.gov.pt/171-2/historia-2/>), somewhen during its construction (1682–1966).

13.6 Sensors for temperature and luminosity

This chapter is still being written.

13.7 Sensors for pH and concentration

In the picture: National Pantheon, or Church of Saint Engratia, Lisbon (source: <http://www.panteaonacional.gov.pt/171-2/historia-2/>), somewhen during its construction (1682–1966).

Glossary

Ἐὰν οὖν μὴ εἰδῶ τὴν δύναμιν τῆς φωνῆς, ἔσομαι τῷ λαλοῦντι βάρβαρος
καὶ ὁ λαλῶν ἐν ἐμοὶ βάρβαρος.

Saint PAUL of Tarsus (5 BC? — †67?), *First Epistle to the Corinthians* (c. 53–55?), xiv 11

anemometer anemómetro
encoder contador incremental
extensometer extensómetro
speedometer velocímetro

Exercises

1. According to European standards, car speedometers must never indicate a speed below the actual speed, and the indicated speed must not exceed the actual speed by more than 10% of the actual value plus 4 km/h.

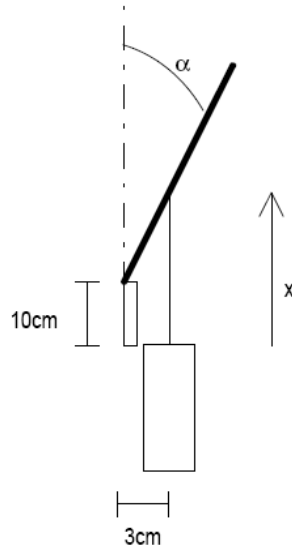


Figure 13.22: Angular position measurement system from Exercise 2.

- (a) Let v be the actual speed and v_m the indicated speed. Plot the maximum and minimum admissible values of v_m as functions of v , together in the same plot.
 - (b) Plot the maximum and minimum admissible values of v as functions of v_m , together in another different plot.
2. The angular position α in Figure 13.22 is measured using a laser sensor, able to provide readings in the [100 mm, 200 mm] range, with a resolution of 1 mm and a precision of 5 mm.
- (a) What are the resolution and the precision in relation to the 100 mm width of the measuring range?
 - (b) Show that the angular position α and the linear position x are related by

$$\tan \alpha = \frac{d}{x - x_0} \quad (13.13)$$

What are the values of d and x_0 ?

- (c) What are the maximum and minimum values of α that can be measured?
 - (d) Plot $\alpha(x)$ for the entire possible ranges of both variables.
 - (e) What is the precision in the measurement of α ?
 - (f) What values can the resolution take?
3. The angular velocity ω of a rotating shaft is measured with an encoder that provides 1024 pulses per rotation, connected to a counter that uses a 5 Hz sampling frequency. The shaft can rotate up to 7500 rpm.
- (a) Show that the change Δn of the counter reading between two successive sampling instants is given by

$$\Delta n = \lfloor 32.6\omega \rfloor \quad (13.14)$$

when ω is given in rad/s.

- (b) Find the absolute value of the resolution of the angular velocity measurement.
- (c) Find the resolution in relation to the largest possible value of the angular velocity.
- (d) How many 4-bit counters are needed to make up a counter that can read all possible values?

Table 13.1: Three accelerometers for Exercise 4.

	Servo	Piezoelectric	Piezoresistive
Range	10 g	10 g	10 g
Pass band	300 Hz	[1 Hz, 10000 Hz]	1000 Hz
Sensibility	1 mA/g	0.1 V s ² /m	5 μV/V/g
Precision	10 ⁻⁴ g	0.5%	1%
Price	1800 €	180 €	500 €

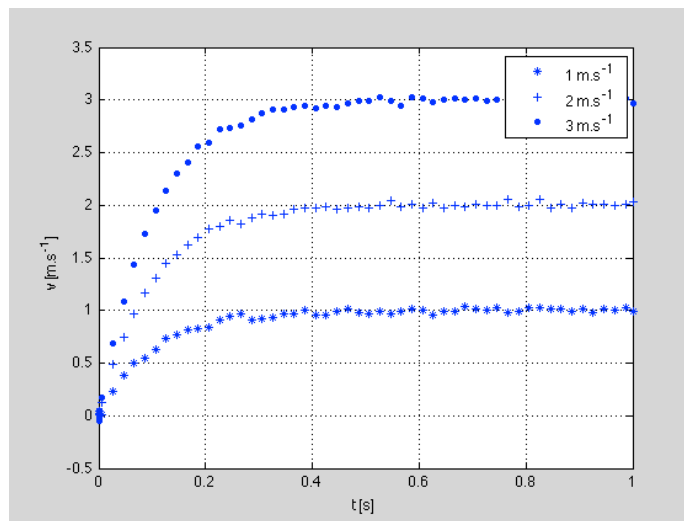


Figure 13.23: Time responses from Exercise 5.

- (e) If only the first 8 bits are considered, what will be the resolution of the angular velocity measurement?
4. An accelerometer is needed to measure accelerations in an automobile, in the [0 g, 2 g] range, with frequencies in the [0 Hz, 50 Hz] range, during 2 hours.
- Which of the three sensors in Table 13.1 would you choose, and why?
 - The AD converter has 8 bits and an input in the [0 V, 5 V] range. Assuming that the power supply for the sensor will be the 12 V DC car battery, design the necessary signal conditioning.
 - What will the resolution be?
 - If data is recorded with a 250 Hz sampling frequency, how large will be the file where measurements are recorded?
5. A velocity sensor was tested for three different constant speeds. The time responses obtained are shown in Figure 13.23.
- Do these time responses support that the sensor measurement is linear?
 - Find a suitable transfer function to model the sensor response.
6. An elevator comprises a 500 kg cabin, a 600 kg counterweight, and an electrical motor to move the steel cable that connects them. A 280 Ω extensometer, with sensibility $\frac{\delta R}{R} = 2$, mounted in a simple bridge powered at 24 V, measures the elastic deformation of the cable, given by (see Figure 13.24)

$$\varepsilon = \frac{F}{SE} \quad (13.15)$$

where the cable's cross-section is $S = 4 \text{ cm}^2$, and the Young modulus is $E = 10^{11} \text{ Pa}$. The objective is to detect a cargo above the maximum admissible value of 150 kg.

- Draw a scheme of the signal conditioning described.

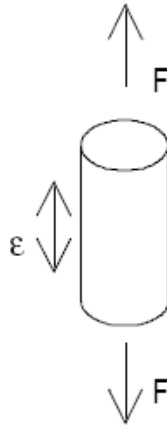


Figure 13.24: The elevator cable from Exercise 6.

- (b) With this signal conditioning, what will the sensibility be, in V/N?
 (c) Design an additional signal conditioning element to sound a buzzer when the cargo is too heavy.
7. The temperature of a motor can assume values in the $[10^{\circ}\text{C}, 180^{\circ}\text{C}]$ range, and is measured with an infrared sensor that works in the $[-18^{\circ}\text{C}, 538^{\circ}\text{C}]$ range. Its output is in the $[0\text{ V}, 5\text{ V}]$ range, and its precision is given by

$$\max\{4^{\circ}\text{C}, 2\%T_F\} \quad (13.16)$$

where T_F is the temperature measured, in $^{\circ}\text{F}$.

- (a) Find the relation between the sensor output e and the temperature T .
 (b) What is the sensor's sensibility?
 (c) Given the range of temperatures being measured, what will be the actual range of e ?
 (d) Plot the precision as function of T_F , in $^{\circ}\text{F}$.
 (e) Given the range of temperatures being measured, what will be the maximum value of the error?
 (f) The sensor is directly connected to an 8-bit AD converter, that receives inputs in the $[0\text{ V}, 5\text{ V}]$ range. Find the AD output as a function of temperature.
 (g) What will be the resolution of the measurement, in $^{\circ}\text{C}$?
 (h) AD converter noise affects 3 LSB. What will be the precision of the measurement, considering both conversion noise and sensor precision?
 (i) The emissivity is 0.6, but estimated as 0.5. How will this affect precision?
8. A sensor outputs a tension in the $[0.2\text{ V}, 3.3\text{ V}]$ range, varying linearly with the relative humidity in the $[0\%, 100\%]$ range.
- (a) Design the signal conditioning that will convert this output into the $[0\text{ V}, 1\text{ V}]$ range. Available tensions are 12 V, -12 V , and 5 V.
 (b) This will be connected to a 10-bit AD converter that receives tensions in the $[0\text{ V}, 1\text{ V}]$ range. What is the resolution of the measurement?
 (c) The precision of the sensor is 1% or less. What is the precision of the measurement, considering both the precision of the sensor and the resolution of the AD converter?
 (d) Figure 13.25 shows a control system of relative humidity $H(s)$, where $H_{ref}(s)$ is the reference for humidity $H(s)$, $P(s)$ is a disturbance, $G_p(s) = \frac{100}{s+100}$ is the process we want to control, $G_s(s)$ is the sensor, and $G_c(s) = \frac{10}{s+10}$ is a controller. Find transfer function $\frac{H(s)}{H_{ref}(s)}$, and plot its Bode diagram.

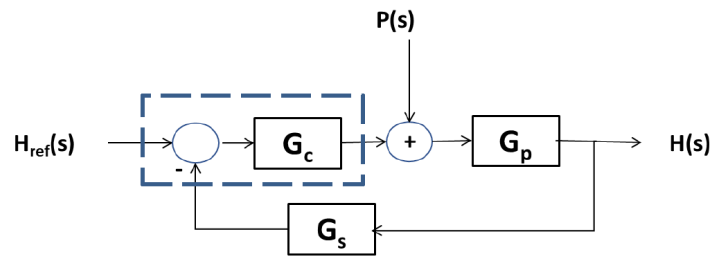


Figure 13.25: Relative humidity control system from Exercise 8.

9. What is the resolution of the position measurement depicted in Figure 13.4?

Chapter 14

Actuators

An engineer is a person who enjoys the feel of tools, the noise of dynamos and smell of oil.

George ORWELL (1903 — †1950), *Benefit of Clergy: Some Notes on Salvador Dali* (1944)

This chapter is still being written.

14.1 Generalities about electric motors

AC, DC, power amplification.

14.2 DC motors

In the picture: National Pantheon, or Church of Saint Engratia, Lisbon (source: <http://www.panteaonacional.gov.pt/171-2/historia-2/>), somewhen during its construction (1682–1966).

14.3 AC motors

This chapter is still being written.

14.4 Generalities about pneumatic and hydraulic actuators

In the picture: National Pantheon, or Church of Saint Engratia, Lisbon (source: <http://www.panteaonacional.gov.pt/171-2/historia-2/>), somewhen during its construction (1682–1966).

14.5 Pneumatic and hydraulic compressors and motors

This chapter is still being written.

14.6 Cylinders and valves

In the picture: National Pantheon, or Church of Saint Engratia, Lisbon (source: <http://www.panteaonacional.gov.pt/171-2/historia-2/>), somewhen during its construction (1682–1966).

14.7 Dimensioning of pneumatic and hydraulic circuits



Glossary

Os bos e xenerosos
 a nosa voz entenden
 e con arroubo atenden
 o noso rouco son,
 máis sóo os ñorantes,
 e féridos e duros,
 imbéciles e escuros
 non os entenden, non.

Eduardo María González-PONDAL Abente (1835 — †1917), *Queixumes dos pinos* (1886), *Os pinos*

alternating current corrente alternada
brushless motor motor sem escovas
compressed air ar comprimido
compressor compressor
coreless motor motor sem núcleo
direct current corrente contínua
double acting valve, double action valve válvula de duplo efecto
power amplifier amplificador de potencia
stepper motor motor passo-a-passo

Exercises

- We want to control the angular position of a brushless DC motor, with the characteristics given in Figure 14.1. An 8-bit DA converter will be used, with the characteristics given in Figures 14.2 and 14.3. The supply voltage VDD of the DA converter will be the highest possible.
 - The motor should turn in both senses, and consequently must receive a voltage in the range from -12 V to $+12\text{ V}$. Design a signal conditioning to connect the output of the DA converter to the motor input.
 - Show how you could implement this signal conditioning using OpAmps. Find reasonable values for resistors and other components you may need. Assume that you have -12 V and $+12\text{ V}$ available.

- (c) Find the resolution of the control action in Volt, after the signal conditioning.
- (d) Tensions -12 V and $+12\text{ V}$ correspond to the motor's nominal rotation velocity of 6070 rpm shown in Figure 14.1 (in different senses of rotation, of course). What is the resolution of the rotation velocity?
- (e) When the DA converter's output is 01101001, what tension will be supplied to the motor, and what will be its rotation speed?
- (f) The angular position of the motor's shaft is being measured with 160 Hz noise. Design a first order filter that brings the amplitude of the noise down to 10% of its original value, with the smallest possible attenuation to low frequencies two decades or more below the frequency of the noise. (There is no need to show how the filter could be implemented with OpAmps.)
- (g) The DC motor takes a time interval τ to reach 95% of the expected rotation velocity when there is a step in the tension supplied. Model this dynamic behaviour with a first order transfer function (with unit gain at low frequencies).

Specifications

- Nominal Voltage: 12V
- No Load RPM: 7000
- No Load Current: 0.65A
- Rated RPM: 6070
- Rated Torque: 10 oz-in
- Stall Current: 43.9A
- Stall Torque: 125.69 oz-in
- Shaft Type: Round

Figure 14.1: Part of a DC motor's datasheet.

7.3 Recommended Operating Conditions

over operating free-air temperature range (unless otherwise noted)

		MIN	NOM	MAX	UNIT
V_{DD}	Positive supply voltage to ground (A_{GND})	1.71		5.5	V
V_{IH}	Digital input high voltage, $1.7\text{ V} < V_{DD} \leq 5.5\text{ V}$	1.62			V
V_{IL}	Digital input low voltage			0.4	V
T_A	Ambient temperature	-40		125	°C

Figure 14.2: Part of a DA converter's datasheet.

By default, the DACx3401 operate with the power-supply pin (V_{DD}) as a reference. Equation 1 shows DAC transfer function when the power-supply pin is used as reference.

$$V_{OUT} = \frac{DAC_DATA}{2^N} \times V_{DD}$$

where:

- N is the resolution in bits, either 8 (DAC43401) or 10 (DAC53401).
- DAC_DATA is the decimal equivalent of the binary code that is loaded to the DAC register.
- DAC_DATA ranges from 0 to $2^N - 1$.
- V_{DD} is used as the DAC reference voltage.

(1)

Figure 14.3: Part of a DA converter's datasheet (continued).

2. To control the angular velocity of a shaft, its angular position is measured with an encoder. Figure 14.4 shows part of the encoder's manual. The sampling frequency will be the highest supported by the encoder, which is 1 kHz. The shaft is actuated by a brushed DC motor with the characteristics shown in Figure 14.5.
 - (a) What values, in both degrees and radians, do the two resolutions mentioned in Figure 14.4 correspond to?
 - (b) Which of the two resolutions allows reading a lower angular velocity of the motor? Find the value of that velocity in radians per second.

The absolute angle position value from the interpolator is output through a parallel binary interface or a serial SSI interface. The relative changes of the angle position are output through incremental A QUAD B encoder signals. The resolution of incremental output is selectable between 128 and 256 counts per turn with an external pin.

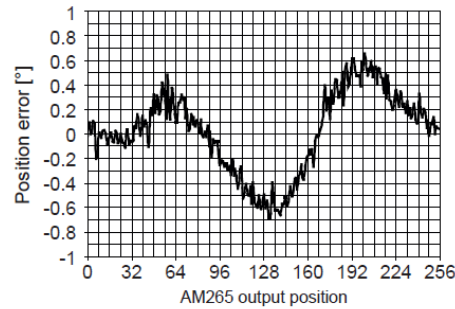


Fig. 9: Typical integral nonlinearity plot if the magnet is on the limit of alignment tolerances

Figure 14.4: Part of an encoder's manual for Exercise 2.

Specifications

- Nominal Voltage: 12V
- No Load RPM: 7000
- No Load Current: 0.65A
- Rated RPM: 6070
- Rated Torque: 10 oz-in
- Stall Current: 43.9A
- Stall Torque: 125.69 oz-in
- Shaft Type: Round

Figure 14.5: Part of a brushed DC motor's manual for Exercise 2.

- (c) The plot in Figure 14.4 shows the non-linearity of the sensor for the worst situation that still respects alignment tolerance, when 256 counts per turn are selected. That largest absolute value of the angle given in the plot correspond to how many LSB?
 - (d) For the rated angular velocity of the motor (do not confound this with the angular velocity when there is no load), and for both resolutions mentioned in Figure 14.4, find the change in the reading of the encoder during a sample time.
 - (e) Between the DA converter providing the control action and the DC motor, there is one of the two signal conditioning circuits in Figure 14.6. For each of the circuits, find what is the largest voltage that the DA can supply, without exceeding the nominal voltage at the input of the motor. (Consider that the power amplifier does not change the voltage.)
3. A pneumatic double acting cylinder is used in a circuit with compressed air at 10 bar. The inside diameter of the cylinder (i.e. the bore) is 5 cm. The diameter of the rod is 1 cm.
 - (a) If this is a cylinder with a through rod, what force can it exert?
 - (b) If this cylinder does not have a through rod, what force can it exert?

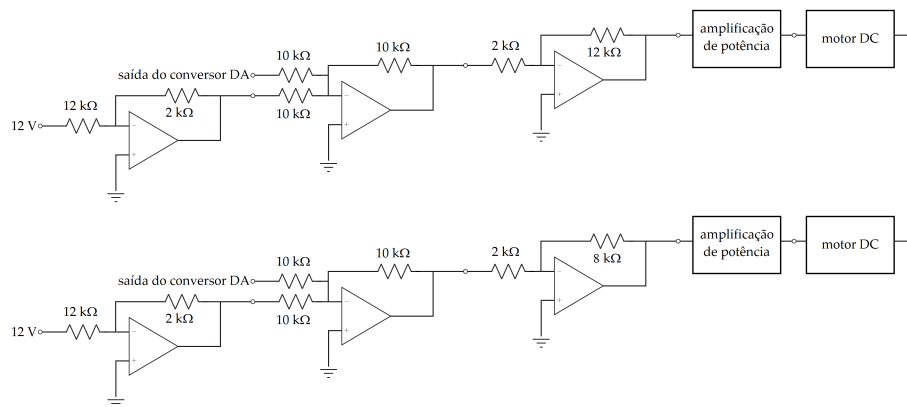


Figure 14.6: Signal conditioning circuits for Exercise 2.

Part IV

Control systems

The thought makes me feel old. I can remember when there wasn't an automobile in the world with brains enough to find its own way home. I chauffeured dead lumps of machines that needed a man's hand at their controls every minute. Every year machines like that used to kill tens of thousands of people.

The automatics fixed that. A positronic brain can react much faster than a human one, of course, and it paid people to keep hands off the controls. You got in, punched your destination and let it go its own way.

We take it for granted now, but I remember when the first laws came out forcing the old machines off the highways and limiting travel to automatics. Lord, what a fuss. They called it everything from communism to fascism, but it emptied the highways and stopped the killing, and still more people get around more easily the new way.

Of course, the automatics were ten to a hundred times as expensive as the hand-driven ones, and there weren't many that could afford a private vehicle. The industry specialized in turning out omnibus-automatics. You could always call a company and have one stop at your door in a matter of minutes and take you where you wanted to go. Usually, you had to drive with others who were going your way, but what's wrong with that?

Isaac ASIMOV (1920 — †1992), *Sally*, *Fantastic*, May-June 1953

In this part of the lecture notes:

- Chapter 15 is about control strategies and controller structures, among which are PID controllers and lead-lag controllers.
- Chapter 16 presents the root locus plot, which is a tool to study closed loop control systems, and how it can be used to design controllers.
- Chapter 18 concerns the Nyquist stability criterion, which are another tool to study closed loop control systems.
- Chapter 17 addresses stability margins, yet another tool to study closed loop control systems.
- Chapter 19 shows the Nichols plot, a further tool for the study of closed loop control systems.
- Chapter 20 studies steady-state errors in closed loop control systems.
- Chapter 21 systematically exposes methods to design controllers of the PID type.
- Chapter 22 systematically exposes methods to design controllers of the lead-lag type.
- Chapter 23 shows how to design controllers using the Internal Model Control methodology.
- Chapter 24 introduces pure delay systems and the problem they cause to control.

Here is what you need to know beforehand to follow these chapters:

- The Laplace and Fourier transforms, from Chapter 2;
- Transfer functions, from Sections 4.1 and 4.2 of Chapter 4;
- System theory, from Part II;
- Filters, from Sections 12.2 and 12.3 of Chapter 12.

Chapter 15

Control strategies and controller structures

—Bon! Peut-être les Sélénites ont-ils poussé plus loin que vous le calcul intégral! Et à propos, qu'est-ce que ce calcul intégral ?

—C'est un calcul qui est l'inverse du calcul différentiel, répondit sérieusement Barbicane.

—Bien obligé.

—Autrement dit, c'est un calcul par lequel on cherche les quantités finies dont on connaît la différentielle.

—Au moins, voilà qui est clair, répondit Michel d'un air on ne peut plus satisfait.

Jules VERNE (1828 — †1905), *Autour de la Lune* (1869), IV

In Chapter 9 we already studied the two basic configurations for control systems:

- open-loop control, in which there is no feedback of the output, and the controller's input is the reference that the output must follow;
- closed-loop control, in which there is output feedback, and the controller's input is the closed-loop error between the reference and the output.

See Section 9.3 again, and in particular Figure 9.13. In this chapter we will study what these configurations can do, and what controllers are used with each.

15.1 Open loop control

Open loop control:

- is conceptually simpler than closed loop control;
- does not require a measurement of the output;
- consequently, it can be implemented even when the output cannot be measured, or is difficult or expensive to measure: no sensor for the output variable is required to implement it;
- can anticipate control actions if the reference is known in advance;
- works well if two conditions are simultaneously met:
 - we know a perfect model of the plant;
 - there are no unknown disturbances (noise) of either the control action or the output.

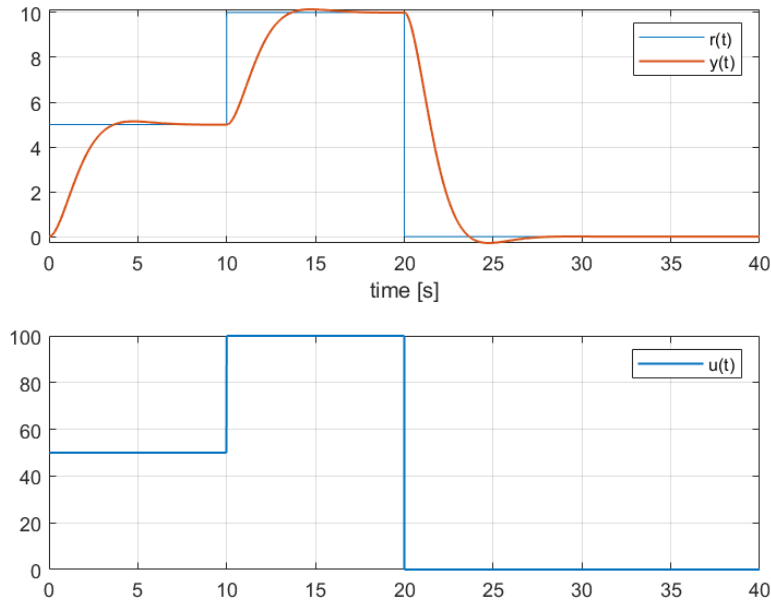


Figure 15.1: Open-loop control of (15.1) from Example 15.1.

Example 15.1. In the open-loop of Figure 9.13, let

$$G(s) = \frac{0.1}{s^2 + 1.5s + 1} \quad (15.1)$$

$$C(s) = 10 \quad (15.2)$$

Plant $G(s)$ is an underdamped second order plant, with damping coefficient $\xi = 0.75$ and steady-state gain 0.1. Controller $C(s)$ ensures that the steady-state of the open-loop $\frac{y(s)}{r(s)}$ is 1. Figure 15.1 shows how a reference consisting of three steps is followed, together with the corresponding control action, which should be always checked to see if control actions are not too large (remember that actuators saturate).

Open-loop control when the reference is known in advance

This is what happens when the reference is known for the time instant when it must be applied. But suppose that the reference is known in advance. In this case, the corresponding control action can begin *before* each step takes place. The best way to do so is to schedule the control action beforehand; in this case, $C(s)$ does not receive the reference as an input, as it must be followed; it just provides the control action chosen beforehand to follow the reference. Of course, in closed-loop control this would make no sense, since the controller needs the error, found from the current value of the output: and if the reference can be known in advance, the output cannot. Figure 15.2 shows this new situation. \square

Open-loop control does not work that well in practice

But,

- if the model of the plant is not perfect, or
- if there is noise,

things do not work that well. And, most importantly, it is nigh to impossible to control unstable plants in open loop.

Example 15.2. Consider the case of Example 15.1 again, with two changes:

- Figure 15.3 shows what happens if there is an error identifying the plant: in particular, a 10% error in the gain. Consequently,

$$G(s) = \frac{0.11}{s^2 + 1.5s + 1} \quad (15.3)$$

but the controller remains the same. Notice that the output has steady-state errors, but, since the controller does not receive any measurement of the output (which may even not be measured at all), nothing is done about it.

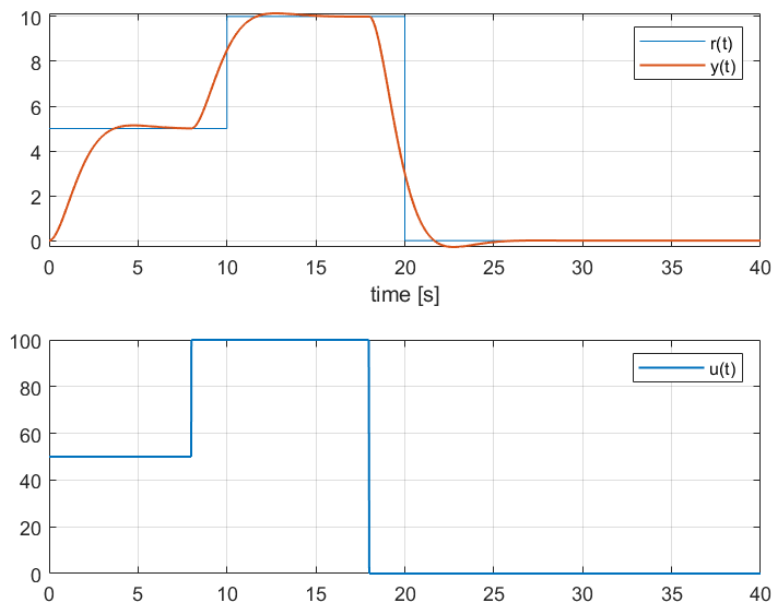


Figure 15.2: Open-loop control of (15.1) from Example 15.1, with an anticipated control action.

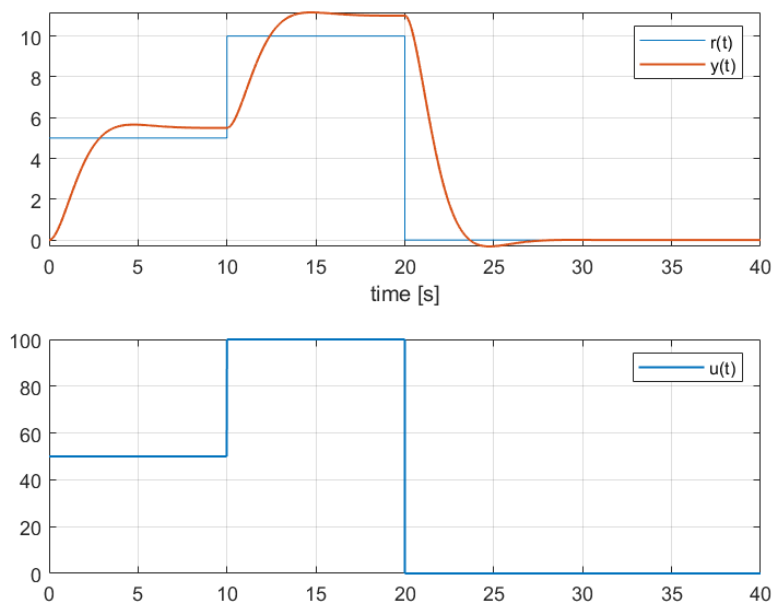


Figure 15.3: Open-loop control of (15.1) from Example 15.1, when the plant is known with a modelling error.

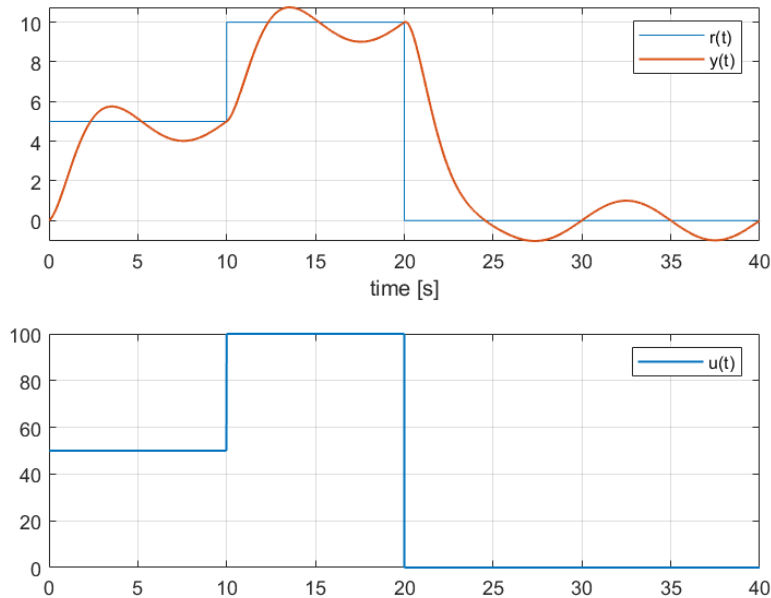


Figure 15.4: Open-loop control of (15.1) from Example 15.1, when the output is affected by sinusoidal disturbance (15.4).

- Figure 15.4 shows what happens if there is a disturbance: in particular, an additive disturbance in the output d_y ; see Figure 9.14. The actual, disturbed output $\tilde{y}(t)$ is given by

$$\tilde{y}(t) = y(t) + \underbrace{\sin \frac{\pi}{5} t}_{y_a(t)} \quad (15.4)$$

where $y(t)$ is the output we expected. Again, because the controller does not receive any measurement of the output, nothing is done about this oscillation which appears in the output. \square

15.2 Closed loop control

Thanks to output feedback, closed loop control can mitigate the effects of noise and wrong models.

Example 15.3. The underdamped second-order plant of Examples 15.1 and 15.2 can be controlled in closed-loop, as seen in Figure 9.13, with

$$C(s) = 55 + \frac{30}{s} + 25s \quad (15.5)$$

This is a type of controller that we will study later on in this chapter. For more realistic simulations of performance, and for fairer comparison with open-loop control, control actions were limited to the $[-500, 500]$ interval. (The effects of this saturation of the control action will be studied in depth in Chapter 28.) Figures 15.5 and 15.6 show that:

- in the absence of modelling errors or disturbances, faster responses are now provided, albeit at the cost of a higher overshoot (overshoots vs. fast responses are a frequent dilemma in closed-loop control, as we shall see in the next chapters);
- the effect of a 10% modelling error, that of (15.3), is nearly eliminated;
- noise (15.4) is significantly attenuated, even if not completely eliminated. \square

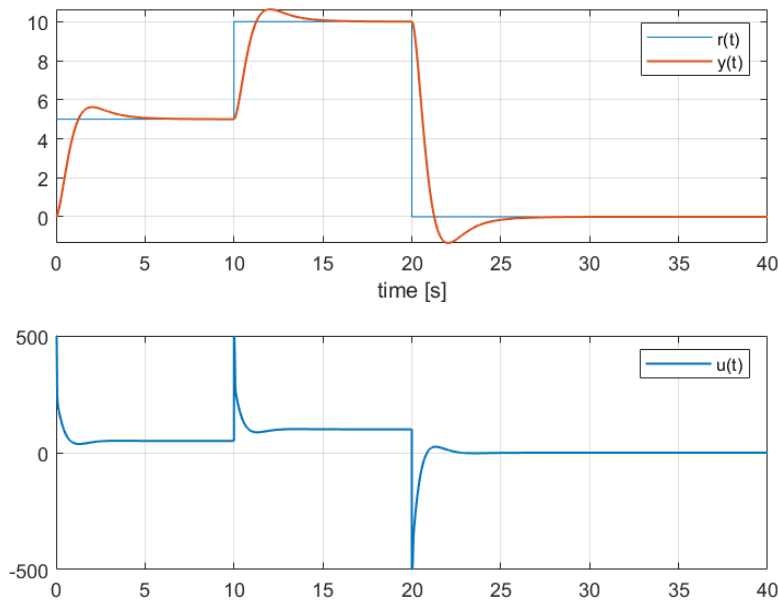


Figure 15.5: Closed-loop control of (15.1) from Example 15.1

A closed loop can also stabilise unstable plants.

Example 15.4. In the closed-loop of Figure 9.13, let

$$G(s) = \frac{1}{s-1} \quad (15.6)$$

$$C(s) = 10 + \frac{5}{s} \quad (15.7)$$

Figure 15.7 shows the response of the closed-loop to the same step-wise reference of the previous examples. We could have known that the closed-loop is stable, even though the plant controlled is not, by finding its poles:

$$\begin{aligned} \frac{y(s)}{r(s)} &= \frac{\left(10 + \frac{5}{s}\right) \frac{1}{s-1}}{1 + \left(10 + \frac{5}{s}\right) \frac{1}{s-1}} \\ &= \frac{\frac{10s+5}{s^2-s}}{1 + \frac{10s+5}{s^2-s}} = \frac{10s+5}{s^2+9s+5} \end{aligned} \quad (15.8)$$

The roots of the denominator are -8.4 and -0.6 : the closed-loop is thereby stable. \square

However, badly designed closed loop control can also make stable plants unstable.

Example 15.5. In the closed-loop of Figure 9.13, let

$$G(s) = \frac{1}{s^3 + 4s^2 + 6s + 4} \quad (15.9)$$

$$C(s) = 25 \quad (15.10)$$

The poles of $G(s)$ are -2 and $-1 \pm j$, and so the plant is stable; but the closed-loop and its poles are

```
>> G = 1/((s+2)*(s+1+1i)*(s+1-1i));
>> closed_loop = feedback(25*G, 1)
```

```
closed_loop =
```

```
25
```

```
-----
```

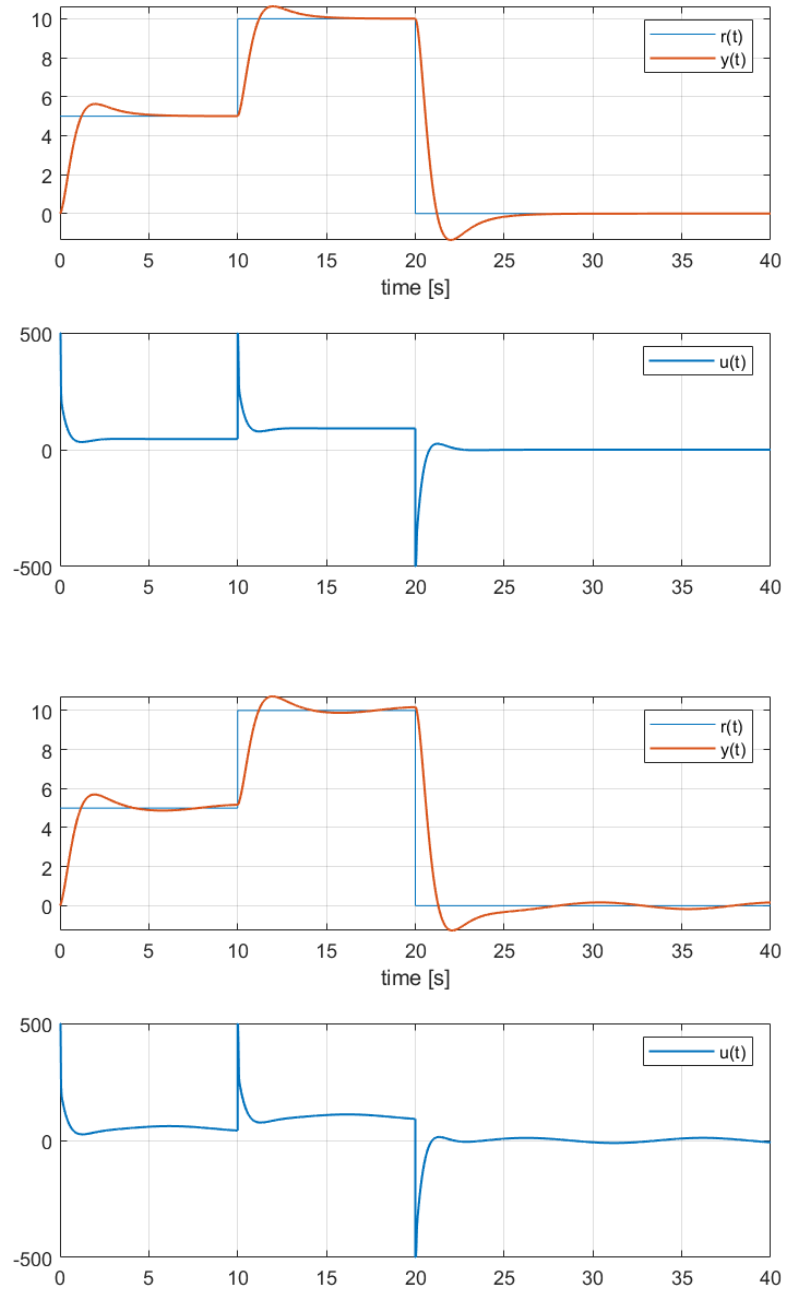


Figure 15.6: Top: closed-loop control of (15.1) from Example 15.1, when the plant is known with a modelling error. Bottom: the same, when the output is affected by sinusoidal disturbance (15.4).

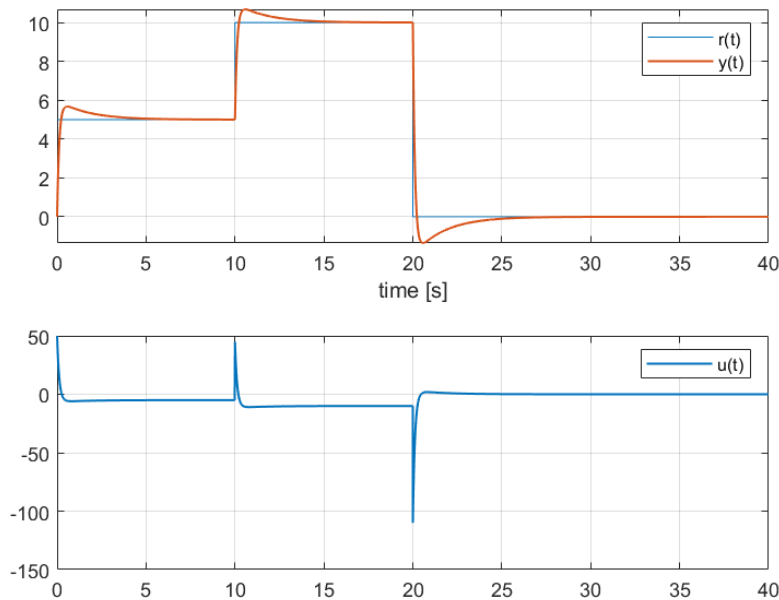


Figure 15.7: Closed-loop control of (15.6) from Example 15.4.

```
s^3 + 4 s^2 + 6 s + 29
```

Continuous-time transfer function.

```
>> pole(closed_loop)
```

```
ans =
```

```
-4.2107 + 0.0000i
0.1054 + 2.6222i
0.1054 - 2.6222i
```

and thus the closed-loop is unstable, because of a badly designed controller. \square

Remark 15.1. Assuming that the controller of Example 15.5 is proportional, *Routh-Hurwitz criterion and proportional control* $C(s) = K \in \mathbb{R}$, it is easy to use the Routh-Hurwitz criterion to find which values of K would ensure a stable closed-loop:

$$\begin{aligned} \frac{y(s)}{r(s)} &= \frac{\frac{K}{s^3+4s^2+6s+4}}{1 + \frac{K}{s^3+4s^2+6s+4}} \\ &= \frac{K}{s^3 + 4s^2 + 6s + (4 + K)} \end{aligned} \quad (15.11)$$

$$\begin{array}{c|cc} s^3 & 1 & 6 \\ s^2 & 4 & 4 + K \\ \hline s & \frac{4 \times 6 - (4 + K)}{4} & \\ 1 & 4 + K & \end{array} \quad (15.12)$$

So as to have all the elements in the first column positive,

$$\frac{24 - 4 - K}{4} > 0 \Rightarrow K < 20 \quad (15.13)$$

$$4 + K > 0 \Rightarrow K > -4 \quad (15.14)$$

Consequently, in Example 15.5, controller $C(s) = K = 25$ caused an unstable closed-loop, since the loop will be stable only if $K \in]-4, 20[$.

Also check Example 11.12 again. \square

Of course, unlike open-loop control, closed-loop control requires a measuring chain for the feedback branch — and, while it is possible to cope with noise, noise still reduces performance, and completely wrong measurements prevent control systems from working.

15.3 Design of open loop controllers

From Figure 9.13 it is clear that

$$y(s) = G(s)C(s)r(s) \quad (15.15)$$

and thus if we want the output $y(s)$ to follow the reference $r(s)$ perfectly, i.e. $y(s) = r(s)$, all we have to do is

$$C(s) = \frac{1}{G(s)} \quad (15.16)$$

and perfect control is achieved.

Open-loop controller design in practice

In practice things are not that simple. As we saw in Chapter 11, plant models ought to be strictly proper, i.e. to have more poles than zeros. If so, applying (15.16) we get a controller $C(s)$ which is not proper, i.e. has more zeros than poles. This controller, as we know, has a behaviour at high frequencies that is impossible. Consequently, additional poles must be added, so that $C(s)$ will be strictly proper, or at least proper. Ideally, additional poles should have a frequency as high as possible, preferably higher than the frequency of every zero and pole in $G(s)$, so as to not alter the desired behaviour in the frequencies where the model is valid (and in which the controller must thus do its job). Of course, higher frequencies will mean faster responses, and faster responses mean higher values of the control actions, so even if the actuator can respond fast enough it may saturate if poles are too far away.

Example 15.6. Suppose that we want to control plant

$$G(s) = \frac{s+1}{s(s+10)} \quad (15.17)$$

in open loop. It has one zero and two poles; so, if the controller is to be proper, it needs an additional pole. One decade above the highest frequency zero or pole would place it at 100 rad/s. But assume that control actions would be too large, or that the actuator does not respond that fast, and we are left with

$$C_1(s) = \frac{20s(s+10)}{(s+1)(s+20)} \quad (15.18)$$

or even with

$$C_2(s) = \frac{10s}{s+1} \quad (15.19)$$

which corresponds to an additional pole at 10 rad/s, which then cancels the zero. Figure 15.8 shows the results obtained with both controllers when the reference is the same of previous examples. \square

Combining open and closed loop control

It is possible to combine open-loop and closed-loop control in a single control system.

Example 15.7. Plant (15.17) from Example 15.6, with a 10% error in the steady-state gain, can be controlled combining open-loop controller (15.19) with a closed-loop proportional controller, as seen in Figure 15.9. The result, when there is a sinusoidal output disturbance given by (15.4), is shown in Figure 15.10, and is clearly better than results got with only one of the controllers, shown in Figure 15.11. \square

15.4 Closed loop controllers

The design of closed loop controllers is not as trivial, and will occupy the remaining chapters of this Part. We will study right away the most common forms that closed loop controllers take.

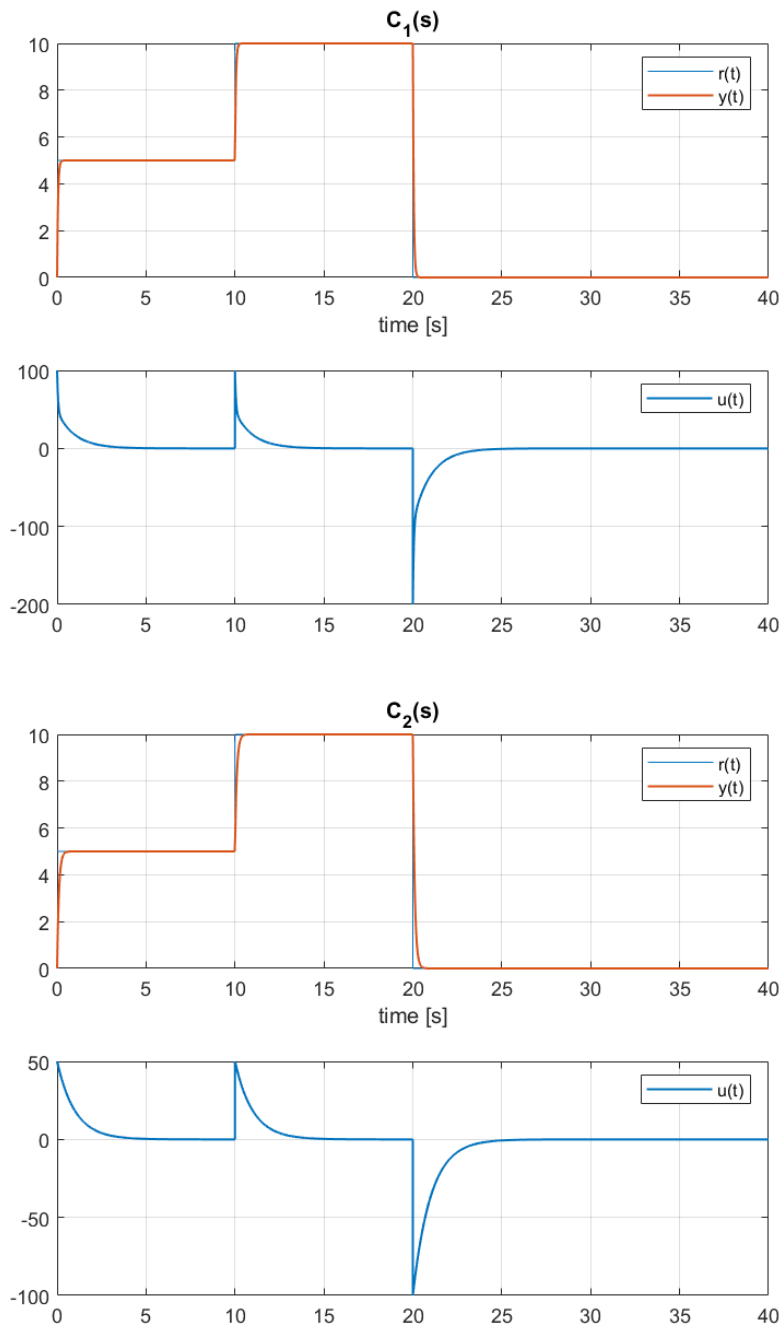


Figure 15.8: Closed-loop control of (15.17) from Example 15.6.

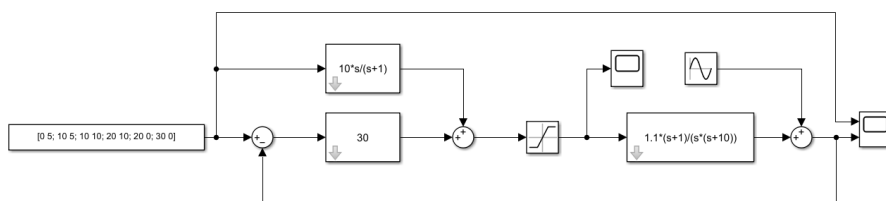


Figure 15.9: Simulink file with which simulation results in Figures 15.10 and 15.11 were obtained, corresponding to Example 15.7.

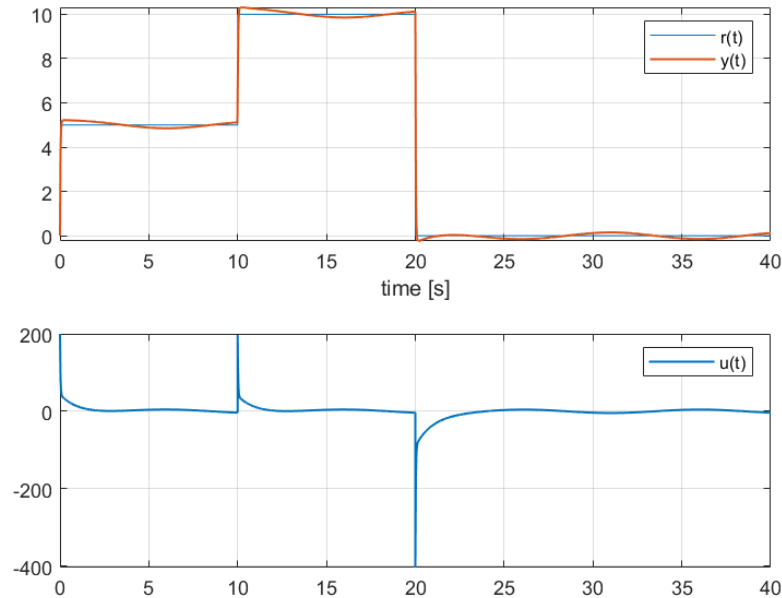


Figure 15.10: Control of (15.17) from Example 15.7, combining open-loop and closed-loop control as seen in Figure 15.9.

P controllers

- **Proportional controllers** are already known to us. They apply a control action

$$u(t) = K e(t), \quad K \neq 0 \quad (15.20)$$

$$\Rightarrow u(s) = K e(s) \quad (15.21)$$

$$\Rightarrow \frac{u(s)}{e(s)} = C(s) = K \quad (15.22)$$

proportional to the loop error: the larger the error, the larger the control action. Notice that a proportional controller is a static system.

PI controllers

- **Proportional–integral controllers** apply a control action which is the sum of two control actions:

- one proportional to the loop error,
- one proportional to the integral of the loop error:

$$u(t) = K_p e(t) + K_i \int_0^t e(t) dt, \quad K_p, K_i \neq 0 \quad (15.23)$$

$$\Rightarrow u(s) = K_p e(s) + \frac{K_i}{s} e(s) \quad (15.24)$$

$$\Rightarrow \frac{u(s)}{e(s)} = C(s) = K_p + \frac{K_i}{s} \quad (15.25)$$

The reasoning behind this second term is that, in this way, if proportional control achieves a steady-state error, since the integral of the error will grow with time, the control action will also keep increasing, eliminating the error. We will see in Chapter 20 when this actually works and when it does not.

PD controllers

- **Proportional–derivative controllers** apply a control action which is the sum of two control actions:

- one proportional to the loop error,
- one proportional to the derivative of the loop error:

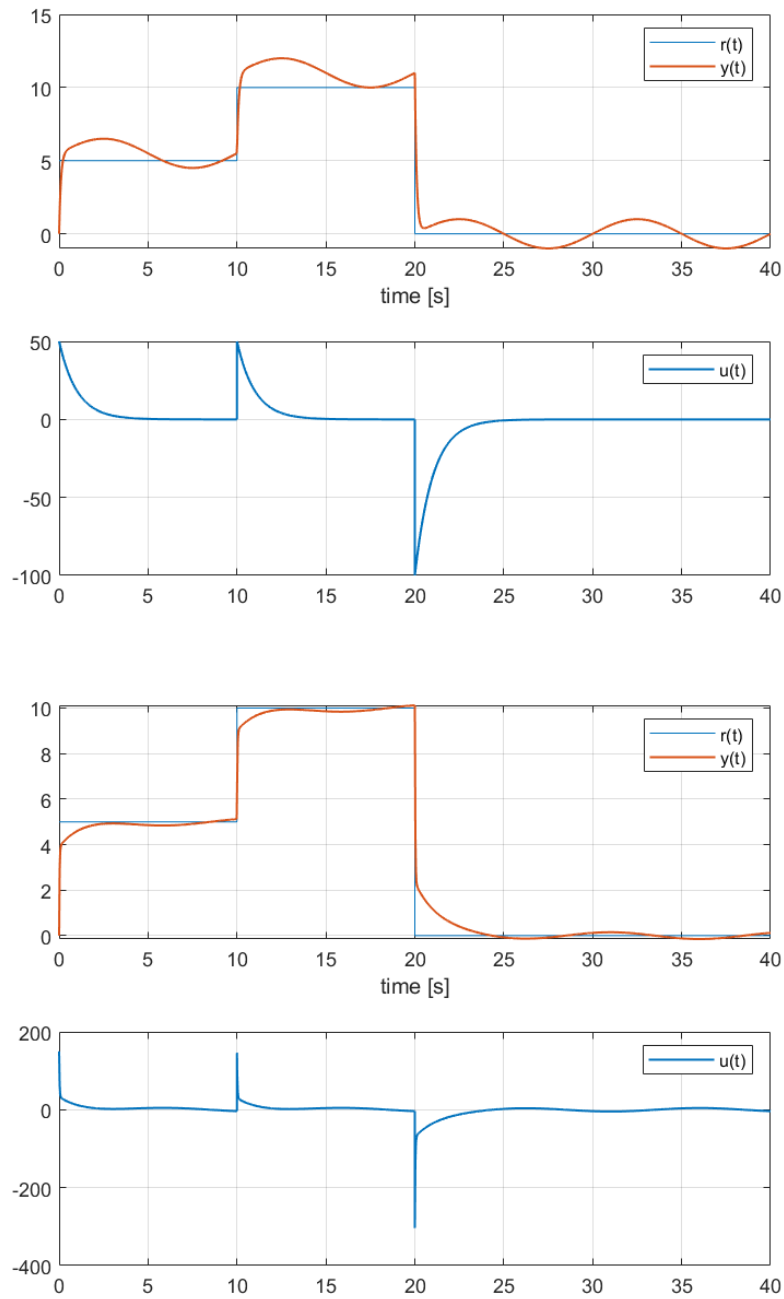


Figure 15.11: Control of (15.17) from Example 15.7. Top: open-loop control only. Bottom: closed-loop control only.

$$u(t) = K_p e(t) + K_d \frac{de(t)}{dt} dt, \quad K_p, K_d \neq 0 \quad (15.26)$$

$$\Rightarrow u(s) = K_p e(s) + K_d s e(s) \quad (15.27)$$

$$\Rightarrow \frac{u(s)}{e(s)} = C(s) = K_p + K_d s e(s) \quad (15.28)$$

The reasoning behind this second term is that, in this way, if there is a sudden increase of the error, the control action will immediately increase to counter it, rather than waiting for a large error (or, even worse, for a large integral of the error) to do so. We will see in the next chapters when this makes sense and when it does on.

PID controllers

- **Proportional–integral–derivative controllers** apply a control action which is the sum of three control actions:

$$u(t) = K_p e(t) + K_i \int_0^t e(t) dt + K_d \frac{de(t)}{dt} dt \quad (15.29)$$

$$\Rightarrow u(s) = K_p e(s) + \frac{K_i}{s} e(s) + K_d s e(s) \quad (15.30)$$

$$\Rightarrow \frac{u(s)}{e(s)} = C(s) = K_p + \frac{K_i}{s} + K_d s \quad (15.31)$$

This combines the desired advantages of both PI and PD control. Of course,

- proportional control is a particular case of PID control, in which $K_i = K_d = 0$;
- PI control is a particular case of PID control, in which $K_d = 0$;
- PD control is a particular case of PID control, in which $K_i = 0$.

PID family

- The **PID family** of controllers includes not only

- proportional controllers,
- PI controllers,
- PD controllers,
- PID controllers,

but also similar controllers with more than one derivative part (though this is seldom found) or (more often) more than one integral part, such as

- PI²D controllers:

$$C(s) = K_p + \frac{K_{i1}}{s} + \frac{K_{i2}}{s^2} + K_d s \quad (15.32)$$

- PI³D controllers:

$$C(s) = K_p + \frac{K_{i1}}{s} + \frac{K_{i2}}{s^2} + \frac{K_{i3}}{s^3} + K_d s \quad (15.33)$$

- PID² controllers:

$$C(s) = K_p + \frac{K_i}{s} + K_{d1} s + K_{d2} s^2 \quad (15.34)$$

and so on. We will see in Chapter 20 why additional integral parts may be needed. Chapter 21 is devoted to design methods for the PID family of controllers.

- **Lead-lag controllers** have one pole, one zero, and a positive gain that tends to zero at either low or high frequencies:

Lead controllers

- Lead controllers are given by

$$C(s) = \frac{\alpha s + a}{s + a}, \quad \alpha > 1, a > 0 \quad (15.35)$$

– Lag controllers are given by

$$C(s) = \frac{s + a}{s + \frac{a}{\alpha}}, \quad \alpha > 1, a > 0 \quad (15.36)$$

So, both (15.35) and (15.36) correspond to particular cases of

$$C(s) = K \frac{s + b}{s + p} \quad (15.37)$$

with some restrictions on the parameters (gain K , pole $-p$, and zero $-b$). We will see in Chapters 18 and 20 why we want these controllers, and will learn how to design them in Chapter 22. In that Chapter we plot the Bode diagrams of (15.35) and (15.36), and see that they have, respectively, phases which are positive (lead) and negative (lag) — hence the names.

In the following chapters we will see abundant examples of controllers of all these types (and of other types too).

Meanwhile notice that

Proper controllers have additional poles

- lead-lag controllers (15.35)–(15.36) and PI controllers (15.25) are proper but not strictly proper,
- PD controllers (15.28) and PID controllers (15.31) are not proper,

and so all require additional poles to be implemented in practice. We will see in Chapter 29 more details on this.

Glossary

I have remarked that the paper had fallen away in parts. In this particular corner of the room a large piece had peeled off, leaving a yellow square of coarse plastering. Across this bare space there was scrawled in blood-red letters a single word:—

RACHE

“What do you think of that?” cried the detective, with the air of a showman exhibiting his show. “This was overlooked because it was in the darkest corner of the room, and no one thought of looking there. The murderer has written it with his or her own blood. See this smear where it has trickled down the wall! That disposes of the idea of suicide anyhow. Why was that corner chosen to write it on? I will tell you. See that candle on the mantelpiece. It was lit at the time, and if it was lit this corner would be the brightest instead of the darkest portion of the wall.”

“And what does it mean now that you *have* found it?” asked Gregson in a depreciatory voice.

“Mean? Why, it means that the writer was going to put the female name Rachel, but was disturbed before he or she had time to finish. You mark my words, when this case comes to be cleared up you will find that a woman named Rachel has something to do with it. It’s all very well for you to laugh, Mr. Sherlock Holmes. You may be very smart and clever, but the old hound is the best, when all is said and done.”

(...)

“One other thing, Lestrade,” he added, turning round at the door, “ ‘Rache’ is the German for ‘revenge,’ so don’t lose your time by looking for Miss Rachel.”

Sir Arthur CONAN DOYLE (1859 — †1930), *A study in scarlet* (1887), I 3

lag controller controlador de atraso
lead controller controlador de avanço
lead-lag controller controlador de avanço-atraso
proportional controller controlador proporcional

proportional–derivative controller controlador proporcional–derivativo
proportional–integrative controller controlador proporcional–integral
proportional–integrative–derivative controller controlador proporcional–
integral–derivativo

Exercises

1. A train can travel at a maximum speed of 115.2 km/h, and its maximum acceleration is ± 0.8 m/s².
 - (a) What is the minimum time that the train takes to travel between two stations 6.4 km apart? Draw the evolution of velocity with time.
 - (b) Assume that larger accelerations (or decelerations) mean a larger consumption of energy. (In fact things are more complicated in real life.) If, due to timetable constraints, the train must take exactly six minutes to travel between the same two stations, how should the velocity change with time so that energy consumption is as low as possible?

Chapter 16

Root locus

Minos' daughter Ariadne was among the spectators and she fell in love with Theseus at first sight as he marched past her. She sent for Daedalus and told him he must show her a way to get out of the Labyrinth, and she sent for Theseus and told him she would bring about his escape if he would promise to take her back to Athens and marry her. As may be imagined, he made no difficulty about that, and she gave him the clue she had got from Daedalus, a ball of thread which he was to fasten at one end to the inside of the door and unwind as he went on. This he did and, certain that he could retrace his steps whenever he chose, he walked boldly into the maze looking for the Minotaur.

Edith HAMILTON (1867 — †1963), *Mythology* (1942), III 9

The **root locus diagram**:

- shows the location (or locus) in the complex plane of the **poles of a closed loop** (i.e. the *roots* of the denominator of the closed loop transfer function),
- when the **gain of the open loop** changes.

It can be used in one of the following situations (see Figure 16.1):

What the root locus diagram is for

- to design a proportional controller,
- to design the gain of a controller that has poles and zeros, once these are known,
- to see what happens to a closed loop control system with a known controller, when the gain of the plant changes for some reason.

As we already saw, the location of the poles of the closed loop control system:

- lets us know if it is stable (remember Section 10.3);
- can give us an idea of how its time responses will be (remember Section 11.6).

16.1 Simple examples

Example 16.1. Let plant $\frac{y(s)}{u(s)} = \frac{1}{s+1}$ be controlled in closed loop by proportional controller $\frac{u(s)}{e(s)} = K$. The closed loop transfer function is

$$\frac{y(s)}{r(s)} = \frac{\frac{K}{s+1}}{1 + \frac{K}{s+1}} = \frac{K}{s+1+K} \quad (16.1)$$

and thus has always only one pole, located at

$$s = -1 - K \quad (16.2)$$

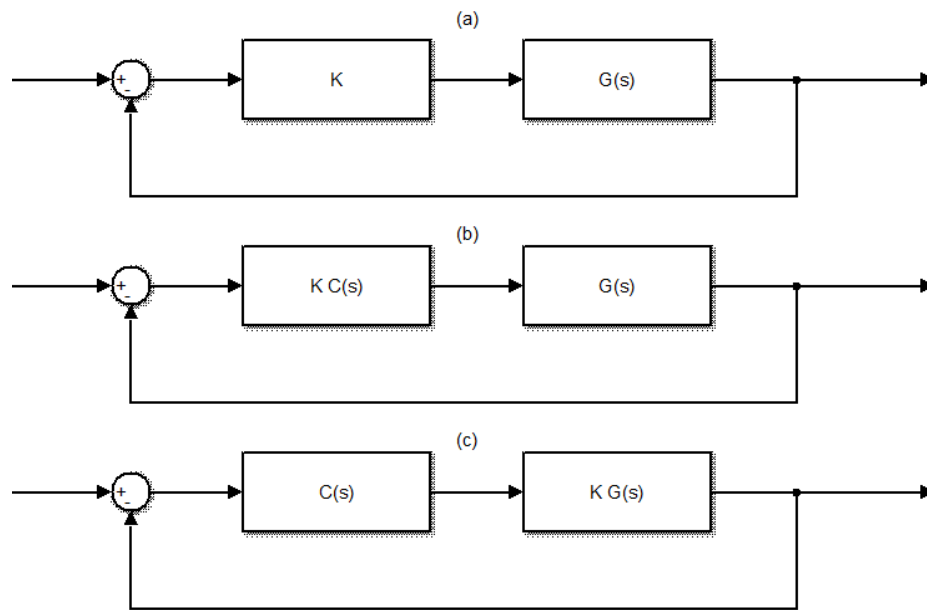


Figure 16.1: Situations in which the root locus diagram is useful. (a) The plant $G(s)$ is known. The controller K is proportional. We want to set a reasonable value for K . (b) The plant $G(s)$ is known. The controller $K \times C(s)$ has known poles and zeros (collected in transfer function $C(s)$). We want to set a reasonable value for controller gain K . (c) The controller $C(s)$ is known. The plant $K \times G(s)$ has known poles and zeros (collected in transfer function $G(s)$). We want to know what happens when the gain of the plant K changes for some reason. (Notice that in this case we cannot change K at will, otherwise it would be part of the controller, not the plant.)

Table 16.1: Variation of the closed loop pole with open loop gain K in Example 16.1.

K	5	4	3	2	1	0	-1	-2	-3	-4	-5
pole	-6	-5	-4	-3	-2	-1	0	1	2	3	4

This pole will be stable if

$$s < 0 \Leftrightarrow -1 - K < 0 \Leftrightarrow K > -1 \quad (16.3)$$

Notice that, if $K = 0$, the closed loop pole (16.2) will be $K = -1$, which is the open loop pole too. But in that case the numerator of the closed loop (16.1) is zero, the control action $u(t)$ will be always zero, and the control system will not work; there is not, properly speaking, a closed loop when $K = 0$.

We can give different values to K and find the values for the closed loop pole in Table 16.1, which are then plotted in Figure 16.2. Plotting more points is easy using two `for` cycles in MATLAB, one for positive and one for negative values of K , with the result also shown in Figure 16.2:

```
G = tf(1,[1 1]);
% positive values of K
poles_closed_loop = [];
for K = 0 : 0.25 : 6
poles_closed_loop = [poles_closed_loop; pole(feedback(K*G, 1))];
end
figure, plot(real(poles_closed_loop), imag(poles_closed_loop), 'bx')
% negative values of K
poles_closed_loop = [];
for K = 0 : -0.25 : -6
poles_closed_loop = [poles_closed_loop; pole(feedback(K*G, 1))];
end
hold on, plot(real(poles_closed_loop), imag(poles_closed_loop), 'rx')
grid on, xlabel('Real axis'), ylabel('Imaginary axis'), legend({'K>0','K<0'})
```

Of course, the **root locus diagram** corresponds to an infinitely small resolution in K . This diagram shows what happens to the closed loop pole with the variation of K . The last plot of Figure 16.2 was obtained with MATLAB using commands

```
>> s = tf('s');
>> figure,rlocus(1/(s+1))
>> hold on, rlocus(-1/(s+1),'g') % the 'g' option forces the green colour
>> legend({'K>0','K<0'})
```

MATLAB *command*
rlocus

Function `rlocus` always assumes a positive open loop gain K ; that is why it was used twice, the second time with the minus sign inserted in the open loop itself. For this simple open loop transfer function with only one pole, the root locus could also have been easily plotted by hand from (16.2). \square

Example 16.2. Let plant $\frac{y(s)}{u(s)} = \frac{1}{(s+4)(s+2)}$ be controlled in closed loop by proportional controller $\frac{u(s)}{e(s)} = K$. The closed loop transfer function is

$$\frac{y(s)}{r(s)} = \frac{\frac{K}{s^2+6s+8}}{1 + \frac{K}{s^2+6s+8}} = \frac{K}{s^2 + 6s + 8 + K} \quad (16.4)$$

There are two poles, located at

$$s = \frac{-6 \pm \sqrt{36 - 32 - 4K}}{2} = -3 \pm \sqrt{1 - K} \quad (16.5)$$

These poles verify the following:

- If $K = 0$, they are those of the open loop, $s = -3 \pm 1 \Leftrightarrow s = -4 \vee s = -2$, though once more the numerator of the closed loop (16.4) is zero, so there is in fact no closed loop at all.

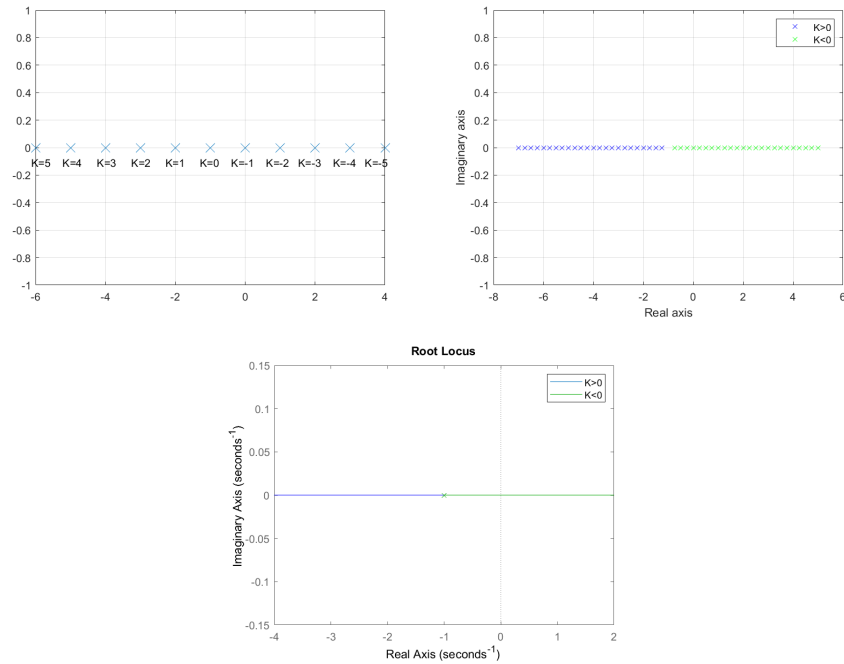


Figure 16.2: Top left: pole of the close loop consisting of plant $\frac{y(s)}{u(s)} = \frac{1}{s+1}$ and a proportional controller K , for the cases in Table 16.1. Top right: the same, calculated with for cycles for more values of K . Bottom: root locus diagram of the plant.

- If $0 < K < 1$, they are real.
- If $K = 1$, they coincide, $s = -1$.
- If $K > 1$, they are imaginary, having a constant real part $\Re(s) = -1$;
- If $K < 0$, they are real. If $-3 + \sqrt{1 - K} > 0 \Rightarrow 1 - K > 9 \Leftrightarrow K < -8$, one of them will be positive.

These conclusions can be qualitatively seen in the root locus diagram, which can be plot as

```
G = tf(1,conv([1 4],[1 2]));
% positive values of K
poles_closed_loop = [];
for K = 0 : 0.2 : 10
    poles_closed_loop = [poles_closed_loop; pole(feedback(K*G, 1))];
end
figure, plot(real(poles_closed_loop), imag(poles_closed_loop), 'bx')
% negative values of K
poles_closed_loop = [];
for K = 0 : -0.2 : -10
    poles_closed_loop = [poles_closed_loop; pole(feedback(K*G, 1))];
end
hold on, plot(real(poles_closed_loop), imag(poles_closed_loop), 'gx')
grid on, xlabel('Real axis'), ylabel('Imaginary axis'), legend({'K>0','K<0'})
```

or better still as

```
s = tf('s');
G = 1/((s+2)*(s+4));
figure,rlocus(G,-G)
legend({'K>0','K<0'})
```

and can be seen in Figure 16.3.

Table 16.2: Rules to plot the root locus of $G(s)$

1. $G(s) = \frac{b_m s^m + b_{m-1} s^{m-1} + \dots + b_0}{a_n s^n + a_{n-1} s^{n-1} + \dots + a_0}$ must be proper, i.e. $n \geq m$. Write its product by the varying gain $K \in \mathbb{R}$ in the form

$$KG(s) = \frac{K b_m}{a_n} \frac{(s - z_1) \dots (s - z_m)}{(s - p_1) \dots (s - p_n)}, \quad K \in \mathbb{R} \quad (16.6)$$

Two cases must be considered, corresponding to $\frac{K b_m}{a_n} > 0$ and $\frac{K b_m}{a_n} < 0$.

2. Plot the n open loop poles on the complex plane using a cross \times , and the m open loop zeros using a circle \circ .
3. The root locus diagram is symmetric in relation to the real axis.
4. If $n - m \geq 2$, the centroid of the root locus is constant, irrespective of the value of K .
5. Points on the real axis belong to the root locus
 - if there is an odd number of open loop poles and zeros to its right, and $\frac{K b_m}{a_n} > 0$;
 - if there is an even number of open loop poles and zeros to its right, or n no poles and zeros at all, and $\frac{K b_m}{a_n} < 0$.
6. The diagram has n branches, each of which corresponds to a closed loop pole.
7. All branches converge to the open loop poles when $K \rightarrow 0$.
8. m branches converge to the open loop zeros when $|K| \rightarrow \infty$.
9. $n - m$ branches diverge to infinity when $|K| \rightarrow \infty$, following asymptotes that make with the positive real axis angles of

$$\gamma = \begin{cases} \frac{180^\circ(2k+1)}{n-m}, & k = 0, \dots, n-m-1, & \text{if } \frac{K b_m}{a_n} > 0 \\ \frac{360^\circ k}{n-m}, & k = 0, \dots, n-m-1, & \text{if } \frac{K b_m}{a_n} < 0 \end{cases} \quad (16.7)$$

and intersect on the real axis at point

$$\sigma = \frac{\sum_{k=1}^n p_k - \sum_{k=1}^m z_k}{n-m} \quad (16.8)$$

10. The branches of the root locus converge or diverge on the real axis, perpendicularly thereto, at the real roots of

$$\frac{d}{ds} \frac{1}{G(s)} = 0 \Leftrightarrow \frac{d}{ds} \frac{(s-p_1) \dots (s-p_n)}{(s-z_1) \dots (s-z_m)} = 0 \quad (16.9)$$

11. In the neighbourhood of a complex open loop pole p_i , branches have an asymptote with an angle ϕ_i given by

$$\phi_i = \begin{cases} \sum_{k=1}^m \psi_k - \sum_{\substack{k=1 \\ k \neq i}}^n \phi_k + 180^\circ, & \text{if } \frac{K b_m}{a_n} > 0 \\ \sum_{k=1}^m \psi_k - \sum_{\substack{k=1 \\ k \neq i}}^n \phi_k, & \text{if } \frac{K b_m}{a_n} < 0 \end{cases} \quad (16.10)$$

where

$$\phi_k = \angle[p_i - p_k], \quad k = 1, \dots, n, \quad k \neq i \quad (16.11)$$

$$\psi_k = \angle[p_i - z_k], \quad k = 1, \dots, m \quad (16.12)$$

as seen in Figure 16.4.

12. In the neighbourhood of a complex open loop zero z_i , branches have an asymptote with an angle ψ_i given by

$$\psi_i = \begin{cases} \sum_{k=1}^n \phi_k - \sum_{\substack{k=1 \\ k \neq i}}^m \psi_k + 180^\circ, & \text{if } \frac{K b_m}{a_n} > 0 \\ \sum_{k=1}^n \phi_k - \sum_{\substack{k=1 \\ k \neq i}}^m \psi_k, & \text{if } \frac{K b_m}{a_n} < 0 \end{cases} \quad (16.13)$$

where

$$\phi_k = \angle[z_i - p_k], \quad k = 1, \dots, n \quad (16.14)$$

$$\psi_k = \angle[z_i - z_k], \quad k = 1, \dots, m, \quad k \neq i \quad (16.15)$$

as seen in Figure 16.4.

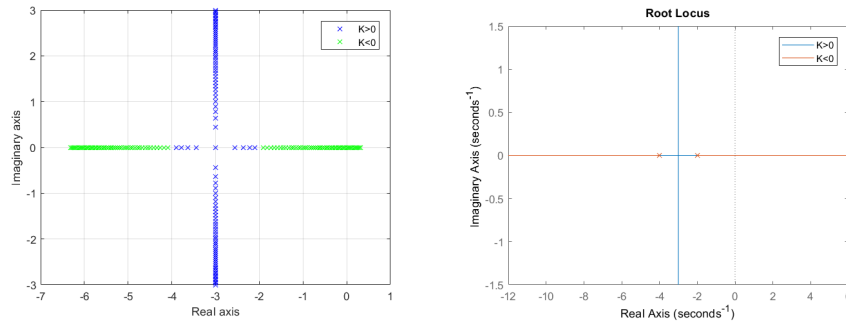


Figure 16.3: Left: root locus of $\frac{1}{(s+4)(s+2)}$ obtained with `for cycles`. Right: the same, obtained with `rlocus`.

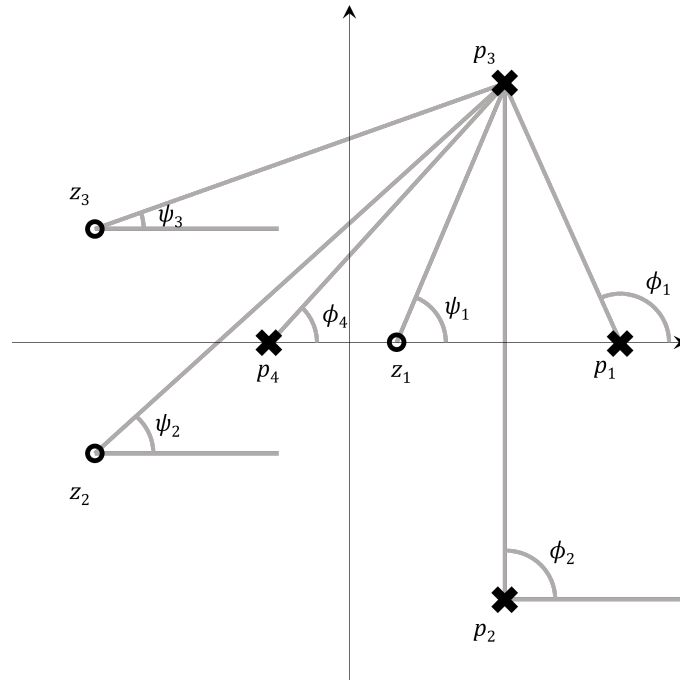


Figure 16.4: Measuring the angles for rules 11 and 12 from Table 16.2.

16.2 Rules for the root locus

Root locus diagrams can be easily found with MATLAB, but it is also possible to draw them by hand. Drawing root locus diagrams by hand is as important as drawing Bode diagrams by hand: of course both can be accurately found using MATLAB, but hand-drawn diagrams help to gain sensibility to the effects of poles and zeros and of their location in the complex plane.

We present a systematic set of rules to draw root locus diagrams in Table 16.2, together with some simple examples. In the next section, we will see where these rules come from.

Example 16.3. Consider plant

$$G(s) = \frac{20}{s^2 + 2s + 2} \quad (16.16)$$

To plot its root locus by hand, we follow the steps in Table 16.2:

1. We rewrite the open loop transfer function as $G(s) = 20K \frac{1}{(s+1+j)(s+1-j)}$. The two cases we must consider are $20K > 0 \Leftrightarrow K > 0$ and $20K < 0 \Leftrightarrow K < 0$.
2. There are $n = 2$ poles, a complex conjugate pair at $-1 \pm j$. There are no zeros, i.e. $m = 0$.
3. We know that the diagram is always symmetric in relation to the real axis.

4. Because $n - m = 2$, the centroid will be constant. So, when a closed loop pole moves in one direction, the other must move in the same direction but in the opposite sense, so that the centroid remains the same.
5. Real points to the right of $s = -1$ have no open loop poles to their right, i.e.

$$\forall s > -1, s > \Re(p_1) \wedge s > \Re(p_2) \quad (16.17)$$

Consequently, that part of the real axis belongs to the root locus when $K < 0$. Real points to the left of $s = -1$ have two open loop poles to their right, i.e.

$$\forall s < -1, s < \Re(p_1) \wedge s < \Re(p_2) \quad (16.18)$$

Since 2 is an even number, that part of the real axis also belongs to the root locus when $K < 0$. No part of the real axis belongs to the root locus when $K > 0$. We can now draw the first, incomplete sketch of the root locus diagram in Figure 16.5.

6. The diagram has $n = 2$ branches, i.e. the closed loop has 2 poles.
7. Both branches converge to $-1 \pm j$ when $K \rightarrow 0$.
8. Since there are no open loop zeros, no branches of the root locus converge to them.
9. $n - m = 2$ branches, i.e. both of them, diverge to infinity when K is large. When $K > 0$, they follow asymptotes with slopes given by the angles

$$\gamma_0 = \frac{180^\circ(2 \times 0 + 1)}{2} = 90^\circ \quad (16.19)$$

$$\gamma_1 = \frac{180^\circ(2 \times 1 + 1)}{2} = 270^\circ \equiv -90^\circ \quad (16.20)$$

It would have sufficed to find $\gamma_0 = 90^\circ$; since the root locus is symmetric in relation to the real axis, the other angle would have to be -90° . When $K > 0$, the asymptotes correspond to angles

$$\gamma_0 = \frac{360^\circ \times 0}{2} = 0^\circ \quad (16.21)$$

$$\gamma_1 = \frac{360^\circ \times 1}{2} = 180^\circ \quad (16.22)$$

These asymptotes are already in the plot, as they are in fact the real axis itself. Asymptotes pass through the real point

$$\sigma = \frac{-1 + j - 1 - j}{2} = -1 \quad (16.23)$$

Notice that we might have considered only the real parts; the imaginary parts are bound to cancel out. We now know that the vertical asymptotes, the ones with the angle $\pm 90^\circ$ for $K > 0$, pass through $\sigma = -1$. We know enough to draw the complete root locus diagram in Figure 16.5. Mind the arrows indicating the increasing values of K .

10. It is obvious that the point of divergence on the real axis is -1 , but we can confirm this calculating

$$\frac{d}{ds} \frac{s^2 + 2s + 2}{20} = 0 \Leftrightarrow 2s + 2 = 0 \Leftrightarrow s = -1 \quad (16.24)$$

11. It is also obvious that the branches arrive at the complex conjugate open loop poles when $K \rightarrow 0^-$ and leave them when K increases from 0 with vertical asymptotes (after all, the vertical asymptotes for $K \rightarrow +\infty$ pass right through the open loop poles themselves). But we can confirm this as follows: let $p_1 = -1 - j$ and $p_2 = -1 + j$; then the angles of the asymptotes around p_2 are found as

$$\phi_1 = 90^\circ \Rightarrow \begin{cases} \phi_2 = 180^\circ - 90^\circ = 90^\circ, & K > 0 \\ \phi_2 = -90^\circ, & K < 0 \end{cases} \quad (16.25)$$

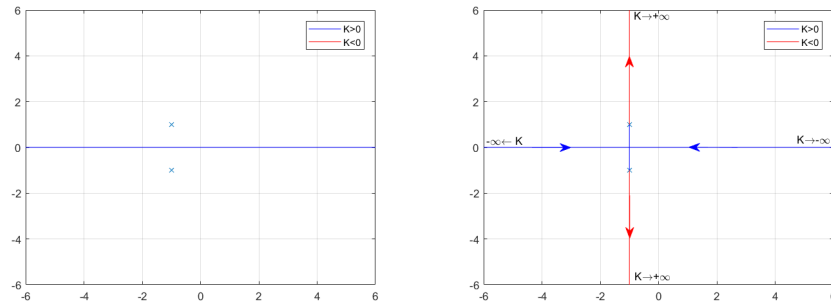


Figure 16.5: Root locus of (16.16) from Example 16.3. Left: incomplete sketch after step 5. Right: complete sketch after step 9.

Because of the symmetry, we can tell right away that

$$\begin{cases} \phi_1 = -90^\circ, & K > 0 \\ \phi_1 = 90^\circ, & K < 0 \end{cases} \quad (16.26)$$

12. There are no complex conjugate open loop poles, so this rule does not apply.

It is clear from the root locus that

- $G(s)$ can be controlled in closed loop with a proportional controller $K > 0$, being always stable;
- there will always be oscillations in the controlled system, since the closed loop poles will always be complex;
- the larger the value of K , the more oscillations the controlled system will have;
- a negative proportional controller may result in real poles, if K is negative enough;
- however, for values of K which are negative and very large, the controlled system will become unstable, since one of the closed loop poles will be positive. We can find the value of K below which the closed loop is unstable applying to the closed loop transfer function

$$\begin{aligned} \frac{y(s)}{r(s)} &= \frac{\frac{20K}{s^2+2s+2}}{1 + \frac{20K}{s^2+2s+2}} \\ &= \frac{20K}{s^2 + 2s + (2 + 20K)} \end{aligned} \quad (16.27)$$

the Routh-Hurwitz criterion:

$$\begin{array}{c|cc} s^2 & 1 & 2 + 20K \\ s & 2 & \\ \hline 1 & 2 + 20K & \end{array} \quad (16.28)$$

We see that the closed loop is stable if

$$2 + 20K > 0 \Leftrightarrow K > -0.1 \quad (16.29)$$

- for this gain value $K = -0.1$ which makes the closed loop marginally stable, there will be a pole at $s = 0$. \square

Example 16.4. The root locus of plant

$$G(s) = \frac{s + 4}{(-s + 1)(s + 3)(s^2 + 4s + 8)} \quad (16.30)$$

is found as follows.

1. We rewrite the open loop transfer function as

$$G(s) = -K \frac{s + 4}{(s - 1)(s + 3)(s + 2 + 2j)(s + 2 - 2j)} \quad (16.31)$$

The two cases we must consider are $-K > 0 \Leftrightarrow K < 0$ and $-K < 0 \Leftrightarrow K > 0$. Notice how the unstable pole causes the two cases to switch in what the sign of K is concerned.

2. There are $n = 4$ poles, one unstable at $s = 1$, a stable real pole at $s = -3$, and a pair of complex conjugate stable poles at $s = -2 \pm 2j$. There is $m = 1$ zero, which is a minimum phase zero, at $s = -4$.
3. We know that the diagram is always symmetric in relation to the real axis.
4. Because $n - m = 3$, the centroid will be constant. So, when a closed loop pole moves in one direction, the others must move in such a way that the centroid remains the same.
5. Real points verify the following:
 - Those to the right of $s = 1$ have no open loop poles to their right, so that part of the real axis belongs to the root locus when $-K < 0 \Leftrightarrow K > 0$.
 - Those in $] - 2, 1[$ have one open loop pole to their right, so as 1 is odd that part of the real axis belongs to the root locus when $-K > 0 \Leftrightarrow K < 0$.
 - Those in $] - 3, -2[$ have three open loop poles to their right, so as 3 is odd that part of the real axis also belongs to the root locus when $-K > 0 \Leftrightarrow K < 0$.
 - Those in $] - 4, -3[$ have four open loop poles to their right, so as 4 is even that part of the real axis belongs to the root locus when $-K < 0 \Leftrightarrow K > 0$.
 - Those to the left of $s = -4$ have to their right four open loop poles and one open loop zero, so as the total 5 is odd that part of the real axis belongs to the root locus when $-K > 0 \Leftrightarrow K < 0$.

6. The diagram has $n = 4$ branches, i.e. the closed loop has 4 poles.
7. The four branches converge to the open loop poles $s = 1$, $s = -3$, $s = -2 \pm 2j$ when $K \rightarrow 0$.
8. One of the branches will converge to the open loop zero $s = -4$ when $|K| \rightarrow \infty$. Since the zero is real, this convergence is already shown in the sketch above.
9. $n - m = 3$ branches diverge to infinity when K is large. When $K > 0$, they follow asymptotes with slopes given by the angles

$$\gamma_0 = \frac{360^\circ \times 0}{2} = 0^\circ \quad (16.32)$$

$$\gamma_1 = \frac{360^\circ \times 1}{3} = 120^\circ \quad (16.33)$$

$$\gamma_2 = \frac{360^\circ \times 2}{3} = 240^\circ \equiv -120^\circ \quad (16.34)$$

Notice that, because the root locus is symmetric in relation to the real axis, it suffices to find the angles between 0° and 180° ; the others are symmetric. When $K < 0$,

$$\gamma_0 = \frac{180^\circ(2 \times 0 + 1)}{3} = 60^\circ \quad (16.35)$$

$$\gamma_1 = \frac{180^\circ(2 \times 1 + 1)}{3} = 180^\circ \quad (16.36)$$

and of course $\gamma_2 = -60^\circ \equiv 300^\circ$. Asymptotes pass through the real point

$$\sigma = \frac{1 - 3 - 2 - 2 - (-4)}{3} = -\frac{2}{3} \quad (16.37)$$

We can now draw the first, incomplete sketch of the root locus diagram in Figure 16.6.

10. The points of divergence or convergence on the real axis are found solving

$$\begin{aligned} & \frac{d}{ds} \frac{\overbrace{(-s+1)(s+3)(s^2+4s+8)}^{-s^4-6s^3-13s^2-4s+24}}{s+4} = 0 \\ \Leftrightarrow & \frac{(-4s^3-18s^2-26s-4)(s+4) + s^4 + 6s^3 + 13s^2 + 4s - 24}{(s+4)^2} = 0 \\ \Leftrightarrow & \frac{-3s^2 - 28s^3 - 85s^2 - 104s - 40}{(s+4)^2} = 0 \end{aligned} \quad (16.38)$$

This is too difficult to solve analytically, but we can make

```
>> roots([3 28 85 104 40])
```

```
ans =
```

```
-4.8449 + 0.0000i
-1.8950 + 0.5908i
-1.8950 - 0.5908i
-0.6985 + 0.0000i
```

Only the real roots matter, viz. -4.8449 and -0.6985 .

11. Let $p_1 = 1$, $p_2 = -2 + 2j$, $p_3 = -2 - 2j$, $p_4 = -3$ and $z_1 = -4$. See Figure 16.7. The angles of the asymptotes around p_2 are found as

$$\begin{aligned} & \begin{cases} \phi_1 = 180^\circ - \arctan \frac{2}{3} = 146.3^\circ \\ \phi_3 = 90^\circ \\ \phi_4 = \arctan \frac{2}{1} = 63.4^\circ \\ \psi_1 = \arctan \frac{2}{2} = 45^\circ \end{cases} \\ \Rightarrow & \begin{cases} \phi_2 = 45^\circ - 146.3^\circ - 90^\circ - 63.4^\circ = -254.7^\circ \equiv 105.3^\circ, K > 0 \\ \phi_2 = -254.7^\circ + 180^\circ = -76.7^\circ, K < 0 \end{cases} \end{aligned} \quad (16.39)$$

Because of the symmetry, we can tell right away that

$$\begin{cases} \phi_3 = -105.3^\circ, K > 0 \\ \phi_3 = 76.7^\circ, K < 0 \end{cases} \quad (16.40)$$

12. There are no complex conjugate open loop poles, so this rule does not apply. With all we know, the second diagram in Figure 16.6 can now be drawn.

It is clear from the root locus that

- $G(s)$ can be controlled in closed loop with a proportional controller K ;
- when $K > 0$, the control loop is always unstable;
- when $K < 0$ is close to 0, the control loop is also unstable;
- when $K < 0$ is very large, the control loop is unstable too;
- when the control loop is stable, there may be two complex conjugate dominant poles, resulting in an oscillating response, or one real dominant pole;
- even when there is a real dominant pole, there will be complex conjugate poles;

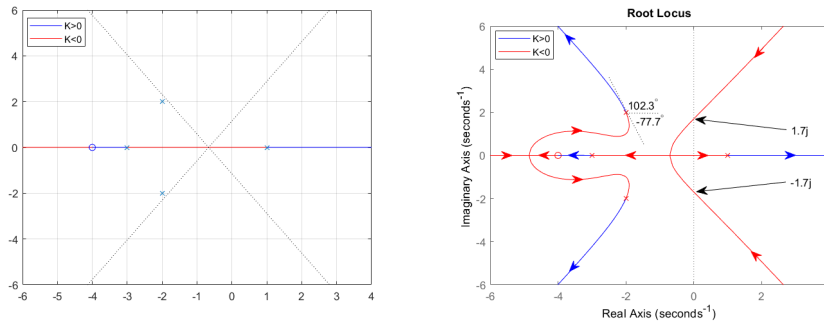


Figure 16.6: Root locus of (16.30) from Example 16.4. Left: incomplete sketch after step 9. Right: complete sketch after step 11.

- the control loop will be stable only for an interval of negative values of K , that can be found applying to the closed loop transfer function

$$\begin{aligned}
 \frac{y(s)}{r(s)} &= \frac{\frac{K(s+4)}{-s^4-6s^3-13s^2-4s+24}}{1 + \frac{K(s+4)}{-s^4-6s^3-13s^2-4s+24}} \\
 &= \frac{Ks + 4K}{-s^4 - 6s^3 - 13s^2 - 4s + 24 + Ks + 4K} \\
 &= \frac{-Ks - 4K}{s^4 + 6s^3 + 13s^2 + s(4 - K) + (-4K - 24)} \quad (16.41)
 \end{aligned}$$

(in which we took the care of letting the first coefficient of the denominator be positive) the Routh-Hurwitz criterion:

$$\begin{array}{c|ccc}
 s^4 & 1 & 13 & -4K - 24 \\
 s^3 & 6 & 4 - K & \\
 \hline
 s^2 & 13 + \frac{K-4}{6} & -4K - 24 & \\
 s & 4 - K + 6\frac{4K+24}{13 + \frac{K-4}{6}} & & \\
 1 & -4K - 24 & &
 \end{array} \quad (16.42)$$

For stability, the entire first column must have the same sign (in this case, positive), and so

$$\begin{aligned}
 &\begin{cases} 13 + \frac{K-4}{6} > 0 \\ 4 - K + 36\frac{4K+24}{78+K-4} > 0 \\ -4K - 24 > 0 \end{cases} \\
 \Rightarrow &\begin{cases} K - 4 > -78 \\ \frac{(-K+4)(K+74)+144K+864}{K+74} > 0 \\ 4K < -24 \end{cases} \\
 \Rightarrow &\begin{cases} K > -74 \\ \frac{K^2-74K-1160}{K+74} < 0 \\ K < -6 \end{cases} \\
 \Rightarrow &\begin{cases} K > -74 \\ -13.3 < K < 87.3 \\ K < -6 \end{cases} \Rightarrow -13.3 < K < -6 \quad (16.43)
 \end{aligned}$$

Notice that these negative values of K , replaced in the closed loop (16.41), result in a positive steady state value;

- when $K = -6$, the marginally stable closed loop has a pole at $s = 0$;
- when $K = -13.3$, the closed loop has two imaginary poles $s = \pm j\omega$. \square

Example 16.5. In the previous example, it may be interesting to know where exactly does the root locus cross the imaginary axis, i.e. the value of $\omega > 0$ for the closed loop poles $s = \pm j\omega$ when $K = -13.3$. This can be found from the

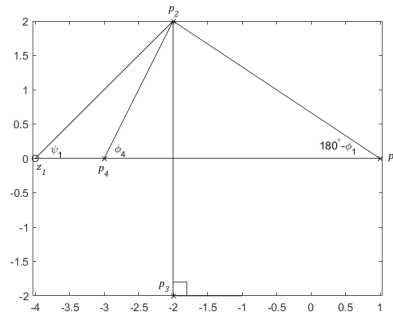


Figure 16.7: Angles needed for the root locus of (16.30) from Example 16.4.

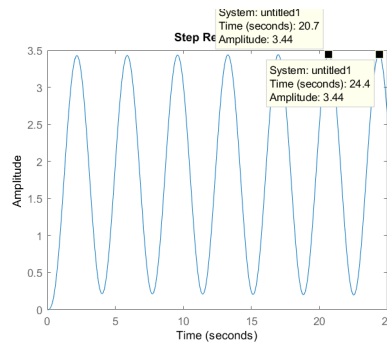


Figure 16.8: Step response of the marginally stable closed loop consisting in plant (16.30) controlled with $K = -13.3$, from Example 16.5.

denominator of the closed loop (16.41), which must be zero at these poles. If we replace $K = -13.3$ and $s = j\omega$, and equal to zero:

$$\begin{aligned}
 & (j\omega)^4 + 6(j\omega)^3 + 13(j\omega)^2 + j\omega(4 + 13.3) + (-4(-13.3) - 24) = 0 \\
 \Leftrightarrow & \omega^4 - j6\omega^3 - 13\omega^2 + j17.3\omega + 29.2 = 0 \\
 \Leftrightarrow & \begin{cases} \omega^4 - 13\omega^2 + 29.2 = 0 \\ -6\omega^3 + 17.3\omega = 0 \end{cases} \\
 \Leftrightarrow & \begin{cases} \omega^2 = \frac{13 \pm \sqrt{169 - 4 \times 29.2}}{2} \\ \omega = 0 \vee -6\omega^2 + 17.3 = 0 \end{cases} \\
 \Leftrightarrow & \begin{cases} \omega^2 = 10.1 \vee \omega^2 = 2.9 \\ \omega = 0 \vee \omega^2 = 2.9 \end{cases} \\
 \Leftrightarrow & \begin{cases} \omega = 3.2 \vee \omega = 1.7 \\ \omega = 0 \vee \omega = 1.7 \end{cases} \tag{16.44}
 \end{aligned}$$

Thus, when $K = -13.3$ the root locus crosses the imaginary axis at $s = \pm 1.7j$, as shown in Figure 16.6. The interest of knowing this is that, as we saw in Chapter 10, this ω is the frequency of an oscillation. The oscillation is that of the step response of the closed loop, which is marginally stable (since it has no unstable poles, and two poles on the imaginary axis). If we make

`figure,step(feedback(-13.3 * (s+4)/((-s+1)*(s+3)*(s^2+4*s+8)), 1), 25)`

we get the response in Figure 16.8. The frequency of the oscillation is, as expected,

$$\omega = \frac{2\pi}{24.4 - 20.7} = 1.7 \text{ rad/s} \quad \square \tag{16.45}$$

16.3 Proofs of rules for the root locus

We now establish the results we used in the last section and incidentally meet the important concept of characteristic equation.

Theorem 16.1. The root locus diagram is symmetric in relation to real axis.

Proof. This is a consequence of poles being either real or pairs of complex conjugates, as we saw in Section 9.1. \square

Definition 16.1. Given a plant with transfer function

Characteristic equation

$$G(s) = \frac{N(s)}{D(s)} \quad (16.46)$$

where $N(s)$ and $D(s)$ are polynomials in s , the equation $D(s) = 0$, the roots of which are the poles of $G(s)$, is called **characteristic equation**. \square

Corollary 16.1. Let the direct branch of any of the closed loop control systems in Figure 16.1 be $K \frac{N(s)}{D(s)}$, where K is the varying gain and polynomials $N(s)$ and $D(s)$ collect the zeros and the poles. Since the transfer function of the closed loop is

$$\frac{y(s)}{r(s)} = \frac{K \frac{N(s)}{D(s)}}{1 + K \frac{N(s)}{D(s)}} \quad (16.47)$$

the characteristic equation of the closed loop is

Closed loop characteristic equation

$$1 + K \frac{N(s)}{D(s)} = 0 \quad (16.48)$$

which can also be written in any of the following forms:

$$D(s) + KN(s) = 0 \quad (16.49)$$

$$K \frac{N(s)}{D(s)} = -1 \quad (16.50)$$

$$\begin{cases} |K| \frac{|N(s)|}{|D(s)|} = 1 \\ \angle \left[K \frac{N(s)}{D(s)} \right] = \pi + 2k\pi, \quad k \in \mathbb{Z} \end{cases} \quad \square \quad (16.51)$$

In what follows we always presume that the open loop transfer function $K \frac{N(s)}{D(s)}$ is proper, i.e. that the order n of polynomial $N(s)$ is not larger than the order m of polynomial $D(s)$, i.e. that there are not more zeros than poles (remember Definition 9.1 and the discussion about the number of poles and zeros in Section 11.4). Also let

$$\begin{aligned} N(s) &= b_m s^m + b_{m-1} s^{m-1} + \dots + b_0 \\ &= b_m (s - z_1) \dots (s - z_m) \end{aligned} \quad (16.52)$$

$$\begin{aligned} D(s) &= a_n s^n + a_{n-1} s^{n-1} + \dots + a_0 \\ &= a_n (s - p_1) \dots (s - p_n) \end{aligned} \quad (16.53)$$

where z_1, \dots, z_m are the zeros of the open loop, and p_1, \dots, p_n the poles.

Theorem 16.2. Points s on the real axis that are not zeros of $G(s)$ belong to the root locus

- if there is an odd number of open loop poles and zeros to its right, and $\frac{Kb_m}{a_n} > 0$;
- if there is an even number of open loop poles and zeros to its right, or no poles and zeros at all, and $\frac{Kb_m}{a_n} < 0$.

Proof. The phase condition of (16.51) can be written as

$$\begin{aligned} \angle \left[K \frac{b_m (s - z_1) \dots (s - z_m)}{a_n (s - p_1) \dots (s - p_n)} \right] &= \pi + 2k\pi \\ \Leftrightarrow \angle \left[\overbrace{\frac{Kb_m}{a_n}}^{\in \mathbb{R}} \right] + \angle[s - z_1] + \dots + \angle[s - z_m] - \angle[s - p_1] - \dots - \angle[s - p_n] &= \pi + 2k\pi \\ \Leftrightarrow \angle[s - z_1] + \dots + \angle[s - z_m] - \angle[s - p_1] - \dots - \angle[s - p_n] &= \underbrace{-\angle \left[\frac{Kb_m}{a_n} \right]}_{\text{either } 0 \text{ or } \pm\pi} + \pi + 2k\pi \end{aligned} \quad (16.54)$$

Let s be real. As we know, poles and zeros may either be real or appear as complex conjugates:

- When a zero z_i or pole p_i is real, the angle $\angle[s - z_i]$ or $\angle[s - p_i]$ is either 0 or $\pm\pi$.
 - It will be 0 if $s - z_i > 0 \Leftrightarrow s > z_i$ or $s - p_i > 0 \Leftrightarrow s > p_i$, i.e. if the pole or zero lies to the left of s .
 - It will be $\pm\pi$ if $s - z_i < 0 \Leftrightarrow s < z_i$ or $s - p_i < 0 \Leftrightarrow s < p_i$, i.e. if the pole or zero lies to the right of s . Notice that it is irrelevant whether we are talking about a pole or a zero, since $+\pi \equiv -\pi$.
- When there is a pair of complex conjugate poles or zeros, since s is real and thus $s = \bar{s}$, the angles must always cancel out:

$$\begin{aligned}\angle[s - z_i] + \angle[s - \bar{z}_i] &= \angle[s - z_i] + \angle[\overline{s - z_i}] \\ &= \angle[s - z_i] - \angle[s - z_i] = 0\end{aligned}\quad (16.55)$$

$$\begin{aligned}-\angle[s - p_i] - \angle[s - \bar{p}_i] &= -\angle[s - p_i] - \angle[\overline{s - p_i}] \\ &= -\angle[s - p_i] - (-\angle[s - p_i]) = 0\end{aligned}\quad (16.56)$$

(This is graphically represented in Figure 16.9, since each of the terms $s - z_1$ to $s - p_n$, which is a complex number, can also be thought of as a vector on the complex plane.) Thus, complex conjugate zeros and poles are irrelevant for (16.54), and each real zero or pole to the right of s contributes with a phase of $\pm\pi$. So:

- If $\frac{Kb_m}{a_n} > 0$, then $-\angle\left[\frac{Kb_m}{a_n}\right] = 0$, and (16.54) becomes

$$\angle[s - z_1] + \dots + \angle[s - z_m] - \angle[s - p_1] - \dots - \angle[s - p_n] = \pi + 2k\pi \quad (16.57)$$

There must be an odd number of real zeros and poles to the right of s , if s is to belong to the root locus.

- If $\frac{Kb_m}{a_n} < 0$, then $-\angle\left[\frac{Kb_m}{a_n}\right] = \pm\pi$, and (16.54) becomes

$$\angle[s - z_1] + \dots + \angle[s - z_m] - \angle[s - p_1] - \dots - \angle[s - p_n] = 2k\pi \quad (16.58)$$

There must be an even number of real zeros and poles to the right of s , if s is to belong to the root locus.

Because adding 2 to an odd number results in an odd number, and adding 2 to an even number results in an even number, the total number of zeros and poles can be considered; pairs of complex conjugates change nothing just the same.

We have shown when s may belong to the root locus because the phase condition of (16.51) is fulfilled. But, if the phase condition is fulfilled, the gain condition of (16.51) is fulfilled making

$$|K| = \frac{|D(s)|}{|N(s)|} \quad (16.59)$$

which is always possible as long as $N(s) \neq 0$. □

Theorem 16.3. The number of poles of the closed loop is the number of poles of the open loop n .

Proof. Since we assumed that $m \leq n$, and K is a scalar, the polynomial in the left member of (16.49) is of order n . □

Theorem 16.4. When $K \rightarrow 0$, the poles of the closed loop converge to the poles of the open loop, i.e. the roots of $D(s)$.

Proof. When $K \rightarrow 0$, (16.49) becomes $D(s) = 0$. □

Theorem 16.5. When $K \rightarrow \pm\infty$, m closed loop poles converge to the zeros of the open loop, i.e. the roots of $N(s)$, and $n - m$ closed loop poles diverge to infinity.

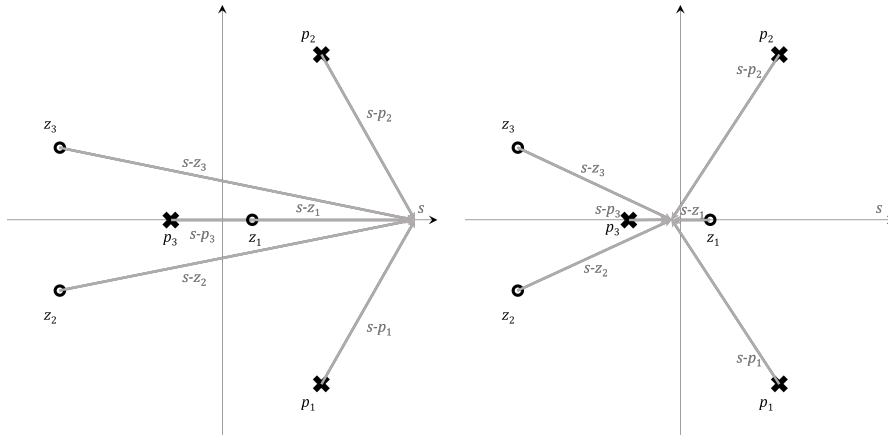


Figure 16.9: Diagrams to illustrate the proof of Theorem 16.2. Left: s has all the open loop zeros and poles to its left. Right: some open loop zeros and poles are on the right side of s , some on the left.

Proof. When $K \rightarrow \pm\infty$, (16.49) can be approximated by $N(s) = 0$. So the roots of $N(s)$, i.e. the zeros of the open loop, will be roots of the closed loop.

If $n = m$, this accounts for all the closed loop poles. If the open loop is strictly proper, there will be $n - m$ closed loop poles unaccounted for. In that case, the characteristic equation (16.49) becomes

$$\begin{aligned} a_n s^n + \dots + a_{m+1} s^{m+1} + a_m s^m + \dots + a_0 + K b_m s^m + \dots + K b_0 &= 0 \\ \Leftrightarrow a_n s^n + \dots + a_{m+1} s^{m+1} + (a_m + K b_m) s^m + \dots + a_0 + K b_0 &= 0 \\ &\Leftrightarrow a_n s^n + \dots + a_{m+1} s^{m+1} = -(a_m + K b_m) s^m - \dots - (a_0 + K b_0) \end{aligned} \quad (16.60)$$

Since $|K| \rightarrow +\infty$, the right member is diverging to infinity. The only way the left member is also diverging to infinity is $|s| \rightarrow +\infty$. (Notice that s is complex and may be diverging to infinity in any direction.) \square

Theorem 16.6. The $n - m$ closed loop poles diverging to infinity have asymptotes making with the positive real axis angles given by

$$\begin{cases} \frac{\pi(2k+1)}{n-m}, & \text{if } \frac{K b_m}{a_n} > 0 \\ \frac{2k\pi}{n-m}, & \text{if } \frac{K b_m}{a_n} < 0 \end{cases} \quad (16.61)$$

Proof. Since $|s| \leftarrow \infty$,

$$\begin{aligned} \lim_{|s| \rightarrow \infty} K \frac{N(s)}{D(s)} &= \lim_{|s| \rightarrow \infty} K \frac{b_m s^m + b_{m-1} s^{m-1} + \dots + b_0}{a_n s^n + a_{n-1} s^{n-1} + \dots + a_0} \\ &= \lim_{|s| \rightarrow \infty} \frac{K b_m}{a_n} \frac{(s - z_1) \dots (s - z_m)}{(s - p_1) \dots (s - p_n)} \\ &= \lim_{|s| \rightarrow \infty} \frac{K b_m}{a_n} \frac{s^m}{s^n} = \lim_{|s| \rightarrow \infty} \frac{K b_m}{a_n} \frac{1}{s^{n-m}} \end{aligned} \quad (16.62)$$

The phase condition of (16.51) becomes

$$\begin{aligned} \angle \left[\frac{K b_m}{a_n} \frac{1}{s^{n-m}} \right] &= \pi + 2k\pi, \quad k \in \mathbb{Z} \\ \Leftrightarrow \underbrace{\angle \left[\frac{K b_m}{a_n} \right]}_{\text{either } 0 \text{ or } \pi} - (n-m) \angle s &= \pi + 2k\pi, \quad k \in \mathbb{Z} \end{aligned} \quad (16.63)$$

If $\frac{K b_m}{a_n} > 0$, its contribution for the phase is 0, and we get

$$\angle s = -\frac{\pi + 2k\pi}{n-m} = \frac{\pi(2k+1)}{n-m}, \quad k \in \mathbb{Z} \quad (16.64)$$

If $\frac{Kb_m}{a_n} < 0$, its contribution for the phase is π , and we get instead

$$\angle s = -\frac{2k\pi}{n-m} = \frac{2k\pi}{n-m}, \quad k \in \mathbb{Z} \quad \square \quad (16.65)$$

□

Theorem 16.7. The branches of the root locus converge or diverge on the real axis at the real roots of

$$\frac{d}{ds} \frac{(s-p_1)\dots(s-p_n)}{(s-z_1)\dots(z-z_m)} = 0 \quad (16.66)$$

Proof. When the branches converge or diverge, the real point s where they do so will correspond to either a maximum or a minimum value of K on the real axis. Indeed, if we move along the real axis with K increasing, and reach a point of divergence s , then K will increase along branches that are no longer real, and to the other side of s gain K will now decrease. And, if we move along the real axis with K decreasing, and reach a point of convergence s , then K will decrease along branches that are no longer real, and to the other side of s gain K will now increase. This is clear by looking at any root locus diagram where such points exist.

Consequently, if K has a local maximum or minimum for a certain real point s , then $\frac{dK}{ds} = 0$. Using (16.50),

$$\begin{aligned} K &= -\frac{D(s)}{N(s)} \Rightarrow \\ \frac{dK}{ds} &= \frac{d}{ds} \left(-\frac{D(s)}{N(s)} \right) = \frac{d}{ds} \frac{D(s)}{N(s)} = 0 \quad \square \end{aligned} \quad (16.67)$$

Theorem 16.8. Rules 11 and 12 from Table 16.2 hold.

Proof. Let s be a point of the root locus in the vicinity of pole p_i , i.e. we make $s \rightarrow p_i$. We take the limit of the phase condition of (16.51)

$$\lim_{s \rightarrow p_i} \angle \left[\frac{K b_m}{a_n} \frac{(s-z_1)\dots(s-z_m)}{(s-p_1)\dots(s-p_n)} \right] = \pi + 2k\pi, \quad k \in \mathbb{Z} \quad (16.68)$$

which becomes (dropping the $2k\pi$ periodicity)

$$\begin{aligned} &\overbrace{\angle \left[\frac{K b_m}{a_n} \right]}^{\text{either } 0 \text{ or } \pm\pi} + \lim_{s \rightarrow p_i} \overbrace{\angle [s-z_1]}^{\angle [p_i-z_1]=\psi_1} + \dots + \lim_{s \rightarrow p_i} \overbrace{\angle [s-z_m]}^{\angle [p_i-z_m]=\psi_m} \\ &- \lim_{s \rightarrow p_i} \overbrace{\angle [s-p_1]}^{\angle [p_i-p_1]=\phi_1} - \dots - \lim_{s \rightarrow p_i} \overbrace{\angle [s-p_i]}^{\phi_i} - \dots - \lim_{s \rightarrow p_i} \overbrace{\angle [s-p_n]}^{\angle [p_n-p_1]=\phi_n} = 0 \end{aligned} \quad (16.69)$$

Solving for ϕ_i , we get (16.10). The proof for a zero is similar. □

We will not prove:

- expression (16.8) for the intersection of the asymptotes;
- that branches only converge or diverge on the real axis;
- that they converge or diverge perpendicularly to the real axis;
- that the centroid of the root locus is constant if $n-m \geq 2$.

Root locus when the feedback branch is not 1

Remark 16.1. Up until now we have assumed that the feedback branch of the closed loop is 1 (i.e. we have assumed a perfect sensor). If this is not the case, the closed loop in Figure 16.10

$$y = KG(r - Hy) \Rightarrow y(1 + KGH) = KGr \Rightarrow \frac{y}{r} = \frac{KG}{1 + KGH} \quad (16.70)$$

has the characteristic equation

$$1 + KGH = 0 \Leftrightarrow KGH = -1 \Leftrightarrow \begin{cases} |KGH| = 1 \\ \angle [KGH] = \pi + 2k\pi, \quad k \in \mathbb{Z} \end{cases} \quad (16.71)$$

Thus, in this situation, instead of the root locus of $G(s)$, the root locus of $G(s)H(s)$ has to be used instead. □

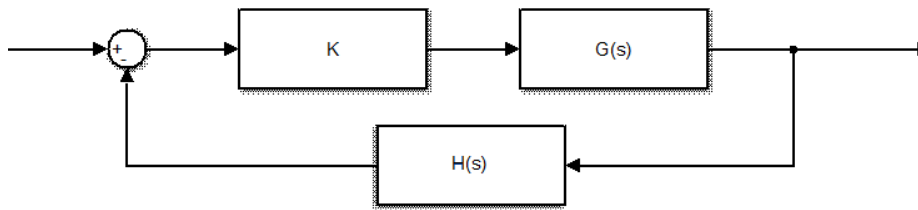


Figure 16.10: Closed loop with non-unitary feedback.

16.4 Finding desired poles from specifications

Suppose that we are given specifications of how the time responses of a controlled system should be. To use the root locus diagram to design a proportional controller, or to adjust the gain of any other type of controller, we

- find from the specifications the zones of the complex plane where the closed loop poles must lie;
- find for which values of gain the root locus is inside such zones. If there are no values of the gain that put all the closed poles in those zones, another controller is necessary.

We know from Section 11.6 that the poles are not the only thing that determines the time response of a system, but reasonable approximations can be found in this way if there is a dominant real pole $s = -a$, or a pair of complex conjugate dominant poles

$$\begin{aligned} s &= -a \pm jb \\ &= -\xi\omega_n \pm j\omega_n\sqrt{1-\xi^2} \end{aligned} \quad (16.72)$$

$$\Rightarrow \begin{cases} |s| = \sqrt{\xi^2\omega_n^2 + \omega_n^2(1-\xi^2)} = \omega_n \\ \angle s = \arctan \frac{\pm\omega_n\sqrt{1-\xi^2}}{-\xi\omega_n} = \arctan \frac{\pm\sqrt{1-\xi^2}}{-\xi} \end{cases} \quad (16.73)$$

(remember (11.43)–(11.44)).

Recall from Sections 11.2 and 11.3 that, in either case,

- the 10% settling time is $t_{s,10\%} = \frac{2.3}{a} \Leftrightarrow -a = -\frac{2.3}{t_{s,10\%}}$;
- the 5% settling time is $t_{s,5\%} = \frac{3}{a} \Leftrightarrow -a = -\frac{3}{t_{s,5\%}}$;
- the 2% settling time is $t_{s,2\%} = \frac{4}{a} \Leftrightarrow -a = -\frac{4}{t_{s,2\%}}$;
- the 1% settling time is $t_{s,1\%} = \frac{4.6}{a} \Leftrightarrow -a = -\frac{4.6}{t_{s,1\%}}$.

Zones of \mathbb{C} for a given settling time

Suppose that a maximum value for the settling time is required. This settling time will correspond to poles on a vertical straight line with real part $-a$ found from the relations above. The dominant poles can be on this line, or, better still, somewhat to its left, in which case they will be faster. See Figure 16.11.

Also recall that:

- if there is a dominant real pole (i.e. without imaginary part), there is no overshoot;
- if there is a pair of complex conjugate dominant poles, the maximum

Zones of \mathbb{C} for a maximum value of M_p

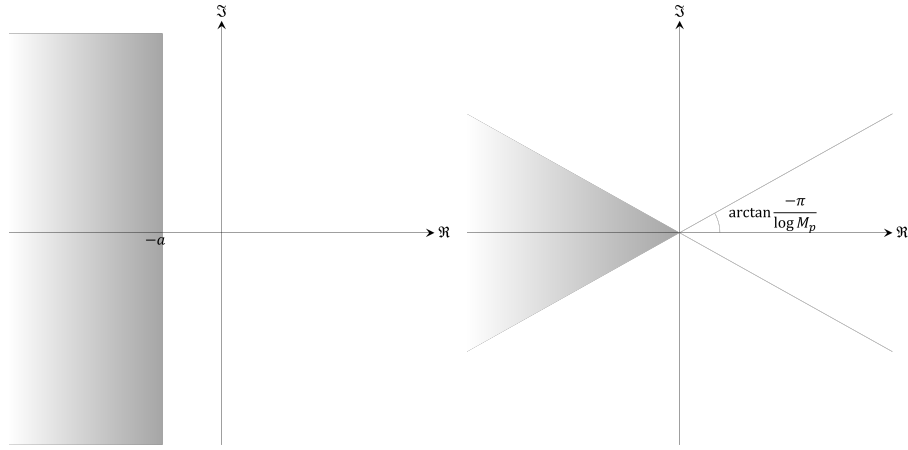


Figure 16.11: Left: zone of the complex plane where dominant poles must be, given a real part $-a$ found from a maximum settling time t_s specification. Right: zone of the complex plane where the dominant poles must be, given a maximum overshoot M_p that is not to be exceeded.

overshoot (as a fraction of the steady-state value) is given by (11.72):

$$\begin{aligned}
 M_p &= e^{\frac{-\xi\pi}{\sqrt{1-\xi^2}}} \\
 \Rightarrow \frac{\xi\pi}{\sqrt{1-\xi^2}} &= -\log M_p \\
 \Rightarrow \xi^2\pi^2 &= (\log M_p)^2(1-\xi^2) \\
 \Rightarrow \xi^2(\pi^2 + (\log M_p)^2) &= (\log M_p)^2 \\
 \Rightarrow \xi^2 &= \frac{(\log M_p)^2}{\pi^2 + (\log M_p)^2} \quad (16.74)
 \end{aligned}$$

Of the two possible values of the square root, one is positive and the other is negative. Usually the maximum overshoot is smaller than 100%, i.e. than 1; thus, it is with the minus sign that we get a positive value for the damping coefficient ξ :

$$\xi = \frac{-\log M_p}{\sqrt{\pi^2 + (\log M_p)^2}} \quad (16.75)$$

So, if one of the specifications is that a certain value of the maximum overshoot cannot be exceeded, (16.75) is used to find the minimum admissible damping factor (ξ can be higher, since larger damping factor correspond to more damped oscillations, i.e. with smaller amplitudes). From (16.73), we know that ξ has no effect on the magnitude of the poles, and suffices to find their phase. Replacing (16.74) and (16.75) in (16.73),

$$\begin{aligned}
 \angle s &= \arctan \frac{\pm \sqrt{1 - \frac{(\log M_p)^2}{\pi^2 + (\log M_p)^2}}}{\frac{-\log M_p}{\sqrt{\pi^2 + (\log M_p)^2}}} \\
 &= \arctan \frac{\pm \sqrt{\frac{\pi^2}{\pi^2 + (\log M_p)^2}}}{\frac{\log M_p}{\sqrt{\pi^2 + (\log M_p)^2}}} \\
 &= \arctan \pm \frac{\pi}{\log M_p} \quad (16.76)
 \end{aligned}$$

Since the poles are stable, these angles are in the second ($\frac{\pi}{2} \leq \angle s \leq \pi$) and third quadrants ($-\pi \leq \angle s \leq -\frac{\pi}{2}$). Poles closer to the real axis, where there are no oscillations, will have lower values of the maximum overshoot (the imaginary parts of the poles are smaller; remember from

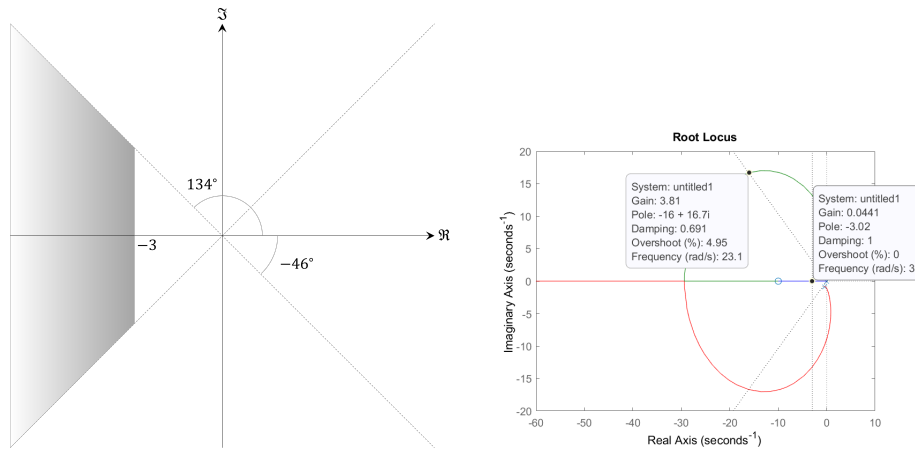


Figure 16.12: Left: zone of \mathbb{C} where dominant poles must be in Example 16.6. Right: root locus of $C(s)G(s)$.

Section 11.6 that the imaginary parts of the poles are what originate the oscillations). Thus, a specification for the largest admissible value of M_p translates into a sector of the complex plane around the negative real axis. See Figure 16.11.

Of course, if there is no clearly dominant pole or pair of complex conjugate poles, the approximations above for the desired locations of poles will be poor.

Example 16.6. Plant $G(s) = \frac{10}{s^2 + s + 1}$ will be controlled in closed loop by a PID controller given by $C(s) = K_p \left(1 + \frac{5}{s} + \frac{s}{20}\right)$. We want

- a maximum overshoot of 5% or less;
- a 5% settling time of 1 s or less.

What should be the value of the PID gain K_p ?

We can find the zone of \mathbb{C} where the dominant pole or poles should be as follows:

- From $M_p \leq 5\% = 0.05$, we get $|\angle s| \geq \left| \arctan \frac{\pi}{\log 0.05} \right| = 134^\circ$. Notice that we chose the angle in the second quadrant; calculating $\arctan \frac{\pi}{\log 0.05}$ will likely gives us -46° , in the fourth quadrant, as the result.
- From $t_{s,5\%} \leq 1$ s, we get $-a \leq -3$.

This zone is shown in Figure 16.12. We now plot the root locus of $C(s)G(s)$ and check the values of K_p for which all poles are inside the desired zone: $K_p = 3.8$, in this case. Choosing the lowest possible of K_p , the control action will be as low as possible. But it may be reasonable to allow for some uncertainty in parameters, and thus use a slightly larger value of K_p to ensure that all poles are indeed inside the desired zone.

The step responses of the closed loop for $K_p = 5$ and $K_p = 30$ are shown in Figure 16.13. Notice how, for $K_p = 5$, the specifications are still not followed, even though all the poles, and consequently the dominant one or dominant ones, are inside the zone of the complex plane corresponding to the specifications. If the control action is feasible, the problem is solved using $C(s) = 30 \left(1 + \frac{5}{s} + \frac{s}{20}\right)$. \square

Remark 16.2. In the Example above, as we do not know what type of system $G(s)$ is (mechanical, thermal, etc.), and what actuator is being used, we can form no idea about whether the control action is reasonably small, or too large. We also do not know if the actuator bandwidth required to implement this control action is feasible. In real life, it is important to know what system is being controlled, so that these matters can be decided. \square

MATLAB includes the app `controlSystemDesigner` (called `sisotool` in older versions) to assist the design of closed loop controllers for SISO plants. Among other functionalities, this app lets us select a plant and change the open loop gain in its root locus diagram, showing interactively how the closed loop step response will be in each case. See Figure 16.14.

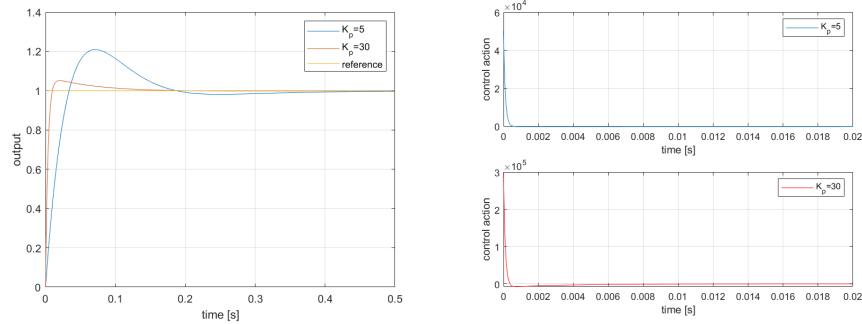


Figure 16.13: Behaviour of the closed loop control system from Example 16.6.

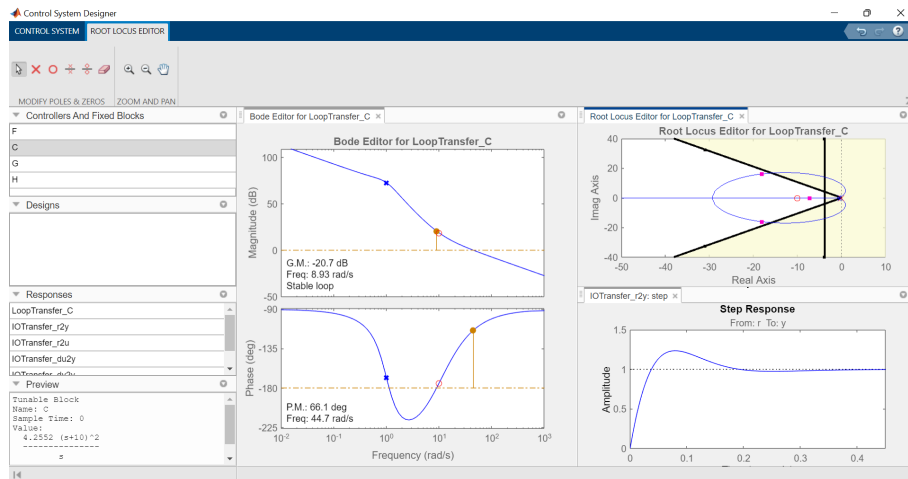


Figure 16.14: controlSystemDesigner from MATLAB.

Glossary

And Eliacim, and Sobna, and Ioahē sayd to Rabsaces: Speake to thy seruants in the Syrian tongue: for we vnderstand it: speake not to vs in the Iewes language in the eares of the people, that is vpon the wal. And Rabsaces sayd to them: Why, did my lord send me to thy lord and to thee, to speake al these wordes; and not rather to the men, that sitte on the wal (...)?

ISAIAH ben Amoz (attrib.; 8th–7th c. BC), *Isaiah*, xxxvi 11–12, Douay-Rheims version (1610)

branch ramo

characteristic equation equação característica

root locus lugar geométrico das raízes (LGR)

Exercises

1. Consider the plants with the following open-loop transfer functions. Admit both cases $K > 0$ and $K < 0$. Plot their root-locus, determining all the relevant points, the asymptotes, the departure and arrival angles, the values of K ensuring stability (resorting to the Routh-Hurwitz criterion as needed), and the points where the root locus crosses the imaginary axis.

$$(a) G_1(s) = \frac{1}{s+10}$$

$$(b) G_2(s) = \frac{1}{s-10}$$

$$(c) G_3(s) = \frac{s+30}{s+10}$$

$$(d) G_4(s) = \frac{s+10}{s+30}$$

$$\begin{aligned} \text{(e)} \quad G_5(s) &= \frac{1}{(s+10)(s+30)} \\ \text{(f)} \quad G_6(s) &= \frac{1}{(s+10)(s-30)} \\ \text{(g)} \quad G_7(s) &= \frac{1}{(s-10)(s+30)} \\ \text{(h)} \quad G_8(s) &= \frac{1}{(s-10)(s-30)} \\ \text{(i)} \quad G_9(s) &= \frac{s+20}{(s+10)(s+30)} \\ \text{(j)} \quad G_{10}(s) &= \frac{1}{(s+10)(s+20)(s+30)} \\ \text{(k)} \quad G_{11}(s) &= \frac{1}{s^2+20s+200} \\ \text{(l)} \quad G_{12}(s) &= \frac{s+20}{s^2+20s+200} \end{aligned}$$

2. Consider the plants with the following open-loop transfer functions. Admit both cases $K > 0$ and $K < 0$. Plot their root-locus, determining all the relevant points, the asymptotes, the departure and arrival angles, the values of K ensuring stability (resorting to the Routh-Hurwitz criterion as needed), and the points where the root locus crosses the imaginary axis.

$$\begin{aligned} \text{(a)} \quad G_1(s) &= \frac{(s+1)}{s^2(s+2)} \\ \text{(b)} \quad G_2(s) &= \frac{(s-2)}{s^2(s+1)(s+3)} \\ \text{(c)} \quad G_3(s) &= \frac{1}{s(s^2+0.2s+1)} \\ \text{(d)} \quad G_4(s) &= \frac{(s+1)}{s(s^2+3s+9)} \\ \text{(e)} \quad G_5(s) &= \frac{1}{s(s+1)^3} \\ \text{(f)} \quad G_6(s) &= \frac{(s^2+2s+4)}{s(s+3)(s^2+2s+4)} \\ \text{(g)} \quad G_7(s) &= \frac{(s+1)}{s^2(s+3)(s+4)} \\ \text{(h)} \quad G_8(s) &= \frac{(s+1)}{s^2(s+10)} \\ \text{(i)} \quad G_9(s) &= \frac{(s+2)}{s(s^2+2s+2)} \\ \text{(j)} \quad G_{10}(s) &= \frac{(-10s+40)}{s(s^2+40s+1025)} \end{aligned}$$

3. Find the zones of the complex plane corresponding to the following specifications:

- Settling time under 5 min; no overshoot.
- Settling time under 1 min; overshoot under 5%.
- Settling time under 10 ms; overshoot under 25%.

4. Find a proportional controller for the following plants and specifications, whenever possible:

$$\begin{aligned} \text{(a)} \quad G(s) &= \frac{s+100}{s^2+100s+2600}, \text{ settling time under 50 s, overshoot under 1\%.} \\ \text{(b)} \quad G(s) &= \frac{s+100}{s^2+100s+2600}, \text{ settling time under 5 s, overshoot under 15\%.} \end{aligned}$$

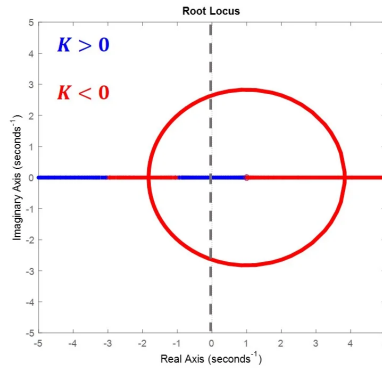


Figure 16.15: Root locus from Exercise 5.

- (c) $G(s) = \frac{100}{s^2 + 100s + 2600}$, settling time under 50 s, overshoot under 1%.
5. The plant with the root locus in Figure 16.15 will be controlled in closed loop with a proportional controller K .
- Is the loop always stable when $K > 0$?
 - Is the loop always unstable when $K > 0$?
 - Is the loop always stable when $K < 0$?
 - Is the loop always unstable when $K < 0$?
 - Is the loop stable if $|K|$ is small enough?
 - Is there any value of K for which the loop will be critically damped?

Chapter 17

The Nyquist stability criterion

‘I suppose there are two views about everything,’ said Mark.
‘Eh? Two views? There are a dozen views about everything until you know the answer. Then there’s never more than one.’

C. S. LEWIS (1868 — †1963), *That hideous strength* (1945), 3

The Nyquist criterion is another tool that, just like the root locus, can be used

- to design a proportional controller,
- to design the gain of a controller that has poles and zeros, once these are known,
- to see what happens to a closed loop control system with a known controller, when the gain of the plant changes for some reason.

This criterion is better visualised in the Nyquist diagram, and to study the Nyquist diagram it is convenient to study the polar diagram first.

17.1 The polar diagram

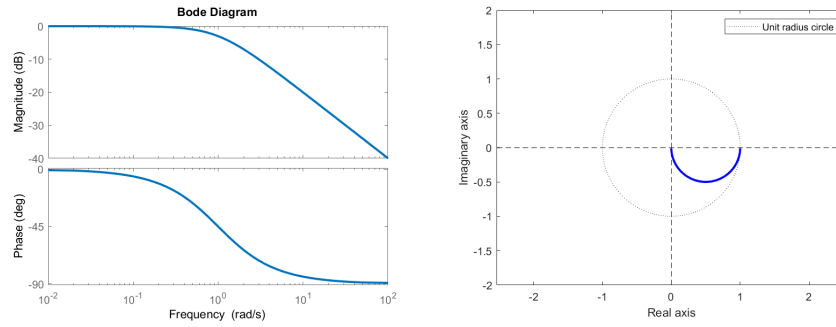
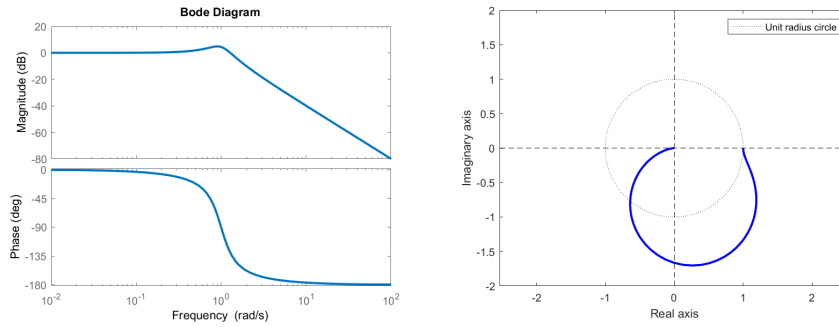
In the **Bode diagram**, the *frequency response* of a plant $G(s)$ is represented as a function of frequency $\omega > 0$; the *gain* $|G(j\omega)|$, in decibel, and the phase $\angle G(j\omega)$, are plotted separately (with a semilogarithmic scale). (Remember section 10.5.)

In the **polar diagram**, $G(j\omega)$ is represented in the complex plane; i.e. *Polar diagram shows* $\Im[G(j\omega)]$ is plotted in the y -axis as function of $\Re[G(j\omega)]$ in the x -axis. *$G(j\omega)$ in the complex plane* Frequency is not explicitly shown. It is obvious that, given any point of a polar diagram,

- the **gain** $|G(j\omega)|$ (in absolute value, not in dB) is its distance to the origin,
- the **phase** $\angle G(j\omega)$ is its phase.

Example 17.1. The polar diagram of $G(s) = \frac{1}{s+1}$ is shown in Figure 17.1 together with its Bode diagram. Notice that

- for low frequencies, $\lim_{\omega \rightarrow 0^+} G(j\omega) = 1$, and so the polar diagram begins at 1; *Where the polar plot begins*
- for high frequencies, $\lim_{\omega \rightarrow +\infty} G(j\omega) = 0$, and so the polar diagram ends at the origin; *Where the polar plot ends*
- for high frequencies, $\lim_{\omega \rightarrow +\infty} \angle G(j\omega) = -90^\circ$, and so the polar diagram approaches the origin from below, tangent to the imaginary axis, where the phase is -90° , remaining always to the right of the imaginary axis because the phase is never below -90° . \square

Figure 17.1: Bode diagram and polar plot for $G(s) = \frac{1}{s+1}$.Figure 17.2: Bode diagram and polar plot for $G(s) = \frac{1}{s^2+0.6s+1}$.

Example 17.2. The polar diagram of $G(s) = \frac{1}{s^2+0.6s+1}$ is shown in Figure 17.2 together with its Bode diagram. Notice that

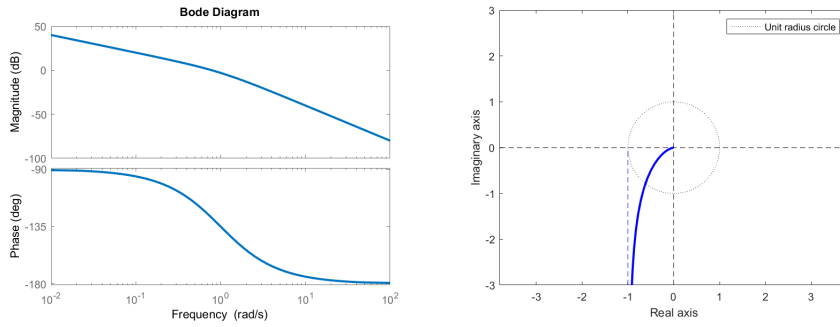
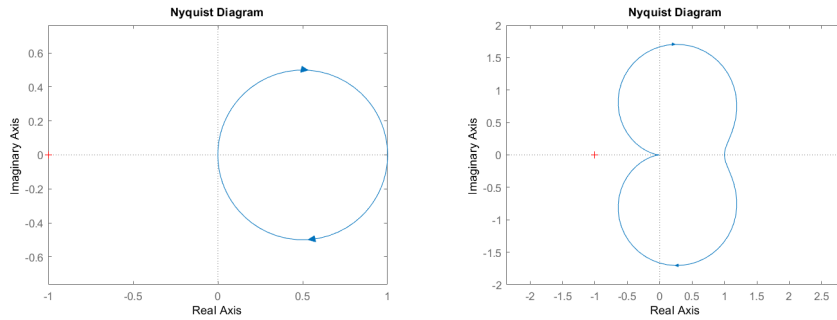
- for low frequencies, $\lim_{\omega \rightarrow 0^+} G(j\omega) = 1$, and so the polar diagram begins at 1;
- for high frequencies, $\lim_{\omega \rightarrow +\infty} G(j\omega) = 0$, and so the polar diagram ends at the origin;
- because $\xi = 0.3 < \frac{\sqrt{2}}{2}$, there is a resonance peak, and thus the gain is partly outside a unit radius circle;
- for high frequencies, $\lim_{\omega \rightarrow +\infty} \angle G(j\omega) = -180^\circ$, and so the polar diagram approaches the origin from the left, tangent to the real axis, where the phase is -180° , remaining always below the real axis because the phase is never below -180° . \square

Example 17.3. The polar diagram of $G(s) = \frac{1}{s(s+1)}$ is shown in Figure 17.3 together with its Bode diagram. The behaviour for low frequencies can be explained as follows.

- Because $\lim_{\omega \rightarrow 0^+} |G(j\omega)| = +\infty$, the polar diagram begins at an infinite distance from the origin.
- Because $\lim_{\omega \rightarrow 0^+} \angle G(j\omega) = -90^\circ$, the polar diagram begins vertically, from below. These points are in the third quadrant, and have phases between -180° and -90° . For arbitrarily large distances from the origin, the real part is neglectable, and the imaginary part very large; thus, the phases approach -90° .

Finding an asymptote of the polar plot

- This can be seen more accurately calculating


 Figure 17.3: Bode diagram and polar plot for $G(s) = \frac{1}{s(s+1)}$.

 Figure 17.4: Left: Nyquist diagram of $\frac{1}{s+1}$. Right: Nyquist diagram of $\frac{1}{s^2+0.6s+1}$.

$$\begin{aligned}
 G(j\omega) &= \frac{1}{j\omega(j\omega + 1)} \\
 &= \frac{1}{-\omega^2 + j\omega} \\
 &= \frac{-\omega^2 - j\omega}{(-\omega^2 + j\omega)(-\omega^2 - j\omega)} \\
 &= \frac{-\omega^2 - j\omega}{\omega^4 + \omega^2} = \frac{-1}{\omega^2 + 1} + j\frac{-1}{\omega^3 + \omega} \quad (17.1)
 \end{aligned}$$

$$\lim_{\omega \rightarrow 0^+} \Re[G(j\omega)] = \lim_{\omega \rightarrow 0^+} \frac{-1}{\omega^2 + 1} = -1 \quad (17.2)$$

$$\lim_{\omega \rightarrow 0^+} \Im[G(j\omega)] = \lim_{\omega \rightarrow 0^+} \frac{-j}{\omega^3 + \omega} = -j\infty \quad (17.3)$$

So (17.2)–(17.3) show that the diagram begins with an asymptote given by $\Re[z] = -1$, approaching the origin from below. The angle of the asymptote is, as expected, -90° . \square

17.2 The Nyquist diagram when there are no poles on the imaginary axis

The simplest description of the Nyquist diagram is that it consists in a transfer function's polar diagram plus its complex conjugate (i.e. the polar diagram turned upside down, so to say).

Example 17.4. Figure 17.4 shows the Nyquist diagram of $\frac{1}{s+1}$ and $\frac{1}{s^2+0.6s+1}$, as plotted by MATLAB function `nyquist` (and after command `axis equal` to make the scales of the x -axis and the y -axis the same). Compare this with Figures 17.1 and 17.2. \square

The formal definition is more complicated, but necessary for the Nyquist stability criterion we are about to study. We will postpone the case of plants with poles on the imaginary axis (either at 0, or pairs of complex conjugates) to Section 17.4.

The Nyquist diagram also shows the complex conjugate of the frequency response
 MATLAB function `nyquist`
 MATLAB command `axis equal`

First notice that the curve showing the frequency response in the polar diagram can be seen as the positive imaginary axis $j \times]0, +\infty[$ transformed by the complex-valued function of complex value $G(s)$. We say that the frequency response curve is a **mapping** of the positive imaginary axis, using function $G(s)$.

Mapping

$$G(-j\omega) = \overline{G(j\omega)}$$

Theorem 17.1. The complex conjugate of the frequency response $\overline{G(j\omega)}$ is the mapping of the negative imaginary axis, $G(-j\omega)$, $\omega > 0$.

Proof. Let $G(s) = \frac{N(s)}{D(s)}$. Then

$$\begin{aligned} G(j\omega) &= \frac{N(j\omega)}{D(j\omega)} \\ &= \frac{N(j\omega) \overline{D(j\omega)}}{\underbrace{D(j\omega) \overline{D(j\omega)}}_{\in \mathbb{R}}} \end{aligned} \quad (17.4)$$

The numerator $N(j\omega) \overline{D(j\omega)}$ is a polynomial on ω :

- Its real part will consist in
 - the independent term,
 - the term on $(j\omega)^2 = -\omega^2$,
 - the term on $(j\omega)^4 = \omega^4$,
 - the term on $(j\omega)^6 = -\omega^6$,

$\Re[G(j\omega)]$ is even

and so on, i.e. the terms on even powers of $j\omega$. Thus, $\Re[N(j\omega) \overline{D(j\omega)}]$ is an even function of ω , and so is $\Re[G(j\omega)]$. That is to say, $\Re[G(j\omega)] = \Re[G(-j\omega)]$.

- Its imaginary part will consist in
 - the term on $j\omega$,
 - the term on $(j\omega)^3 = -j\omega^3$,
 - the term on $(j\omega)^5 = j\omega^5$,

$\Im[G(j\omega)]$ is odd

and so on, i.e. the terms on odd powers of $j\omega$. Thus, $\Im[N(j\omega) \overline{D(j\omega)}]$ is an odd function of ω , and so is $\Im[G(j\omega)]$. That is to say, $\Im[G(j\omega)] = -\Im[G(-j\omega)]$.

Consequently,

$$\begin{aligned} \overline{G(j\omega)} &= \overline{\Re[G(j\omega)] + j\Im[G(j\omega)]} \\ &= \Re[G(j\omega)] - j\Im[G(j\omega)] \\ &= \Re[G(-j\omega)] + j\Im[G(-j\omega)] = G(-j\omega) \quad \square \end{aligned} \quad (17.5)$$

As a result, given a transfer function $G(s)$, its polar plot plus its complex conjugate are the mapping, using $G(s)$, of $j] - \infty, 0[\cup]0, +\infty[$. In fact we can throw in the origin too, and map the entire imaginary axis $j\omega$, $\omega \in] - \infty, +\infty[$, since we are assuming no poles on the imaginary axis: this means that $D(j\omega) \neq 0$ whatever the value of ω (including $\omega = 0$), and so $G(0) = \frac{N(0)}{D(0)}$ is finite.

The Nyquist diagram is defined as the mapping, not of the imaginary axis alone, but of a **contour** in \mathbb{C} .

Contour in \mathbb{C}

Definition 17.1. A contour is a closed curve in \mathbb{C} , oriented clockwise or counter-clockwise, such that:

- there is only a finite number of points of the curve where it is not differentiable;
- the curve can be completely traversed without passing twice by any point (this precludes closed curves that cross themselves, for instance). \square

Definition 17.2. The Nyquist diagram of a transfer function $G(s)$ that has no poles on the imaginary axis is the mapping, using $G(s)$, of the **Nyquist contour**, or Nyquist path, that consists of

Nyquist contour

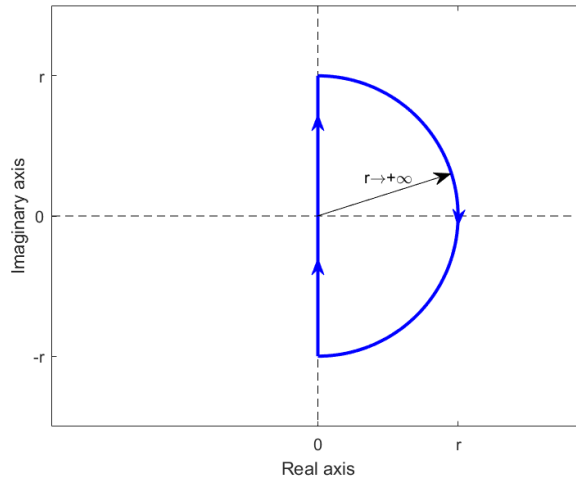


Figure 17.5: The Nyquist contour (blue) in the complex plane. This contour is mapped in a Nyquist diagram, when $G(s)$ has no poles on the imaginary axis.

- a straight line $s = j\omega$, $\omega \in [-r, +r]$, and
- a semi-circle $s = re^{j\theta}$, $\theta \in [-\frac{\pi}{2}, +\frac{\pi}{2}]$,

as $r \rightarrow +\infty$, and is clockwise oriented. See Figure 17.5. \square

Remark 17.1. As $r \rightarrow +\infty$, the semi-circle approaches infinity.

If $G(s)$ is strictly proper, this semi-circle of infinite radius is mapped to the origin. In fact,

$$\lim_{|s| \rightarrow +\infty} G(s) = \lim_{|s| \rightarrow +\infty} \frac{b_m s^m + \dots + b_0}{a_n s^n + \dots + a_0} = 0 \quad (17.6)$$

since $n > m$.

If $n = m$, the semi-circle of infinite radius is mapped to a point other than the origin. In fact,

$$\lim_{|s| \rightarrow +\infty} G(s) = \lim_{|s| \rightarrow +\infty} \frac{b_n s^n + \dots + b_0}{a_n s^n + \dots + a_0} = \frac{b_n}{a_n} \quad \square \quad (17.7)$$

17.3 The Nyquist criterion

The Nyquist stability criterion is based upon a theorem that we will not prove.

Theorem 17.2. Let $f(s) = \frac{f_N(s)}{f_D(s)}$ be a complex-valued rational function of complex variable s , and \mathcal{C} a clockwise (counter-clockwise) contour in \mathbb{C} that

- does not pass on the poles of $f(s)$,
- encircles P poles of $f(s)$,
- encircles Z zeros of $f(s)$.

Then the mapping of contour \mathcal{C} by $f(s)$ is a closed line $f(\mathcal{C})$ that encircles $Z - P$ times the origin, clockwise (counter-clockwise). \square

Remark 17.2. The mapped contour $f(\mathcal{C})$ need not be a contour. It may, for instance, cross itself, or even have coincident lines. \square

Example 17.5. We can verify this theorem numerically as shown in Figures 17.6 and 17.7. On the left, a counter-clockwise complex contour is shown, together with the zeros and poles of a transfer function, which will serve as $f(s)$. On the right, the contour mapped by the the transfer function is shown.

- In Figure 17.6, top, the contour encircles $Z = 0$ zeros and $P = 0$ poles of the transfer function. Thus, the mapped contour encircles $Z - P = 0$ times the origin, counter-clockwise.

- In Figure 17.6, centre, the contour encircles $Z = 0$ zeros and $P = 1$ poles of the transfer function. Thus, the mapped contour encircles $Z - P = -1$ times the origin, counter-clockwise, i.e. it encircles the origin once, clockwise.
- In Figure 17.6, bottom, the contour encircles $Z = 0$ zeros and $P = 2$ poles of the transfer function. Thus, the mapped contour encircles $Z - P = -2$ times the origin, counter-clockwise, i.e. it encircles the origin twice, clockwise.
- In Figure 17.7, top, the contour encircles $Z = 1$ zeros and $P = 2$ poles of the transfer function. Thus, the mapped contour encircles $Z - P = -1$ times the origin, counter-clockwise, i.e. it encircles the origin once, clockwise.
- In Figure 17.7, centre, the contour encircles $Z = 2$ zeros and $P = 4$ poles of the transfer function. Thus, the mapped contour encircles $Z - P = -2$ times the origin, counter-clockwise, i.e. it encircles the origin twice, clockwise.
- In Figure 17.7, bottom, a different contour is used with the same transfer function as above, showing that changing the shape of the contour is irrelevant for the result. The number of encirclements, is, as expected, the same: twice, clockwise.

The plots of Figure 17.6, top, can be drawn with the following code. The next cases are similar.

```

s1 = 1+1i*(-1:0.01:1);
s2 = 1i+(1:-0.01:-1);
s3 = -1+1i*(1:-0.0025:-1);
s4 = -1i+(-1:0.01:1);
G = @(s) 1./(s+2);
% transfer function defined as a function handle for complex variables
figure, plot(real(s1),imag(s1), real(s2),imag(s2),...
    real(s3),imag(s3), real(s4),imag(s4), -2,0,'x')
axis([-2 2 -2 2]), xlabel('Real axis'), ylabel('Imaginary axis')
title('1/(s+2)')
figure, plot(real(G(s1)),imag(G(s1)), real(G(s2)),imag(G(s2)),...
    real(G(s3)),imag(G(s3)), real(G(s4)),imag(G(s4)), 0,0,'+')
axis([-0.5 1.5 -1 1]), xlabel('Real axis'), ylabel('Imaginary axis')
title('1/(s+2)')

```

The plots of Figure 17.7, bottom, use the contour

```
sr = 1.5*exp(1i*(-pi:pi/100:pi));
```

□

Nyquist criterion

Theorem 17.3. Let $G(s) = \frac{N(s)}{D(s)}$ be a transfer function with P unstable poles. Let N be the number of clockwise encirclements of its Nyquist diagram around point -1 . Then $G(s)$ in closed loop, as seen in Figure 17.8, will have $Z = N + P$ unstable poles.

Proof. The characteristic equation of the closed loop is $1 + G(s) = 0$. We will apply Theorem 17.2 to function $f(s) = 1 + G(s)$. The contour \mathcal{C} will be the Nyquist path. We know that the mapping of the Nyquist path using $f(s)$ will have a number of enrolments around the origin N given by $N = Z - P \Leftrightarrow Z = N + P$. But:

- Since $G(s)$ is linear, $f(s)$ is also linear. So, given a contour \mathcal{C} , its mapping by $f(s)$ will be $f(\mathcal{C}) = 1 + G(\mathcal{C})$, i.e. the mapping by $G(s)$ shifted to the right by 1. Thus, the number of enrolments of $f(\mathcal{C})$ around the origin N is also the number of enrolments of $G(\mathcal{C})$ around -1 . And the mapping $G(\mathcal{C})$ is the Nyquist plot of $G(s)$.

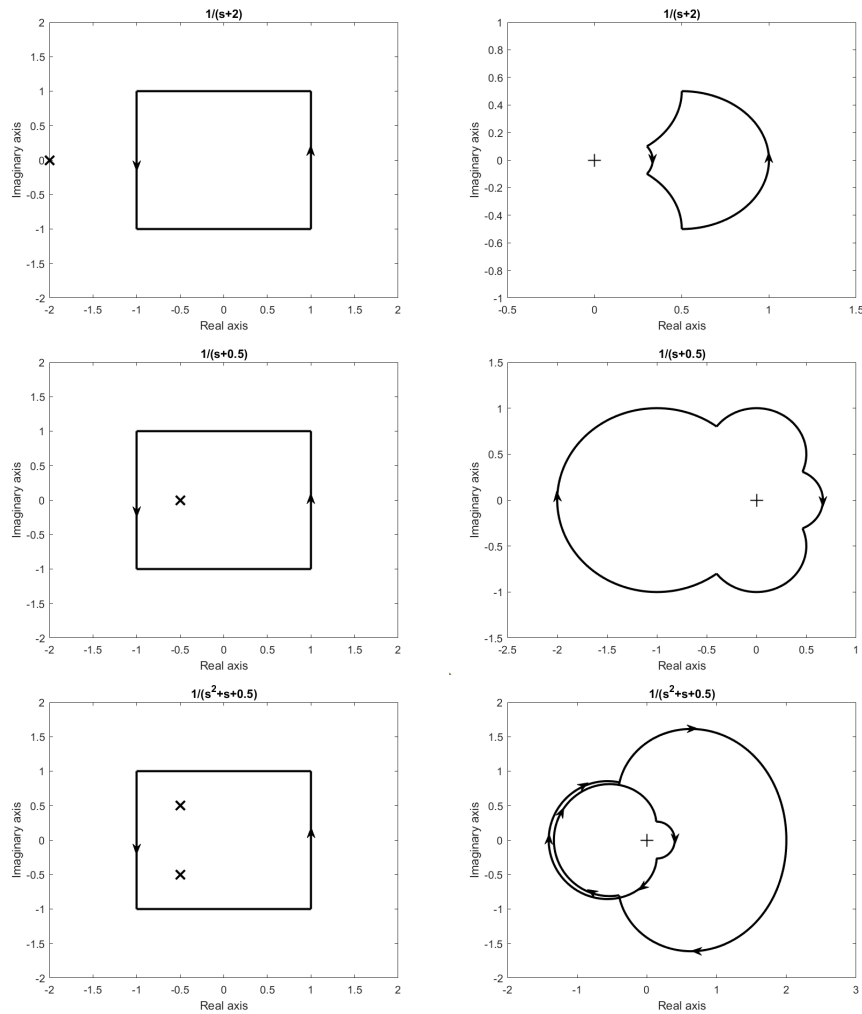


Figure 17.6: Example 17.5 illustrates Theorem 17.2. Left: complex contour and the zeros and poles of a transfer function. Right: the contour mapped by the the transfer function (the origin is marked with a cross). Bottom right figure: the two leftmost curves are coincident; the outside curve was enlarged to show more clearly the shape of the mapped contour. Continues in Figure 17.7.

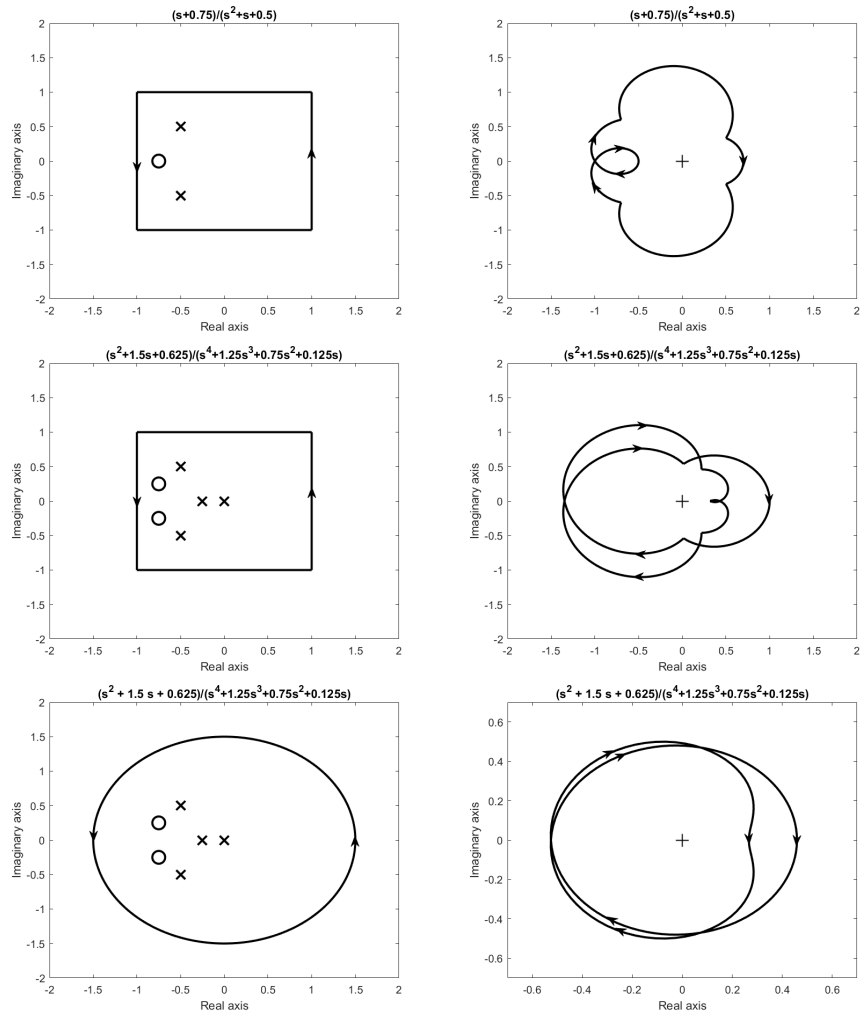


Figure 17.7: Figure 17.6 continued. Example 17.5 illustrates Theorem 17.2. Left: complex contour and the zeros and poles of a transfer function. Right: the contour mapped by the the transfer function (the origin is marked with a cross).

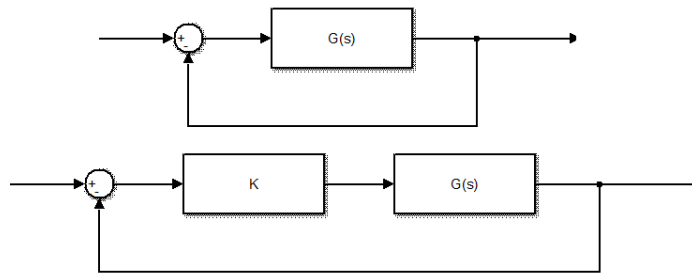


Figure 17.8: Top: block diagram for the Nyquist criterion. Bottom: block diagram for the Nyquist criterion, with a variable open loop gain K .

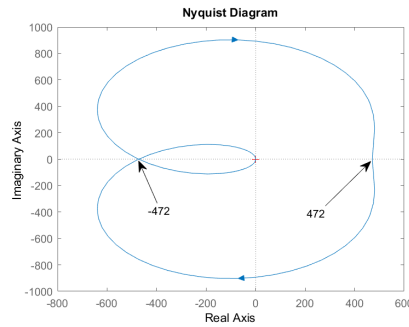


Figure 17.9: Nyquist diagram of model (17.8) of a boost converter, from Example 17.6.

- Z is the number of zeros of $f(s)$ inside \mathcal{C} . Since the Nyquist path covers the entire right half of the complex plane, Z is the number of roots of $f(s)$ such that $s > 0$. The roots of $f(s)$ are the roots of the closed loop characteristic equation, i.e. the unstable poles of the closed loop.
- P is the number of poles of $f(s) = 1 + G(s) = 1 + \frac{N(s)}{D(s)} = \frac{N(s)+D(s)}{D(s)}$ inside \mathcal{C} . Since the Nyquist path covers the entire right half of the complex plane, the poles of $f(s)$ are the roots of $D(s)$ such that $s > 0$, i.e. the unstable poles of $G(s)$.

□

Theorem 17.4. Let $G(s) = \frac{N(s)}{D(s)}$ be a transfer function with P unstable poles. Let N be the number of clockwise encirclements of its Nyquist diagram around point $-\frac{1}{K}$. Then $G(s)$ controlled in closed loop by proportional controller K , as seen in Figure 17.8, will have $Z = N + P$ unstable poles. *Nyquist criterion when K varies*

Proof. The characteristic equation of the closed loop is $1 + KG(s) = 0$. We will apply Theorem 17.2 to function $f(s) = 1 + KG(s)$. The proof is similar to that of the previous theorem, with the following change: given a contour \mathcal{C} , its mapping by $f(s)$ will be $f(\mathcal{C}) = 1 + KG(\mathcal{C})$, i.e. the mapping by $G(s)$ shifted to the right by $\frac{1}{K}$. Thus, the number of enrolments of $f(\mathcal{C})$ around the origin N is also the number of enrolments of $G(\mathcal{C})$ around $-\frac{1}{K}$. □

This result lets us use the Nyquist diagram of a transfer function to find the values of a proportional controller K that can stabilise it in closed loop, just as we might do using Routh’s criterion.

Example 17.6. A boost converter (a device that converts a DC voltage into a larger DC voltage) is modelled by transfer function

$$C(s) = \frac{-418040s + 9.523 \times 10^8}{s^2 + 884.6s + 2.015 \times 10^6} \tag{17.8}$$

The corresponding Nyquist diagram is shown in Figure 17.9. What values of a proportional controller K can stabilise this transfer function in closed loop?

The poles of $C(s)$ are $-442.3 \pm 1348.8j$. They are both stable, thus $P = 0$. We count the number of clockwise enrolments around the points on the real axis, and build the following table:

\mathbb{R}	$-\infty$	-472	0	472	$+\infty$
$-\frac{1}{K} \in$	$]-\infty, -472[$	$]-472, 0[$	$]0, 472[$	$]472, +\infty[$	
$K \in$	$]0, 0.0021[$	$]0.0021, +\infty[$	$]-\infty, -0.0021[$	$]-0.0021, 0[$	
N	0	2	1	0	
P	0	0	0	0	
Z	0	2	1	0	

(17.9)

From the table it is clear that there will be no closed loop unstable poles ($Z = 0$) if $-0.0021 < K < 0$ or if $0 < K < 0.0021$. So the open loop will be stable for

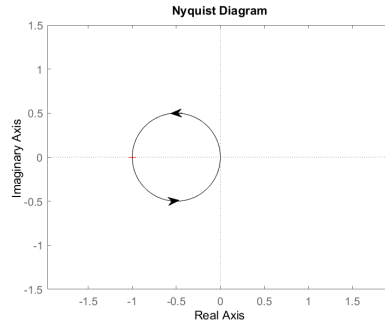


Figure 17.10: Nyquist diagram of $\frac{1}{s-1}$, from Example 17.7.

$-0.0021 < K < 0.0021$. (We can add $K = 0$ to the interval since there will be no feedback and the output will be always zero: control will be a failure, but the output *is* stable.) \square

Example 17.7. Figure 17.10 shows the Nyquist diagram of $\frac{1}{s-1}$. What proportional controllers can stabilise this plant in closed loop?

Counter-clockwise encirclements are negative clockwise encirclements $P = 1$, and real points from -1 to 0 are encircled once counter-clockwise, i.e. they are encircled clockwise -1 times. From table

\mathbb{R}	$-\infty$	-1	0	$+\infty$	
$-\frac{1}{K} \in$	$] -\infty, -1[$	$] -1, 0[$	$] 0, +\infty[$		
$K \in$	$] 0, 1[$	$] 1, +\infty[$	$] -\infty, 0[$		(17.10)
N	0	-1	0		
P	1	1	1		
Z	1	0	1		

we see that we must have $K > 1$. \square

What Nyquist diagram should be plot **Remark 17.3.** Remember that if, instead of the cases of Figure 17.8, the control loop is different, the characteristic equation is not the same. Thus:

- in case (b) of Figure 16.1, when a plant $G(s)$ is controlled in closed loop by a controller $KC(s)$ with variable gain K , we must find the Nyquist diagram of $C(s)G(s)$;
- in case (c) of Figure 16.1, when a plant $KG(s)$ with variable gain K is controlled in closed loop by a controller $C(s)$, we must find the Nyquist diagram of $C(s)G(s)$;
- in the case of Figure 16.10, when a plant $G(s)$ is controlled in closed loop by a proportional controller K , and there is in the feedback branch a sensor $H(s)$, we must find the Nyquist diagram of $G(s)H(s)$. \square

17.4 The Nyquist diagram when there are poles on the imaginary axis

Nyquist contour for poles at the origin or imaginary poles When there are poles on the imaginary axis, the Nyquist contour must be modified with semi-circles with vanishing radius $r' \rightarrow 0^+$, as shown in Figure 17.11 for a situation in which there is a pole at the origin and a pair of complex conjugate imaginary poles. The number and location of the semi-circles depends, of course, of the number and location of the poles on the imaginary axis. The semi-circles are on the right side of the imaginary axis so as to always have $s > 0$, so that only unstable poles of the transfer function are ever encircled.

These vanishing semi-circles near poles originate curves with a radius that increases to infinity as $r' \rightarrow 0^+$. MATLAB function `nyquist` does *not* plot these so-called *curves at infinity* (only the polar plot and its complex conjugate), though we can find them numerically using small values for r' .

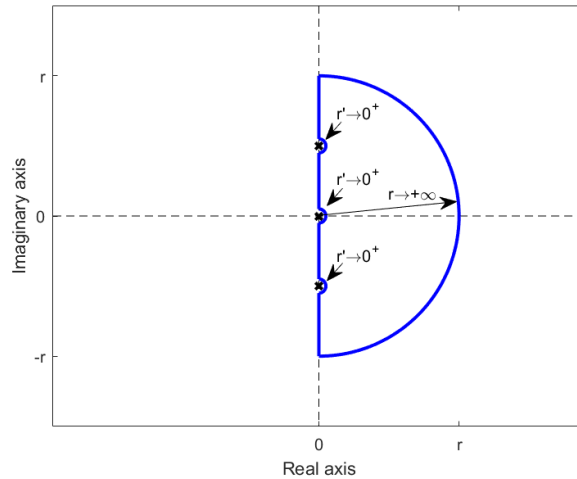


Figure 17.11: The Nyquist contour (blue) in the complex plane, when there are poles on the imaginary axis.

Example 17.8. Figure 17.12 shows the Nyquist diagrams of plants

$$G_1(s) = \frac{s + 0.1}{s(s + 1)} \quad (17.11)$$

$$G_2(s) = \frac{s + 0.1}{s(s - 1)} \quad (17.12)$$

$$G_3(s) = \frac{s - 0.1}{s(s + 1)} \quad (17.13)$$

as MATLAB plots them, and then with the curves at infinity added. The same Figure shows the mapping of two approximations of the Nyquist path: one with $r = 10$, $r' = \frac{1}{10}$, and another with $r = 100$, $r' = \frac{1}{100}$, which is of course a better approximation of the Nyquist diagram. The first approximation already shows where the curves at infinity will be; the second is quite clear. \square

To know without numerical calculations whether a curve in the Nyquist plot caused by a pole at the origin turns clockwise or counter-clockwise, it suffices to see what happens with the only real point of the vanishing semi-circle of the Nyquist path, $\epsilon \rightarrow 0^+$. See Figure 17.13.

Finding curves at infinity

Example 17.9. For transfer functions (17.11)–(17.13), we have

$$\lim_{\epsilon \rightarrow 0^+} G_1(s)|_{s=\epsilon} = \lim_{\epsilon \rightarrow 0^+} \frac{\epsilon + 0.1}{\epsilon(\epsilon + 1)} = +\infty \quad (17.14)$$

$$\lim_{\epsilon \rightarrow 0^+} G_2(s)|_{s=\epsilon} = \lim_{\epsilon \rightarrow 0^+} \frac{\epsilon + 0.1}{\epsilon(\epsilon - 1)} = -\infty \quad (17.15)$$

$$\lim_{\epsilon \rightarrow 0^+} G_3(s)|_{s=\epsilon} = \lim_{\epsilon \rightarrow 0^+} \frac{\epsilon - 0.1}{\epsilon(\epsilon + 1)} = -\infty \quad (17.16)$$

Thus:

- The Nyquist diagram of $G_1(s)$ will have a curve at infinity on the right side of the complex plane; we might, with some abuse of terminology, say that the curve passes through $+\infty$.
- The Nyquist diagrams of $G_2(s)$ and $G_3(s)$ will have curves at infinity on the left side of the complex plane; we might, with some abuse of terminology, say that in each diagram the curve passes through $-\infty$. \square

Notice that multiple poles at the origin result in an increasingly large curve at infinity: *Multiple poles at the origin*

- 1 pole at the origin $\frac{1}{s}$ originates a curve with 180° ;
- 2 poles at the origin $\frac{1}{s^2}$ originate a curve with 360° ;

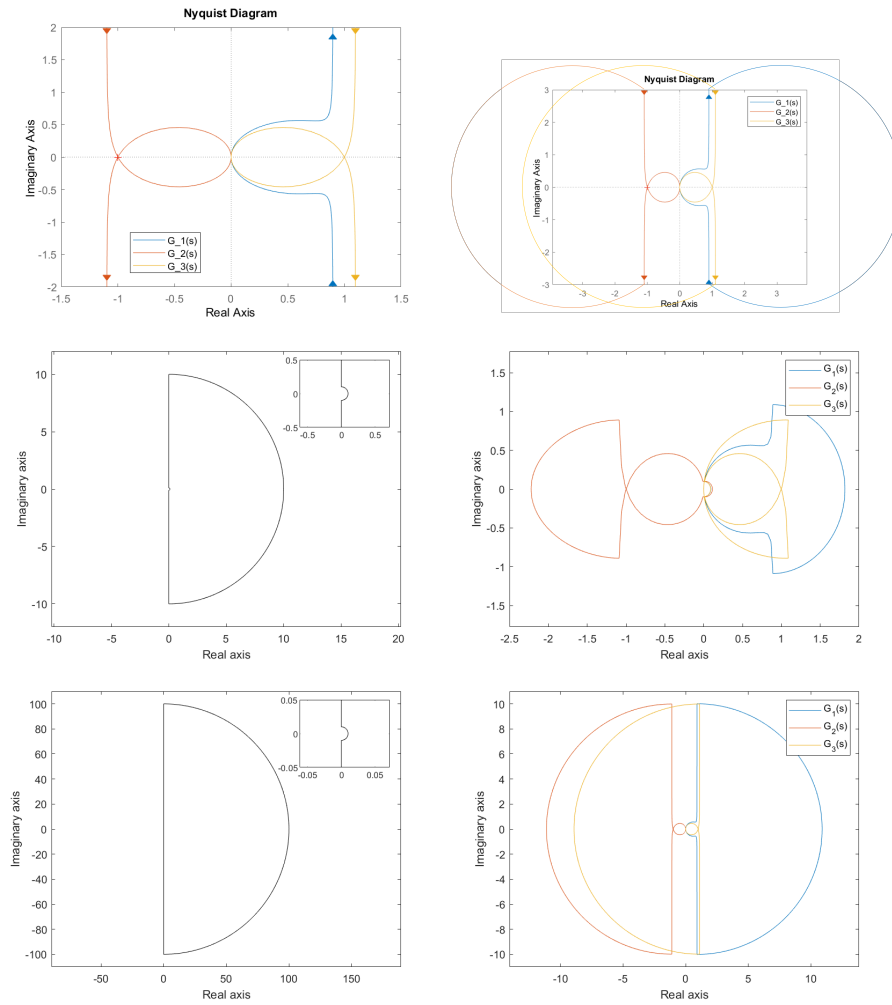


Figure 17.12: Nyquist diagrams of (17.11)–(17.13). Top left: diagrams plotted with `nyquist`. Top right: Nyquist diagrams, obtained adding the curves at infinity to the plots of `nyquist`. Middle left: Nyquist path with a finite value of $r = 10$ and a non-vanishing value of $r' = \frac{1}{10}$. Middle right: approximations of the Nyquist diagrams, that map the contour on the left. Bottom left: Nyquist path with a larger $r = 100$ and a smaller $r' = \frac{1}{100}$. Bottom right: approximations of the Nyquist diagrams, that map the contour on the left.

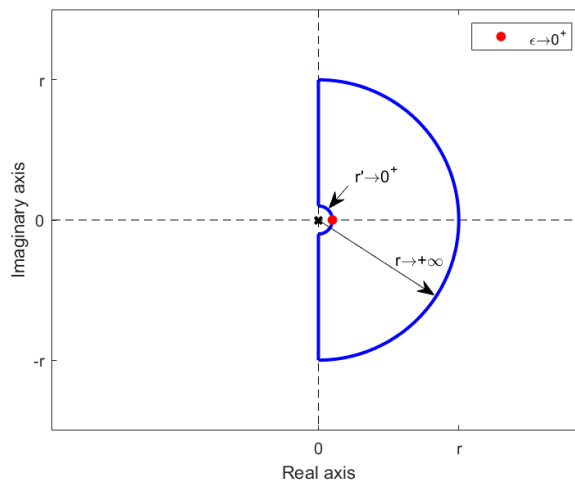


Figure 17.13: The Nyquist contour (blue) in the complex plane, when there is a pole at the origin, evidencing the only vanishing real point $\epsilon \rightarrow 0^+$.

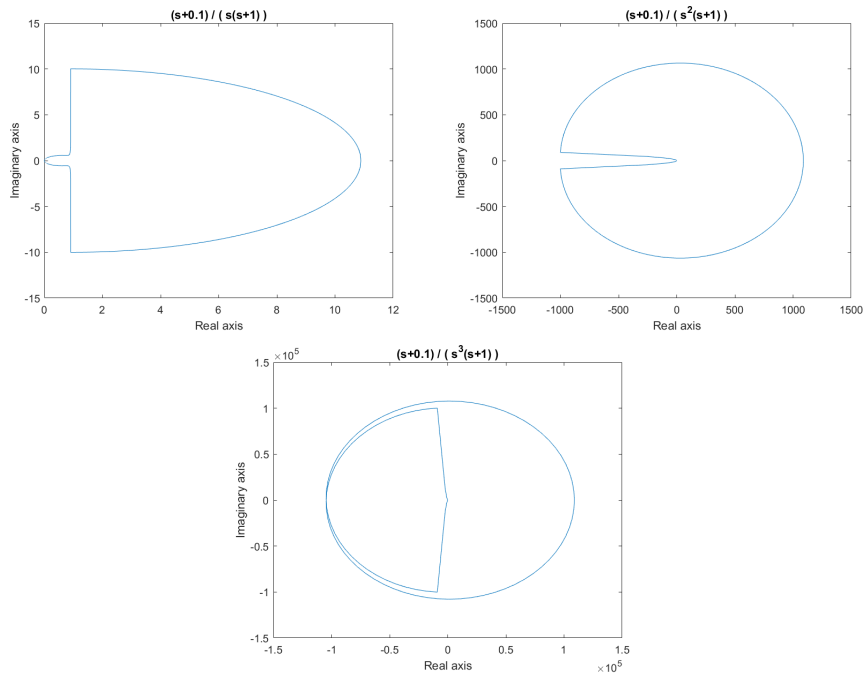


Figure 17.14: Approximations of the Nyquist diagrams of (17.17)–(17.19). The curve at infinity cause by n poles at the origin has an angle given by $n 180^\circ$.

- 3 poles at the origin $\frac{1}{s^3}$ originate a curve with 540° ;
- and in general n poles at the origin $\frac{1}{s^n}$ originate a curve with $n 180^\circ$.

Example 17.10. Figure 17.14 shows fairly good approximations of the Nyquist plots of

$$G_1(s) = \frac{s + 0.1}{s(s + 1)} \quad (17.17)$$

$$G_2(s) = \frac{s + 0.1}{s^2(s + 1)} \quad (17.18)$$

$$G_3(s) = \frac{s + 0.1}{s^3(s + 1)} \quad (17.19)$$

obtained mapping the Nyquist path from Figure 17.12, bottom left ($r = 100$ and $r' = \frac{1}{100}$). \square

Glossary

“Oltre la porta si scopre un sepolcro a sette lati e sette angoli, illuminato prodigiosamente da un sole artificiale. Nel mezzo, un altare rotondo, ornato da vari motti o emblemi, del tipo NEQUAQUAM VACUUM. . .”

“Ne quà quà? Firmato Donald Duck?”

“È latino, hai presente? Vuol dire il vuoto non esiste.”

“Meno male, altrimenti sai che orrore.”

Umberto Eco (1932 — †2016), *Il pendolo di Foucault* (1988), IV 29

clockwise no sentido retrógrado, no sentido horário, no sentido dos ponteiros do relógio

contour contorno

counter-clockwise no sentido direto, no sentido anti-horário, no sentido contrário ao dos ponteiros do relógio

curve at infinity curva no infinito

encirclement enrolamento

mapping mapeamento

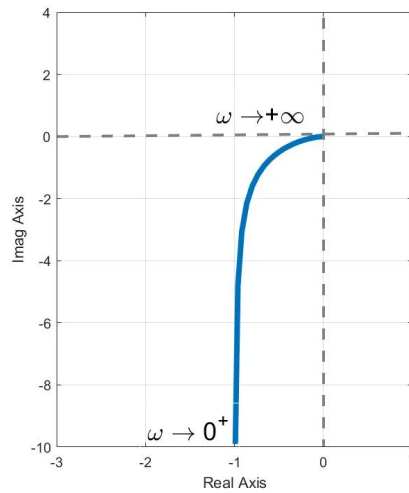
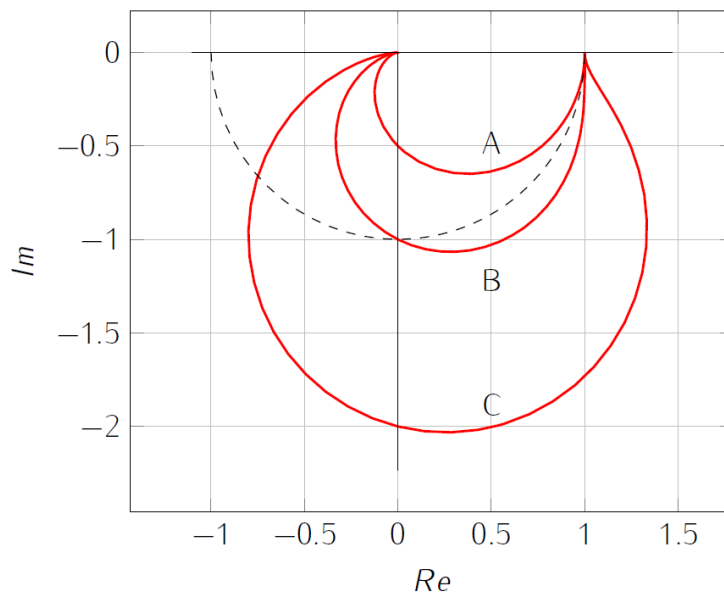


Figure 17.15: Polar diagram of the plant from Exercise 1.

Figure 17.16: Polar diagram of $G(s)$, from Exercise 2.

Exercises

- Figure 17.15 shows the polar plot of a plant.
 - What is the type of the plant?
 - How many poles and zeros does the plant have?
- Figure 17.16 shows the polar diagram of the second-order plant

$$G(s) = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}, \quad (17.20)$$

where $\zeta \in \{0.25, 0.5, 1\}$. Find the correspondence between curves A, B and C, and the three possible values of ζ .

- The closed loop in Figure 17.17 has the open-loop transfer function

$$G(s)H(s) = \frac{K(s+1)}{s(s+2)(s+3)}, \quad K \geq 0. \quad (17.21)$$

- Plot its root locus.
- Plot its Nyquist diagram.
- Analyse the closed loop's stability for different values of K .

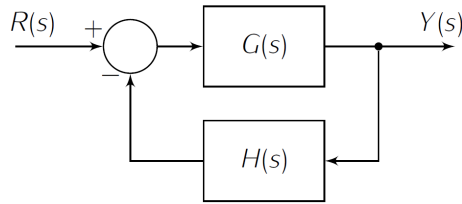


Figure 17.17: Control loop from Exercise 3.

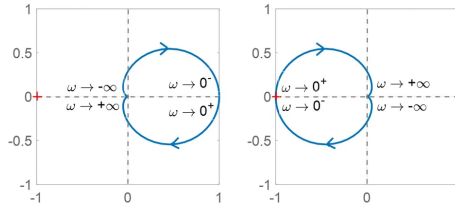


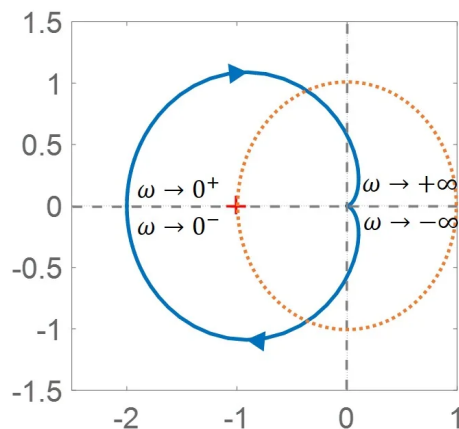
Figure 17.18: Nyquist diagrams from Exercise 4.

4. Which of the two Nyquist diagrams of Figure 17.18 corresponds to transfer function $\frac{-10}{(s+1)(s+10)}$?
5. Figure 17.19 shows the Nyquist diagram of plant $\frac{-20}{(s+1)(s+10)}$.
 - (a) Is the plant stable in closed loop?
 - (b) If the plant is controlled in closed loop with a proportional controller K , for which values of K will the loop be stable?
 - (c) If plant $\frac{-10}{(s+1)(s+10)}$ is controlled in closed loop with a proportional controller K , for which values of K will the loop be stable?
6. The attitude θ of a space station is related to the thrust δ of two motors that can be used to correct it by

$$G(s) = \frac{\Theta(s)}{\Delta(s)} = \frac{15}{s(s^2 + 15s + 75)} \quad (17.22)$$

The attitude will be controlled in closed loop with a proportional controller K .

- (a) Plot its Nyquist diagram.
- (b) Apply Nyquist's criterion and find the values of K for which the control loop is stable.
- (c) Confirm that these values of K stabilise the plant using the Routh-Hurwitz criterion.
- (d) Plot the root-locus of $G(s)$ and relate it to the results you found.


 Figure 17.19: Nyquist diagram of plant $\frac{-20}{(s+1)(s+10)}$ from Exercise 5.

Chapter 18

Stability margins

In eo flumine pons erat. Ibi praesidium ponit et in altera parte fluminis Q. Titurium Sabinum legatum cum sex cohortibus relinquit; castra in altitudinem pedum XII vallo fossaque duodeviginti pedum muniri iubet.

Caius Iulius CÆSAR (100 BC — †24 BC), *Commentarii de bello Gallico* (c. 50 BC), II 5, 6

Stability margins are another tool that, just like the root locus, can be used

- to design a proportional controller,
- to design the gain of a controller that has poles and zeros, once these are known,
- to see what happens to a closed loop control system with a known controller, when the gain of the plant changes for some reason.

18.1 Stability margins in the Nyquist diagram

Consider the first block diagram of Figure 17.8; that is to say, if there is a controller $C(s)$ and a plant $G_p(s)$, both are merged in $G(s) = C(s)G_p(s)$. Of course, this is the same as the second block diagram of Figure 17.8 with $K = 1$. Suppose that $G(s)$ has no unstable poles. Then the Nyquist criterion (Theorem 17.3) shows that the closed loop will have no unstable poles if there are no encirclements around -1 , since in that case $Z = N + P = 0 + 0$.

Example 18.1. Figure 18.1 shows three Nyquist diagrams exemplifying three possible cases:

- On the left,

$$G(s) = \frac{1}{(s+2)(s^2+0.3s+1)} \quad (18.1)$$

There are no encirclements of point -1 , so the closed loop is stable.

- On the centre,

$$G(s) = \frac{1.68}{(s+2)(s^2+0.3s+1)} \quad (18.2)$$

The diagram passes through point -1 , so the closed loop is marginally stable.

- On the right,

$$G(s) = \frac{3}{(s+2)(s^2+0.3s+1)} \quad (18.3)$$

There are two encirclements, so the closed loop is unstable. □

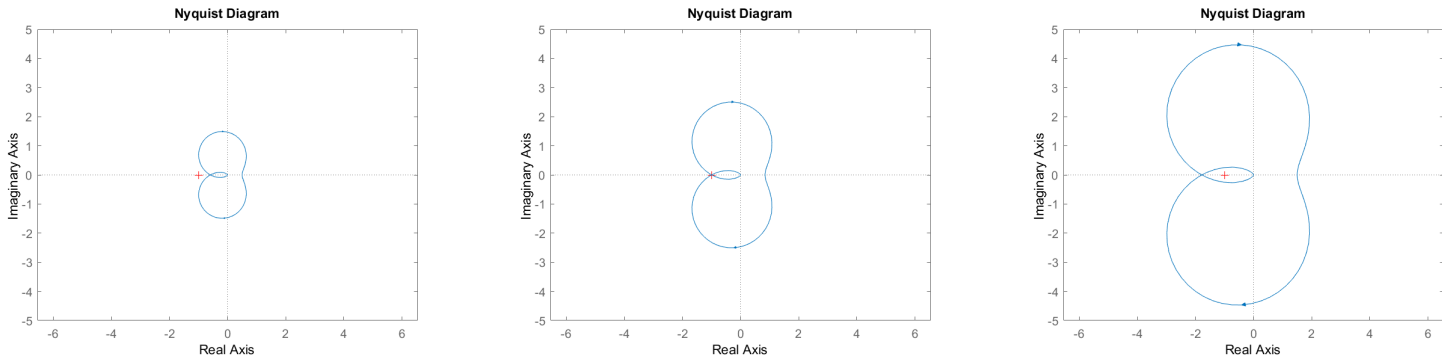


Figure 18.1: Nyquist diagrams of (18.1)–(18.3), from Example 18.1.

The **gain margin** (GM) and the **phase margin** (PM), known together as stability margins, show us how far the frequency response of $G(s)$ is from the situation when the closed loop is critically stable:

1. When the closed loop is stable:

- The gain margin shows us how much the gain of the frequency response of $G(s)$ could increase until the closed loop becomes critically stable. This is the same as increasing open loop gain K . Thus, the gain margin is a value larger than 1; this gain increase is usually given in dB, and so it has a positive value. See Figure 18.2.
- The phase margin shows us how much the phase of the frequency response of $G(s)$ could decrease until the closed loop becomes critically stable. This decrease has a positive value. As is clear from Figure 18.3, since at critical stability the frequency response should pass through, but not encircle, point -1 , we need only consider the phase decrease as taking place on a unit radius circle, i.e. for $|G(j\omega)| = 1$.

If the closed loop is stable, $GM > 0$ and $PM > 0$

GM is usually given in dB

2. When the closed loop is unstable:

- The gain margin shows us how much the gain of the frequency response of $G(s)$ must decrease until the closed loop becomes critically stable. This is the same as decreasing the open loop gain K . Thus, the gain margin is a value smaller than 1; this phase decrease, given in dB, has a negative value. See Figure 18.4.
- The phase margin shows us how much the phase of the frequency response of $G(s)$ must increase until the closed loop becomes critically stable. This phase *increase* is given as a phase *decrease* with a *negative* value. Once more, and as is clear from Figure 18.5, since at critical stability the frequency response should go through -1 , we need only consider the phase increase as taking place on a unit radius circle.

If the closed loop is unstable, $GM < 0$ and $PM < 0$

In short, if an open loop has no unstable poles, then the corresponding closed loop will be

- stable, if both GM and PM are positive;
- unstable, if both GM and PM are negative.

Let us define quantitatively the stability margins:

- On the Nyquist diagram, the gain margin can be seen on the negative real axis, on which lies point -1 . The negative real axis consists of points where

$$\arg G(j\omega) = \dots, +540^\circ, +180^\circ, -180^\circ, -540^\circ, \dots \quad (18.4)$$

The gain margin is read at a phase crossover frequency

So, the frequency response on the negative real axis will correspond to an output and input in phase opposition. Remember from Definition 10.10 that a frequency ω_{pc} at which this happens is called a **phase crossover frequency**.

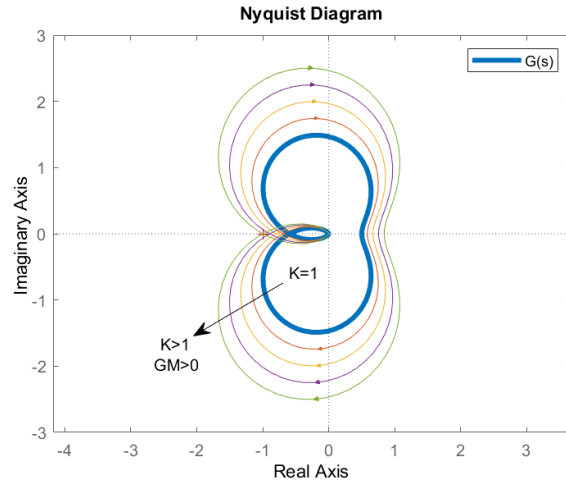


Figure 18.2: The positive gain margin of $G(s)$ shows how much the open loop gain K could increase until the closed loop becomes critically stable.

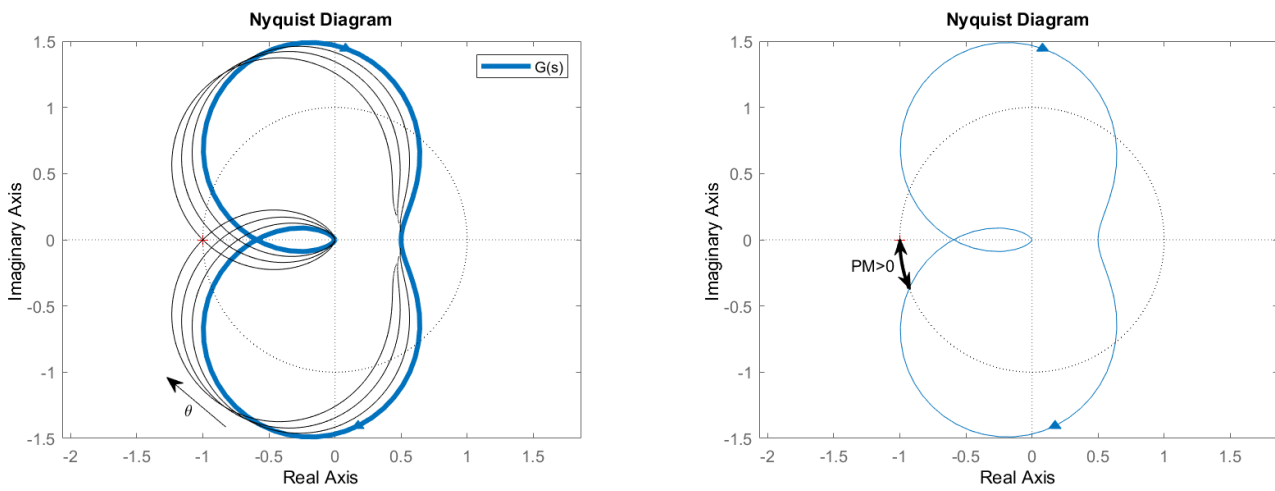


Figure 18.3: The positive phase margin of $G(s)$ shows us how much the phase could decrease until the closed loop becomes critically stable. $G(s)$ is given by (18.1). Left: the whole phase of $G(s)$ is decreased, and thus the entire frequency response is rotated clockwise around 0° , while its complex conjugate is rotated counter-clockwise around 0° . Right: the phase margin can be read on the unit radius circle.

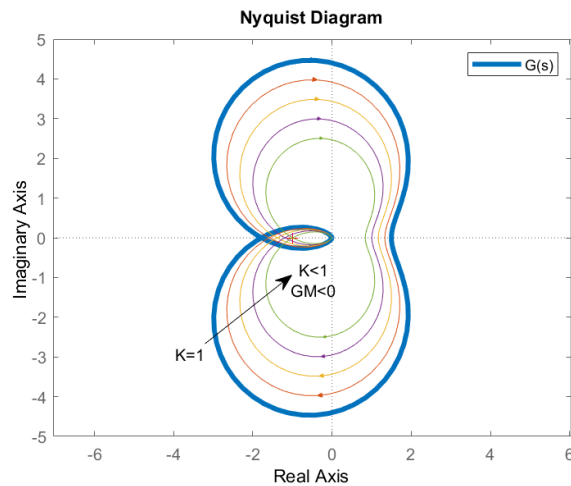


Figure 18.4: The negative gain margin of $G(s)$ shows how much the open loop gain K must decrease until the closed loop becomes critically stable.

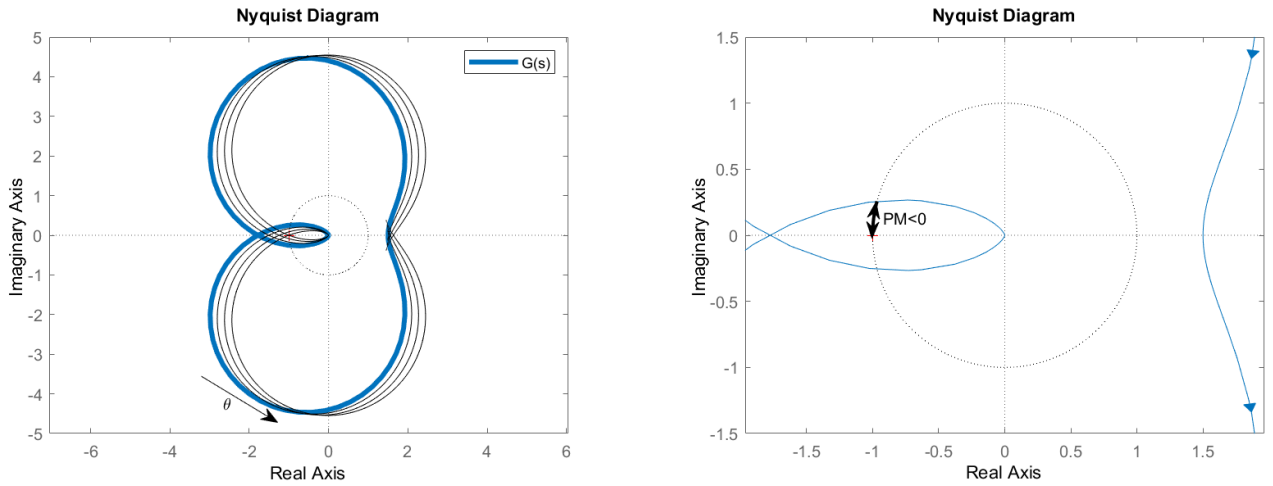


Figure 18.5: The negative phase margin of $G(s)$ shows us how much the phase must increase until the closed loop becomes critically stable (it is negative because the phase margin is given as a phase *decrease*). $G(s)$ is given by (18.3). Left: the whole phase of $G(s)$ is increased, and thus the entire frequency response is rotated counter-clockwise around 0° , while its complex conjugate is rotated clockwise around 0° . Right: the phase margin can be read on the unit radius circle.

- If the gain at the crossover frequency $|G(j\omega_{pc})|$ is multiplied by its inverse $\frac{1}{|G(j\omega_{pc})|}$, then it becomes 1. A gain of 1 with a phase given by (18.4) means that the frequency response goes through point -1 , and the closed loop is marginally stable. So, the gain margin is $\frac{1}{|G(j\omega_{pc})|}$, or, in dB,

$$\text{GM} = 20 \log_{10} \frac{1}{|G(j\omega_{pc})|} = -20 \log_{10} |G(j\omega_{pc})| \quad (18.5)$$

Notice that the minus sign in this definition confirms what we saw above:

- if the closed loop is stable, then $|G(j\omega_{pc})| < 1$, and $\text{GM} = -20 \log_{10} |G(j\omega_{pc})| > 0$ dB;
- if the closed loop is unstable, then $|G(j\omega_{pc})| > 1$, and $\text{GM} = -20 \log_{10} |G(j\omega_{pc})| < 0$ dB.

The phase margin is read at a gain crossover frequency

- Remember from Definition 10.9 that, if at some frequency ω_{gc} the frequency response crosses the unit radius circle on the Nyquist diagram over which the phase margin is read, i.e. if $|G(j\omega_{gc})| = 1$, we have a **gain crossover frequency**.
- The phase margin, as shown in Figures 18.3 and 18.5, is given by

$$\text{PM} = 180^\circ + \arg G(j\omega_{gc}) \quad (18.6)$$

However, notice that there is some ambiguity in this expression, since the crossover frequency need not be at -180° : it can take place at any value $-180^\circ + k360^\circ$, $k \in \mathbb{Z}$. It is better to write

$$\text{PM} = 180^\circ + k360^\circ + \arg G(j\omega_{gc}) \quad (18.7)$$

with k chosen such that a phase increase of PM will cause an encirclement of -1 in the Nyquist plot.

Expressions (18.5) and (18.6)–(18.7) have to be complemented with a consideration of the cases when there are several gain or phase crossover frequencies, or none at all.

GM = $+\infty$ dB if ω_{pc} does not exist

- If there is no phase crossover frequency, the gain can be arbitrarily increased without causing any encirclement of -1 . Thus, the gain margin is infinite, $\text{GM} = +\infty$ dB. (Of course, in practice gains cannot be increased arbitrarily, actuators saturate sooner or later, at some point the models are no longer valid, etc..) See an example in Figure 18.6.

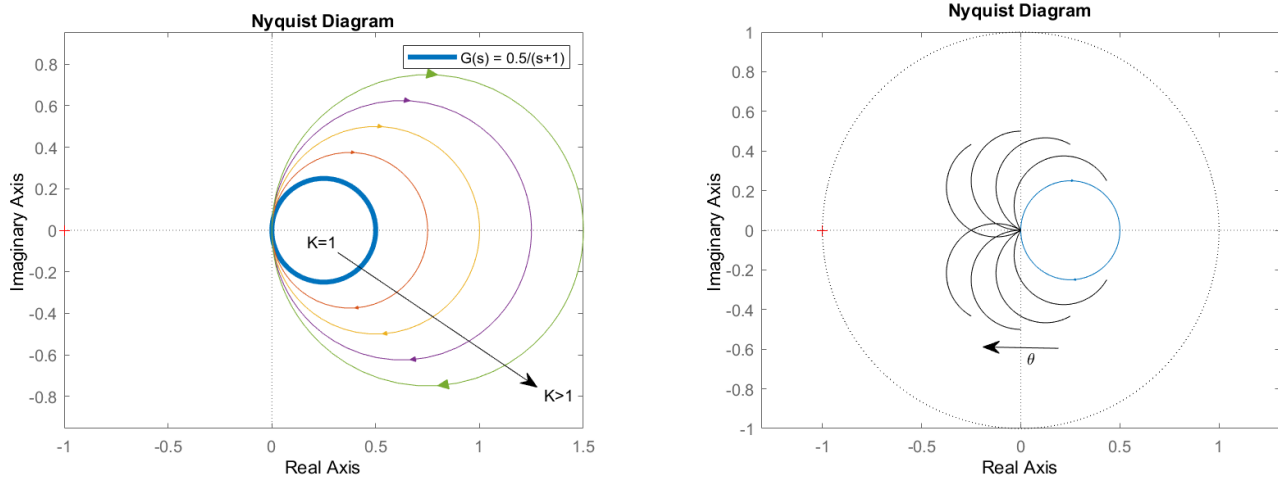


Figure 18.6: Nyquist plot of $\frac{0.5}{s+1}$. Left: The phase never reaches the negative real axis. So, the gain can be arbitrarily increased without the appearance of encirclements of -1 : the gain margin is infinite. Right: the gain is always below 1. So, the phase can be arbitrarily decreased without the appearance of encirclements of -1 : the phase margin is not well defined, or is infinite.

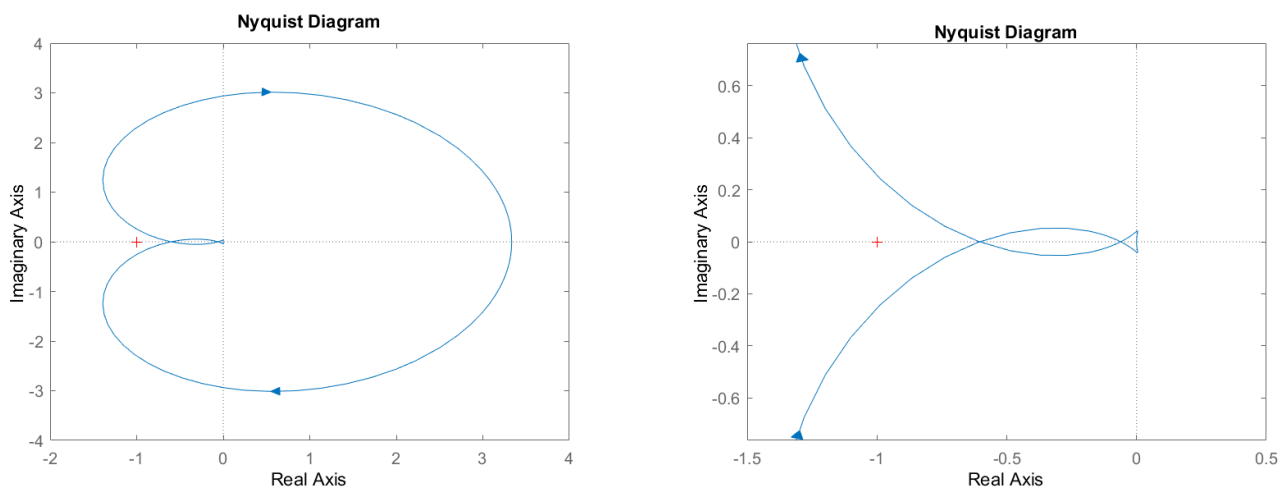


Figure 18.7: Nyquist plot of $\frac{s^2+s+10}{3(s^2+s+1)(s+1)}$, with three phase crossover frequencies. It is the first that determines the gain margin. Right: detail of the plot on the left.

$= +\infty$, or is not defined, if ω_{gc} does not exist

- If there is no gain crossover frequency, there is no point on the unit radius circle at which the phase can be measured for the phase margin, and so this margin can be said not to be defined. Or it can be said that the entire phase of $G(s)$ may decrease arbitrarily, without any encirclement of -1 , and thus the phase margin can be said to be infinite, $PM = +\infty$. See an example in Figure 18.6.
- If there are several phase crossover frequencies, the one that ought to be considered is the first that will cause encirclements of -1 as the gain increases. As the gain must decrease for high frequencies in a strictly proper transfer function, this phase crossover frequency is usually the first, corresponding to the lowest possible gain increase. See an example in Figure 18.7
- If there are several gain crossover frequencies, the one that ought to be considered is the first that will cause encirclements of -1 as the phase decreases. If the phase decreases for high frequencies, this will be the last crossover, as in Figures 18.3 and 18.5 for plants (18.1) and (18.3).

GM when there are several ω_{pc}

PM when there are several ω_{gc}

Remark 18.1. Always remember that we find the GM and PM of the *open loop* determine closed loop stability

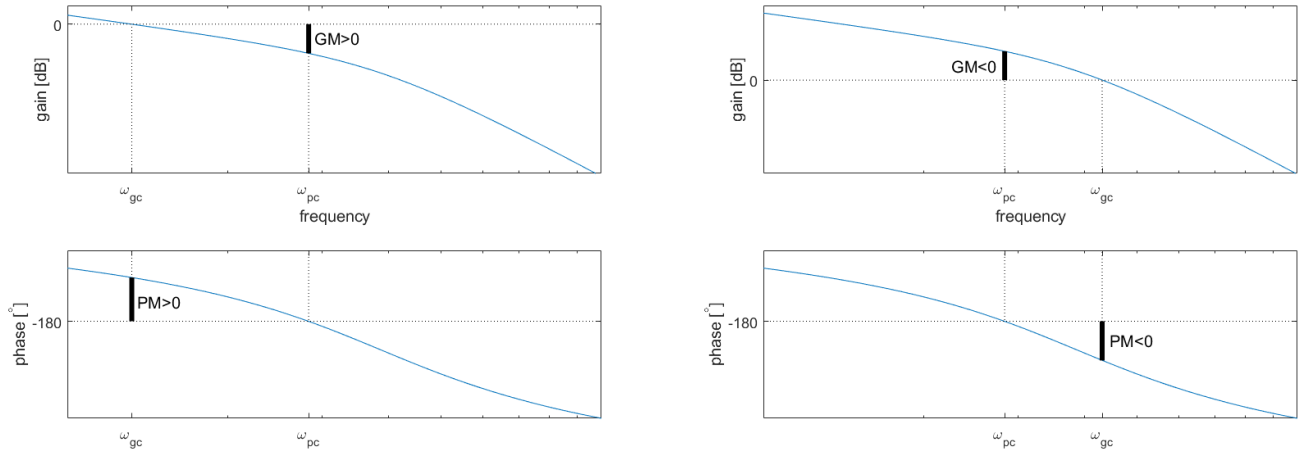


Figure 18.8: Reading the stability margins in the Bode diagram. Left: stable closed loop. Right: unstable closed loop.

loop to learn about the stability of the *closed loop*. Also remember that what is included in the open loop depends on the characteristic equation of the closed loop:

- in case (a) of Figure 16.1, when a plant $G(s)$ is controlled in closed loop by a proportional controller K , we must find the stability margins of $G(s)$;
- in case (b) of Figure 16.1, when a plant $G(s)$ is controlled in closed loop by a controller $KC(s)$ with variable gain K , we must find the stability margins of $C(s)G(s)$;
- in case (c) of Figure 16.1, when a plant $KG(s)$ with variable gain K is controlled in closed loop by a controller $C(s)$, we must find the stability margins of $C(s)G(s)$;
- in the case of Figure 16.10, when a plant $G(s)$ is controlled in closed loop by a proportional controller K , and there is in the feedback branch a sensor $H(s)$, we must find the stability margins of $G(s)H(s)$. \square

18.2 Stability margins in the Bode diagram

The stability margins can also be shown in the Bode diagram. In fact, it is much easier to read them in the Bode diagram than on the Nyquist diagram (although it is on the Nyquist diagram that it can be seen why the closed loop is stable if the margins are positive, and unstable if negative). Figure 18.8 illustrates how the gain and phase margins are read in a generic Bode diagram.

Example 18.2. We want to control plant $G(s) = \frac{10}{s(s+10)^2}$ with a proportional controller $C(s) = 10$. Will the closed loop be stable?

Writing the open loop as

$$C(s)G(s) = \frac{1}{s} \underbrace{\frac{10}{s+10} \frac{10}{s+10}}_{\text{cut-off at } 10 \text{ rad/s}} \underbrace{\frac{100}{10 \times 10}}_1 \quad (18.8)$$

we can easily plot the asymptotic Bode diagram in Figure 18.9, and see that:

- there is one gain crossover frequency at $\omega_{gc} = 1 \text{ rad/s}$;
- at frequency ω_{gc} , the phase of $C(s)G(s)$ is close to -90° , per the asymptotes;
- consequently the phase margin is $\text{PM} = -90^\circ + 180^\circ = 90^\circ$;
- in reality $\angle C(j\omega_{gc})G(j\omega_{gc})$ is somewhat below, since the asymptotic phase behaviour is only an approximation;

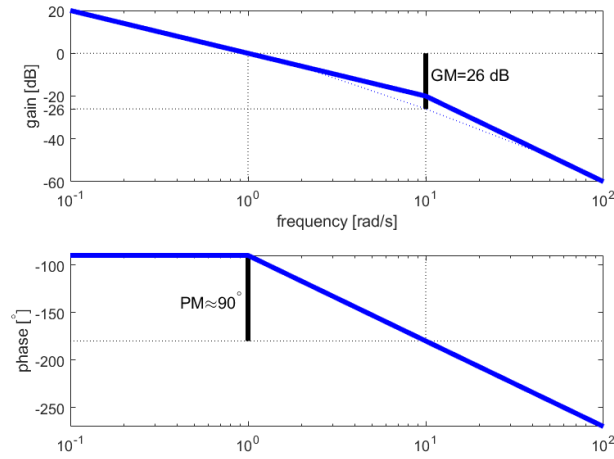


Figure 18.9: Asymptotic Bode diagram of open loop $C(s)G(s) = 10 \frac{10}{s(s+10)^2}$, from Example 18.2.

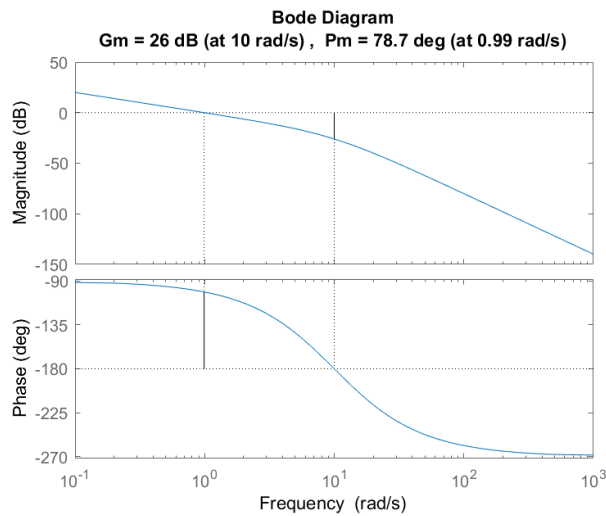


Figure 18.10: Stability margins of open loop $C(s)G(s) = 10 \frac{10}{s(s+10)^2}$, from Examples 18.2 and 18.3.

- in reality PM is somewhat less than 90° ;
- there is one phase crossover frequency at $\omega_{pc} = 10$ rad/s;
- at frequency ω_{pc} , the gain of $C(s)G(s)$ is -20 dB, per the asymptotes;
- remembering (11.90), we see that the gain value is in fact $|C(j\omega_{gc})G(j\omega_{gc})| = -20 - 3 - 3 = -26$ dB, as shown in Figure 18.9;
- consequently the gain margin is $PM = 26$ dB.

The easiest way of finding stability margins in MATLAB is using function `MATLAB function margin`, which is similar to function `bode`.

Example 18.3. The margins of Example 18.2 are found as:

```
>> s=tf('s'); figure, margin(10*10/(s*(s+10)^2))
```

and we obtain Figure 18.10. Function `margin` can also return the margins, if used with output arguments. \square

Example 18.4. Since the gain margin of the last two examples is 26 dB, or, in absolute value, $10^{26/20} = 19.95$, we could still increase the gain up to 19.95 times, and the closed loop would still be stable. This is clear from the Bode

diagram in Figure 18.10: if the gain goes up less than 26 dB, we still have $GM > 0$ and $PM > 0$.

Replacing controller $C(s) = 10$ with controller $C(s) = 10 \times 19.95 = 199.5$, the loop would be marginally stable. \square

Example 18.5. The stability margins of (18.1) and (18.3) can be found and plotted with

```
s = tf('s');
G1 = 1/( (s+2)*(s^2+0.3*s+1) );
figure, margin(G1)
xlim([.1 10]) % the x axis begins at 0.1 rad/s and ends at 10 rad/s
G3 = 3/( (s+2)*(s^2+0.3*s+1) );
figure, margin(G3)
xlim([.1 10])
```

See Figure 18.11. The values returned can be found from the definitions; consider for instance (18.1):

- To find the gain crossover frequency, we first calculate

$$G(s) = \frac{1}{(s+2)(s^2+0.3s+1)} = \frac{1}{s^3+2.3s^2+1.6s+2} \quad (18.9)$$

$$G(j\omega) = \frac{1}{-j\omega^3 - 2.3\omega^2 + 1.6j\omega + 2} = \frac{1}{(2 - 2.3\omega^2) + j(1.6\omega - \omega^3)} \quad (18.10)$$

$$\begin{aligned} |G(j\omega)| &= \frac{1}{\sqrt{(2 - 2.3\omega^2)^2 + (1.6\omega - \omega^3)^2}} \\ &= \frac{1}{\sqrt{\omega^6 + \omega^4(2.3^2 - 3.2) + \omega^2(1.6^2 - 9.2) + 4}} \end{aligned} \quad (18.11)$$

We could now find ω_{gc} by solving

$$|G(j\omega)| = 1 \Leftrightarrow \omega^6 + \omega^4(2.3^2 - 3.2) + \omega^2(1.6^2 - 9.2) + 4 = 1 \quad (18.12)$$

which will have to be done numerically, but since we now know that $\omega_{gc} = 1.13$ rad/s we can just verify that

$$1.13^6 + 1.13^4(2.3^2 - 3.2) + 1.13^2(1.6^2 - 9.2) + 4 = 1 \quad (18.13)$$

Notice from the Bode diagram in Figure 18.11 that there are, in fact, two gain crossover frequencies, and that we are considering the largest, since the phase is closer to -180° .

- From $\omega_{gc} = 1.13$ rad/s, the phase margin can be found:

$$\arg G(j\omega) = \arg 1 - \arg [(2 - 2.3\omega^2) + j(1.6\omega - \omega^3)] = -\arctan \frac{1.6\omega - \omega^3}{2 - 2.3\omega^2} \quad (18.14)$$

$$\arg G(j\omega_{gc}) = -\arctan \frac{1.6 \times 1.13 - 1.13^3}{2 - 2.3 \times 1.13^2} = -158.7^\circ \quad (18.15)$$

(Using MATLAB, the angle is found with `atan2d`, so that the result is in $]-180^\circ, +180^\circ]$.) So $PM = 180^\circ - 158.7^\circ = 21.3^\circ$, there being a small difference to the value in Figure 18.11 due to numerical approximations.

- To find the phase crossover frequency, we could solve

$$\arg G(j\omega) = -\arctan \frac{1.6\omega - \omega^3}{2 - 2.3\omega^2} = -180^\circ \quad (18.16)$$

It is of course easier to notice that this can only happen when the imaginary part of $\arg G(j\omega)$ is zero, i.e.

$$1.6\omega - \omega^3 = 0 \Leftrightarrow \omega = 0 \vee \omega = \pm\sqrt{1.6} = \pm 1.26 \quad (18.17)$$

Consequently, the only phase crossover frequency is 1.26 rad/s.

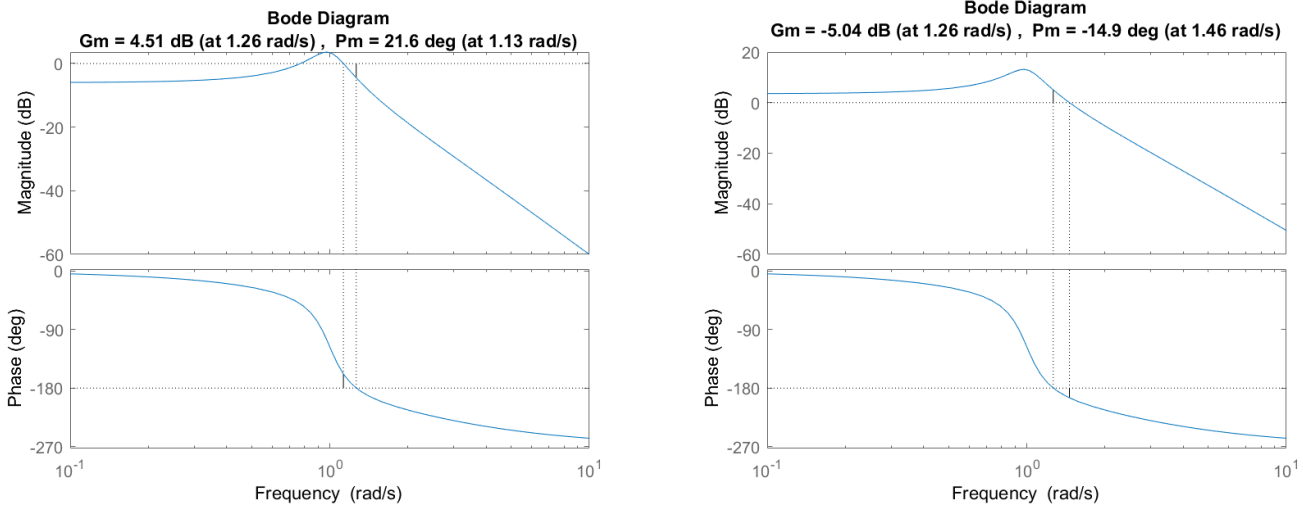


Figure 18.11: Bode diagrams of (18.1) and (18.3), showing the gain and phase margins.

- From $\omega_{pc} = 1.26$ rad/s, the gain margin can be found:

$$|G(j\omega_{pc})| = 20 \log_{10} \frac{1}{\sqrt{1.26^6 + 1.26^4(2.3^2 - 3.2) + 1.26^2(1.6^2 - 9.2) + 4}} = -4.4 \text{ dB} \quad (18.18)$$

and thus GM= 4.4 dB, there being a small difference to the value in Figure 18.11 due to numerical approximations.

- The gain margin could also have been found using the Routh-Hurwitz criterion. The closed loop which has gain K and $G(s)$ in the direct branch is given by

$$\begin{aligned} F(s) &= \frac{\frac{K}{s^3+2.3s^2+1.6s+2}}{1 + \frac{K}{s^3+2.3s^2+1.6s+2}} \\ &= \frac{K}{s^3 + 2.3s^2 + 1.6s + 2 + K} \end{aligned} \quad (18.19)$$

and thus the Routh-Hurwitz table is

$$\begin{array}{c|cc} s^3 & 1 & 1.6 \\ s^2 & 2.3 & 2 + K \\ \hline s & 1.6 - \frac{2+K}{2.3} & \\ 1 & 2 + K & \end{array} \quad (18.20)$$

Consequently, the closed loop will be stable if

$$\begin{cases} 1.6 - \frac{2+K}{2.3} > 0 \\ 2 + K > 0 \end{cases} \Rightarrow \begin{cases} K < 1.6 \times 2.3 - 2 = 1.68 \\ K > -2 \end{cases} \quad (18.21)$$

Since the gain cannot increase more than 1.68 times without the closed loop becoming unstable, we see that the gain margin is $\text{GM} = 20 \log_{10} 1.68 = 4.51$ dB. \square

18.3 Stability margins with opposite signs

Until now, we have seen that the closed loop will be stable or unstable depending on whether the stability margins of the open loop are respectively positive or negative. Remember that we assumed no unstable poles in the open loop, and thus there can be no encirclements of -1 in the Nyquist diagram if the closed loop is to be stable.

If the stability margins found from (18.4)–(18.7) or read in a Bode diagram turn out to have opposite signs, this means that the frequency response is such

If GM and PM have opposite signs, no conclusion about stability can be taken

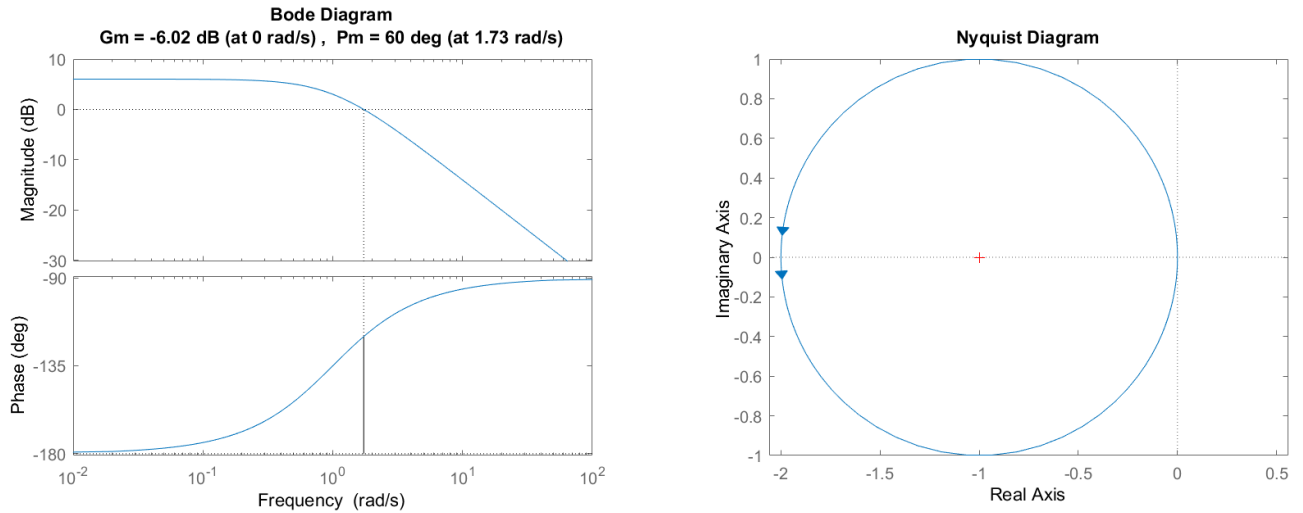


Figure 18.12: Left: Bode diagram of (18.22), from Example 18.6, with the stability margins. Right: Nyquist diagram.

that the margins give no information about encirclements of -1 in the Nyquist plot for an open loop without unstable poles. Thus, in that case closed loop stability cannot be determined from the gain and phase margins.

This can happen because there *are* unstable poles in the open loop, in which case the reasoning about encirclements of -1 no longer applies as assumed until now. But there are still other cases in which gain and phase margins have opposite signs.

Example 18.6. The stability margins of

$$G(s) = \frac{2}{s-1} \quad (18.22)$$

are shown in Figure 18.12. Since one is positive and the other negative, we cannot conclude whether the closed loop in Figure 17.8 with this transfer function will be stable or not. The Nyquist diagram shows the reason why: there is an unstable pole, and the counter-clockwise encirclement of -1 is needed so that the Nyquist stability criterion will return $Z = 1 - 1 = 0$. \square

Example 18.7. The stability margins of

$$G(s) = \frac{(s+1)^2}{s^3} \quad (18.23)$$

are shown in Figure 18.13. This time, $G(s)$ has no poles with positive real parts, but, again, we cannot conclude from the gain and phase margins whether the closed loop in Figure 17.8 with this transfer function will be stable or not. The reason why this case is complicated can be seen in the Nyquist diagram: the frequency response has a vertical asymptote and begins with a phase of -270° , so the plot encircles -1 counter-clockwise; then the three poles at the origin originate a curve at infinity with 540° that encircles -1 clockwise. The net result is zero encirclements, and the Nyquist criterion shows that the closed loop has $Z = 0 + 0 = 0$ unstable poles. (You can check this finding its transfer function, and the corresponding poles, or drawing the root-locus plot.) The stability margins of $G(s)$ tell us nothing, because the net number of encirclements is difficult to determine. \square

18.4 Stability margins and the root locus diagram

The root locus diagram can also be used to justify why the stability margins of an open loop transfer function determine closed loop stability (when they do, i.e. when they have the same sign). Let us consider a plant $G(s) = \frac{N(s)}{D(s)}$ controlled

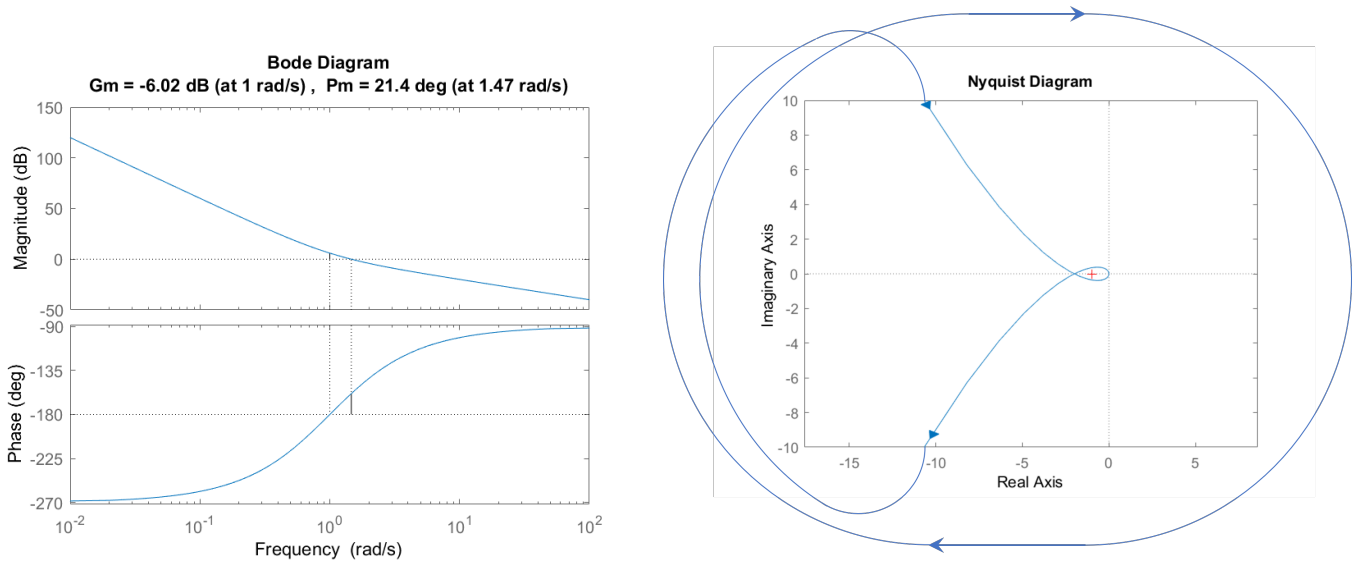


Figure 18.13: Left: Bode diagram of (18.23), from Example 18.7, with the stability margins. Right: Nyquist diagram.

in closed loop by a proportional controller $K > 0$, as in Figure 16.1, case (a). We saw in Section 16.3 that closed loop poles must verify the characteristic equation

$$1 + K \frac{N(s)}{D(s)} = 0 \Rightarrow K \frac{N(s)}{D(s)} = -1 \Rightarrow \begin{cases} K \frac{|N(s)|}{|D(s)|} = 1 \\ \angle \left[K \frac{N(s)}{D(s)} \right] = \pi + 2k\pi, k \in \mathbb{Z} \end{cases} \quad (18.24)$$

(where we made $|K| = K$ because $K > 0$) and we know that marginally stable poles are on the imaginary axis: $s = j\omega$, $\omega \in \mathbb{R}$. Thus, marginally stable closed loop poles verify

$$\begin{aligned} & K \frac{|N(j\omega)|}{|D(j\omega)|} = 1 \\ \Leftrightarrow & \underbrace{20 \log_{10} K}_{K \text{ in dB}} + \underbrace{20 \log_{10} |G(j\omega)|}_{\text{gain in dB at frequency } \omega} = 0 \text{ dB} \end{aligned} \quad (18.25)$$

and

$$\begin{aligned} & \angle \left[K \frac{N(j\omega)}{D(j\omega)} \right] = \pi + 2k\pi \\ \Leftrightarrow & \underbrace{\angle K}_{0 \text{ rad}} + \angle G(j\omega) = \underbrace{\pi + 2k\pi, k \in \mathbb{Z}}_{\dots, 3\pi, \pi, -\pi, -3\pi, \dots} \\ \Leftrightarrow & \angle G(j\omega) = \underbrace{180^\circ + k 360^\circ, k \in \mathbb{Z}}_{\dots, 540^\circ, 180^\circ, -180^\circ, -540^\circ, \dots} \end{aligned} \quad (18.26)$$

In other words, ω must be both a **gain crossover frequency** and a **phase crossover frequency** of $KG(j\omega)$ (remember Definitions 10.9 and 10.10).

Figure 18.14 shows examples of this situation, for stable, minimum phase plants. Notice that, in all cases, lower values of K result in stable closed loops, and higher values of K in unstable closed loops. This is because there are closed loop poles diverging to infinity, which have positive real parts when the gain is large enough. If there are three or more such poles, this is inevitable, as can be seen from the angles of the asymptotes of the root locus; when there are only two branches diverging to infinity, they may or may not become unstable (since the asymptotes have $\pm 90^\circ$ angles).

Thus, notice that

- for lower values of K , when the closed loop is stable,
 - at the phase crossover frequency, the gain is now below 0 dB,

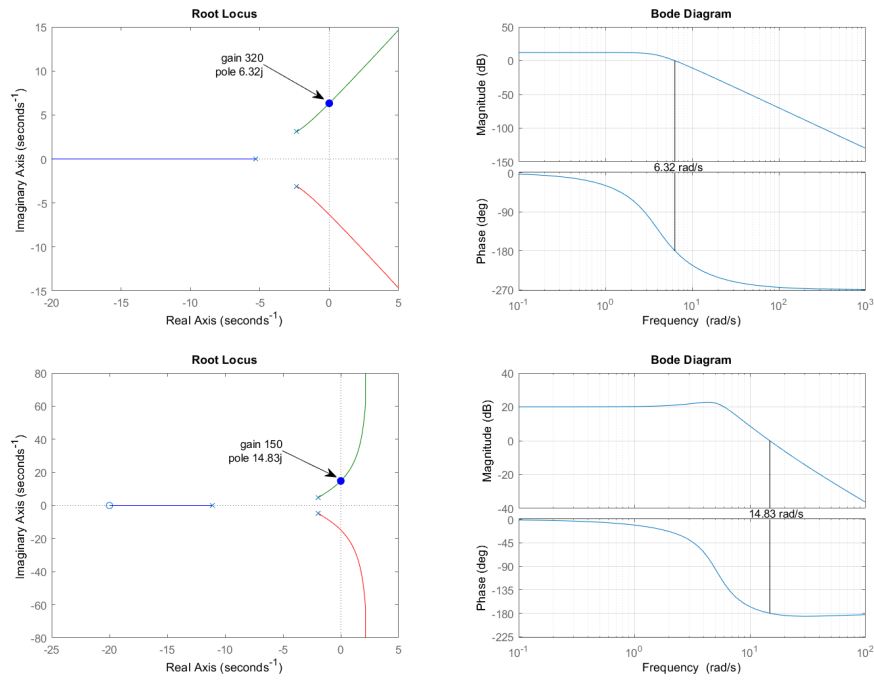


Figure 18.14: Top left: root locus of $G_1(s) = \frac{1}{s^3 + 10s^2 + 40s + 80}$; top right: Bode diagram of $\frac{320}{s^3 + 10s^2 + 40s + 80}$; a closed loop with controller $K = 320$ and plant $G_1(s)$ is marginally stable. Bottom left: root locus of $G_2(s) = \frac{s+20}{s^3 + 15s^2 + 70s + 300}$; top right: Bode diagram of $\frac{150(s+20)}{s^3 + 15s^2 + 70s + 300}$; a closed loop with controller $K = 150$ and plant $G_2(s)$ is marginally stable.

- at the new, lower gain crossover frequency, the phase is above -180° ;
- for higher values of K , when the closed loop is unstable,
 - at the phase crossover frequency, the gain is now above 0 dB,
 - at the new, higher gain crossover frequency, the phase is below -180° .

These are the same conclusions we reached about stability margins using the Nyquist diagram.

Glossary

—Então duvida que se falasse latim?—perguntou Henrique, sorrindo.

—Eu duvido. Não sei como os homens se podessem entender com aquela endiabrada contradança de palavras, com aquela desafinação que faz dar volta ao juízo de uma pessoa. Sabe o senhor o que é uma casa desarranjada, onde ninguém se lembra onde tem as suas coisas quando precisa d'ellas e passa o tempo todo a procural-as? Pois é o que é o latim. Abre a gente um livro e põe-se a traduzir e vae dizendo: «As armas, o homem e eu, canto, de Troia, e primeiro, das praias.» Quem percebe isto! Ora agora peguem n'estas palavras e em outras, que elles punham ás vezes em casa do diabo, e façam uma coisa que se entenda! É quasi uma adivinha. Ora adeus! E depois—continuou elle, entusiasmado com o riso de Henrique, suppondo-o de aprovação—e depois as diferentes maneiras de chamar a um objecto? Isso tambem tem graça. Nós cá dizemos por exemplo: «reino e reinos» e está acabado; lá não senhor; diz-se regnum e regna e regni e regno e regnis e até regnorum. Ora venham-me cá elogiar a tal lingua!

gain margin margem de ganho
phase margin margem de fase
stability margins margens de estabilidade

Exercises

- The distance d travelled by a robot controlled with a variable voltage u is given by $G(s) = \frac{d(s)}{u(s)} = \frac{1}{s(s+1)^2}$. This plant is controlled with a proportional controller K . Find the gain K for which the phase margin of the controlled loop is $PM = 60^\circ$.
- Plot the asymptotes of the Bode diagrams of the plants with the open-loop transfer functions given below. Mark the gain and phase margins in your plots. What can you say about the stability of the plants?

$$(a) G_1(s) = \frac{s^2}{(s+0.5)(s+10)}$$

$$(b) G_2(s) = \frac{10s}{(s+10)(s^2+s+2)}$$

$$(c) G_3(s) = \frac{(s+4)(s+20)}{(s+10)(s+80)}$$

- Find the stability margins of the following plants, and take conclusions about whether or not they are stable in closed loop:

$$(a) G_1(s) = \frac{2(s+3)}{s(s+1)(s+2)}$$

$$(b) G_2(s) = \frac{5(s+2)}{(s+1)(s^2+2s+4)}$$

$$(c) G_3(s) = \frac{1}{s(s+1)^2}$$

Chapter 19

The Nichols diagram

Bis über'n Kopf in's Tintenfaß
Tunkt sie der große Nikolas.

Heinrich HOFFMANN (1809 — †1894), *Der Struwwelpeter*, Die Geschichte von dem schwarzen Buben (1844)

We now know two different ways of graphically representing the frequency response of a system with transfer function $G(s)$:

- The **Bode diagram**, where we represent
 - the gain in dB $20 \log_{10} |G(j\omega)|$ as a function of frequency ω , and
 - the phase $\angle G(j\omega)$, usually in degrees, as a function of frequency *omega*.

A logarithmic scale is used for the x -axis with the frequency, usually given in rad/s.

- The **Nyquist diagram**, where we represent $G(j\omega)$ on the complex plane; i.e. we plot the imaginary part $\Im[G(j\omega)]$ on the y -axis, as a function of the real part $\Re[G(j\omega)]$ on the x -axis. There is no explicit representation of frequency ω .

In this chapter, we study a third possible representation, the **Nichols diagram**, *The Nichols diagram is an alternative to the Bode and Nyquist diagrams* in which we represent the gain in dB $20 \log_{10} |G(j\omega)|$ as a function of the phase $\angle G(j\omega)$, usually in degrees. There is no explicit representation of frequency ω .

19.1 Examples

Example 19.1. Figure 19.1 shows the Bode, Nyquist and Nichols diagrams of three first order transfer function without zeros, that amplify low frequencies. *Nichols diagram of a 1st order system* Notice how the Nichols diagram shows the phase decreasing from 0° to -90° , and the gain decreasing from its initial positive value in dB to $-\infty$ dB (i.e. 0 in absolute value). Also notice how the Nyquist and Nichols diagrams are the same for all three transfer functions, as there is no explicit representation of frequency, and the low frequency gain is the same. \square

Example 19.2. Figure 19.2 shows the Bode, Nyquist and Nichols diagrams of an underdamped second order transfer function without zeros, having a unitary gain for low frequencies. *Nichols diagram of a 2nd order system* Notice how the Nichols diagram shows the phase decreasing from 0° to -180° , and the gain increasing from its initial value value of 0 dB revealing a resonance frequency (and thus a damping coefficient lower than $\frac{\sqrt{2}}{2}$). \square

The Nichols diagrams in Figures 19.1 and 19.2 were plot with MATLAB using command `nichols`, which has a syntax similar to `bode` or `nyquist`.

MATLAB *command*
`nichols`

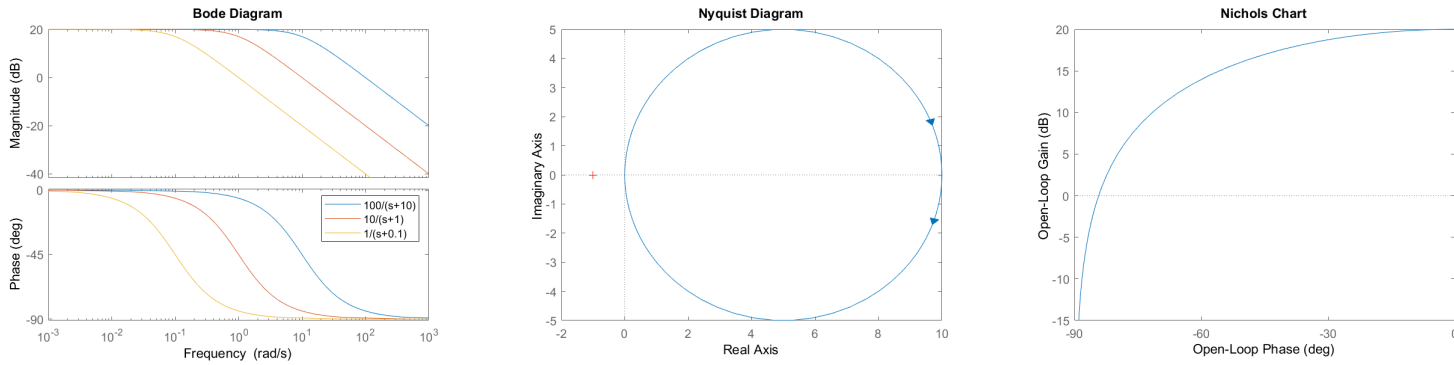


Figure 19.1: Bode, Nyquist and Nichols diagrams of three first order transfer functions.

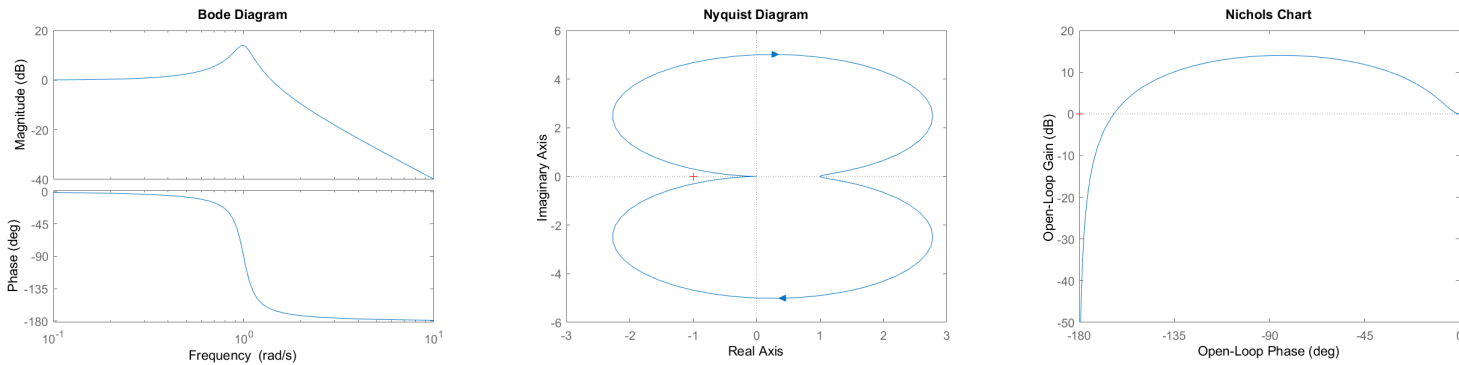


Figure 19.2: Bode, Nyquist and Nichols diagrams of $\frac{1}{s^2+0.2s+1}$.

19.2 Stability margins

Finding crossover frequencies (remember Definitions 10.9 and 10.10) and stability margins (review Section 18.4) in the Nichols diagram is simple:

- There is a **phase crossover frequency** when the curve crosses a vertical line at $-180^\circ \pm k360^\circ$, with $k \in \mathbb{Z}$.
- The corresponding **gain margin** is read at that vertical line: it is positive if the curve is below 0 dB, and negative if above.
- There is a **gain crossover frequency** when the curve crosses the horizontal axis (i.e. 0 dB).
- The corresponding **phase margin** is read over the horizontal axis, and is the distance to the closest vertical line at $-180^\circ \pm k360^\circ$: it is positive if we are to the right of the line, negative if to the left.

Figure 19.3 shows where crossovers are read. MATLAB plots red crosses at $x = 0$ dB, $y = -180^\circ \pm k360^\circ$, so as to make reading stability margins easier.

Example 19.3. Figure 19.4 shows the phase margins of two third order transfer functions: in one case, both are positive; in the other, both are negative. \square

19.3 The N and M curves

We will now study the reason why the Nichols diagram is used. Consider a closed loop as shown in Figure 19.5; if this is a control loop, then $G(s)$ corresponds to both the controller and the plant. Let the closed loop transfer function be $F(s) = \frac{y(s)}{r(s)}$. Also let the absolute value and phase of complex $G(j\omega)$ be

$$G(j\omega) = \underbrace{|G(j\omega)|}_G e^{j \overbrace{\angle G(j\omega)}^\theta} = Ge^{j\theta} = G \cos \theta + j G \sin \theta \quad (19.1)$$

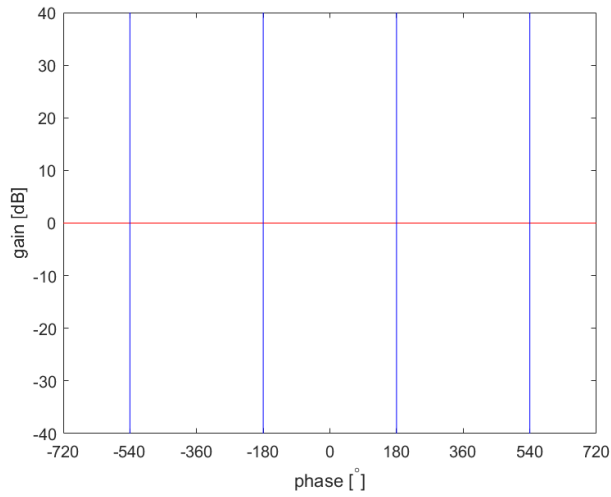


Figure 19.3: In a Nichols diagram, there is a gain crossover when the plot crosses the red line, and a phase crossover when the phase crosses any blue line.

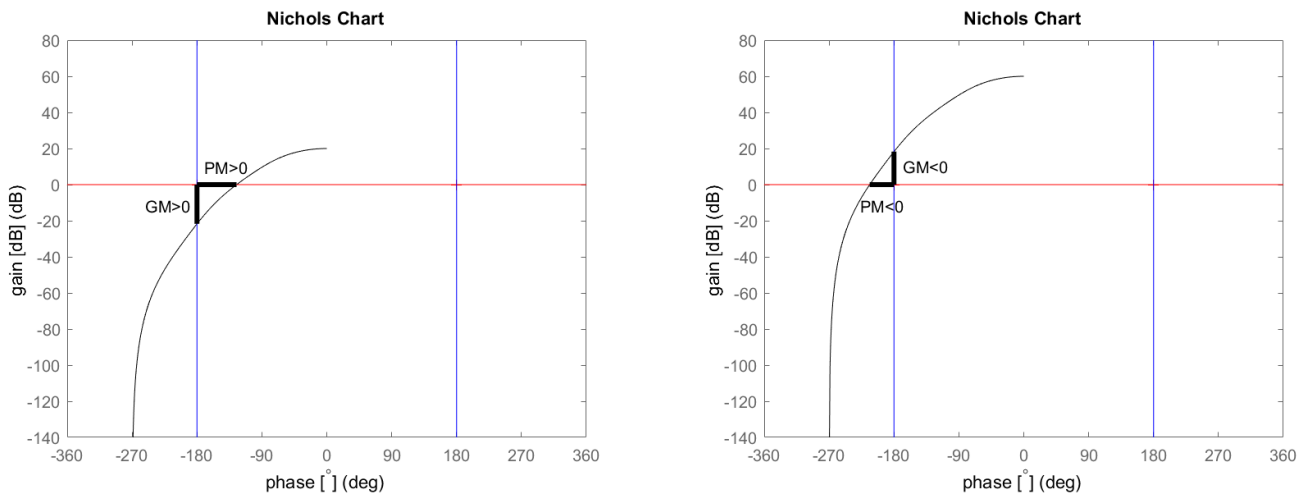


Figure 19.4: Left: Nichols diagram of $\frac{10^4}{(s+1)(s+10)(s+100)}$. Right: Nichols diagram of $\frac{10^6}{(s+1)(s+10)(s+100)}$.

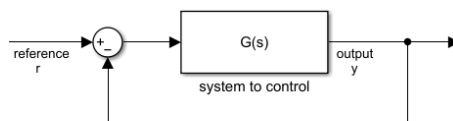


Figure 19.5: Closed loop for Section 19.3.

In other words, the Nichols diagram of $G(s)$ has G in the y -axis, and θ in the x -axis. Thus

$$F(s) = \frac{G(s)}{1 + G(s)} \quad (19.2)$$

$$\begin{aligned} |F(j\omega)| &= \left| \frac{G(j\omega)}{1 + G(j\omega)} \right| = \frac{|G(j\omega)|}{|1 + G(j\omega)|} \\ &= \frac{G}{|1 + G \cos \theta + j G \sin \theta|} = \frac{G}{\sqrt{(1 + G \cos \theta)^2 + G^2 \sin^2 \theta}} \\ &= \frac{G}{\sqrt{1 + 2G \cos \theta + G^2 \cos^2 \theta + G^2 \sin^2 \theta}} = \frac{G}{\sqrt{1 + 2G \cos \theta + G^2}} \end{aligned} \quad (19.3)$$

$$\begin{aligned} \angle F(j\omega) &= \angle \frac{G(j\omega)}{1 + G(j\omega)} = \angle G(j\omega) - \angle [1 + G(j\omega)] \\ &= \theta - \angle [1 + G \cos \theta + j G \sin \theta] = \theta - \arctan \frac{G \sin \theta}{1 + G \cos \theta} \end{aligned} \quad (19.4)$$

It is usual to plot, on Nichols diagrams, curves along which $|F(j\omega)|$ and $\angle F(j\omega)$ have constant values, i.e. level curves of (19.3)–(19.4). These curves are also known as M and N curves, respectively. In this way, looking at a plot that shows us the frequency response of an open loop, we can also see what the frequency response of the closed loop will be.

Finding analytic expressions for these M and N curves is difficult and not necessary; they can be found numerically. MATLAB plots these curves when `grid` is applied to a Nichols plot.

Example 19.4. Figure 19.6 shows the Bode diagrams of an open loop $G(s)$ and the corresponding closed loop $F(s)$, obtained with

```
s = tf('s');
G = (s-10)/(s^2+0.2*s+1);
F = feedback(G,1);
figure,bode(G,F),legend
figure,nichols(G),grid
```

The code above also plots the Nichols diagram of $G(s)$, shown in Figure #. The values of gain and phase of $F(s)$ can be seen from the M and N curves; notice in particular that:

- the gain for low frequencies is positive but not far from 0 dB, and indeed the Nichols diagram of $G(s)$ never goes beyond the 1 dB M curve;
- the gain has no resonance peak, and indeed the Nichols diagram of $G(s)$ never gets closer to the red crosses than it was at its beginning;
- the phase remains close to 0° for a while, and indeed the Nichols diagram of $G(s)$ is for a while almost parallel to an N curve of a very low phase;
- the phase goes down to -90° , and so does the Nichols diagram of $G(s)$.

□

Notice that:

- MATLAB labels the M curves with the corresponding values of gain in dB. N curves are left unlabelled, since the corresponding phase can be read in the x -axis. In fact, when open loop gain G is small, the closed loop phase $\angle F(j\omega)$ given by (19.4) becomes

$$\begin{aligned} \angle F(j\omega) &= \theta - \arctan \underbrace{\frac{G \sin \theta}{1 + G \cos \theta}}_{\approx 1} \\ &\approx \theta - \underbrace{\arctan \overbrace{G \sin \theta}^{\approx 0}}_{\approx 0} \approx \theta \end{aligned} \quad (19.5)$$

which is the open loop phase.

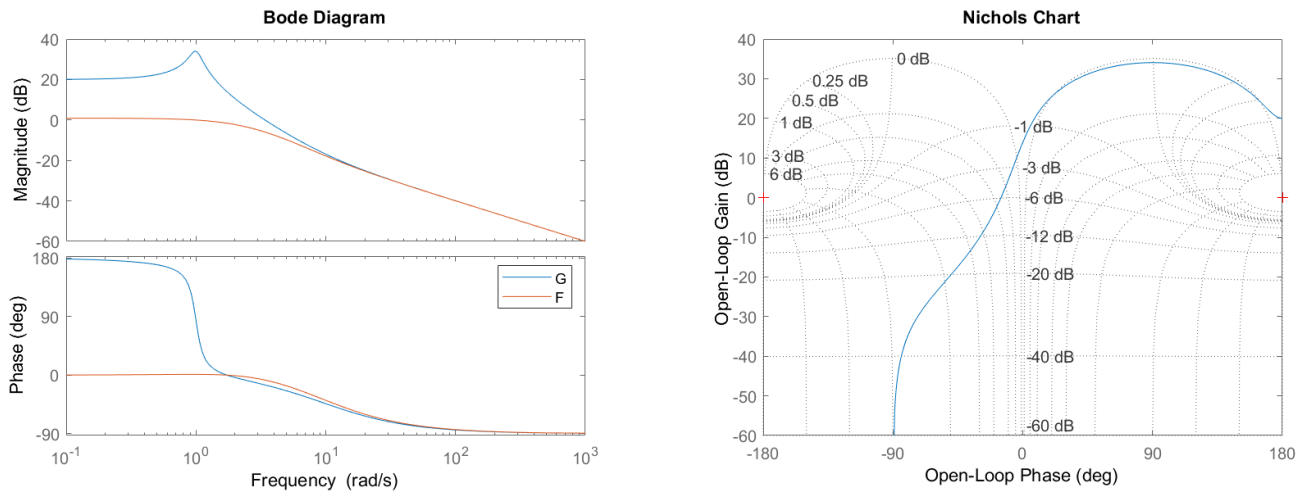


Figure 19.6: Diagrams for Example 19.4. Left: Bode diagram of open loop $G(s)$ and closed loop $F(s)$. Right: Nichols diagram of open loop $G(s)$.

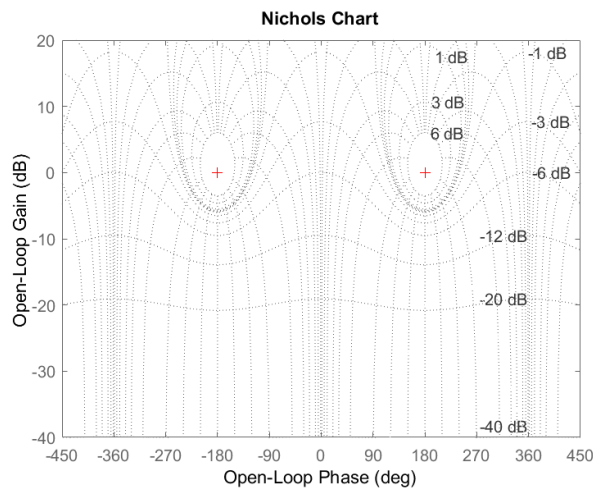


Figure 19.7: Nichols diagram showing the periodicity of the M and N curves in θ .

- M and N curves are periodic in θ , with a period of 360° , since θ only appears in (19.3)–(19.4) together with trigonometric functions, which are periodic. This periodicity is shown in Figure 19.7.
- M and N curves can also be plot in Nyquist diagrams; MATLAB does so when `grid` is applied to a Nyquist plot. However, reading the closed loop gain and phase makes more sense when the diagram shows the open loop gain and phase, not the real and imaginary parts of the open loop frequency response.
- In fact, there are controller design techniques based the Nichols diagram: the controller is designed to avoid the regions of the diagram where the M and N curves have an undesirable behaviour (e.g. a high closed loop gain). We will not study these control design techniques, but will mention one of them below in Section 43.6.

Why the Nichols diagram is used

Glossary

And the Galaadites tooke the fordes of Iordan, by the which Ephraim was to returne. And when there had come to the same one of the number of Ephraim, fleeing, and had said: I besech you let me passe: The Galaadites said to him: Art thou not an Ephraite? Who saying:

I am not: they asked him: Say then Schibboleth, which is interpreted an Eare of corne. Who answered, Sibboleth, not being able by the same letter to expresse, an eare of corne. And immediatly being apprehended they killed him in the very passage of Iordan.

Judges, xii 5–6, Douay-Rheims version (1609)

level curve curva de nivel

Exercises

1. Plot the Nichols diagram of the plant in Exercise 3 from Chapter 18.
2. Repeat Exercise 2 from Chapter 17 using the Nichols plot.
3. Repeat Exercise 3 from Chapter 17 using the Nichols plot.
4. Prove that the M and N curves are circles in the Nyquist diagram. Follow the following steps.

(a) Let the frequency response of a transfer function $G(s)$ be given by $G(j\omega) = X(\omega) + jY(\omega)$, where $X, Y \in \mathbb{R}$, and let the frequency response of the closed loop with $G(s)$ in the direct branch and unit feedback be $F(s)$. Find an expression for $F(j\omega)$ as function of X and Y .

(b) Show that

$$|F(j\omega)|^2 = \frac{X^2 + Y^2}{(1 + X)^2 + Y^2} \quad (19.6)$$

(c) Let $M = |F(j\omega)|$, and show that, if $M = 1$, the curve degenerates into a vertical straight line given by $X = -\frac{1}{2}$.

(d) Let $M \neq 1$, and rearrange terms in (19.6) to arrive at

$$\left(X + \frac{M^2}{M^2 - 1}\right)^2 + Y^2 = \frac{M^2}{(M^2 - 1)^2} \quad (19.7)$$

which is the equation of a circle with centre $\left(-\frac{M^2}{M^2 - 1}, 0\right)$ and radius $\frac{M}{M^2 - 1}$.

(e) Show that

$$\angle F(j\omega) = \arctan \frac{Y}{X} - \arctan \frac{Y}{1 + X} \quad (19.8)$$

(f) Let $N = \tan \angle F(j\omega)$, and rearrange terms in (19.8) to arrive at

$$\left(X + \frac{1}{2}\right)^2 + \left(Y - \frac{1}{2N}\right)^2 = \frac{1}{4} + \left(\frac{1}{2N}\right)^2 \quad (19.9)$$

which is the equation of a circle with centre $\left(-\frac{1}{2}, \frac{1}{2N}\right)$ and radius $\sqrt{\frac{1}{4} + \left(\frac{1}{2N}\right)^2}$.

Chapter 20

Steady-state errors

Einer der Männer feuerte auf sie. Der hellblaue Lichtblitz verfehlte sie um mehrere Meter und schlug eine Stichflamme aus der Wand, und Katt begann Haken zu schlagen und sich auf fast noch unmöglichere Weise zu bewegen. Der nächste Schuss verfehlte sie noch mehr, aber nun eröffneten auch die anderen Männer das Feuer, und Anders hatte ja bereits gesehen, was für ausgezeichnete Schützen sie waren.

Wolfgang HOHLBEIN (1953 — ...), Heike HOHLBEIN (1954 — ...), *Anders* (2004), *Die tote Stadt*, I 6

We saw in Section 10.2 what a steady-state is. Consider now a closed loop control system, as seen in Figure 20.1. As we want the output $y(t)$ to follow reference $r(t)$, we should ideally have an error $e(t) = 0, \forall t$. It of course expectable that this cannot be achieved immediately when the control system is started, or when the reference changes unexpectedly. Still, it can be desirable that

$$\lim_{t \rightarrow +\infty} e(t) = 0 \quad (20.1)$$

In other words, we are requiring that there should be no **steady-state error**. *Steady-state error e_{ss}* If this is impossible, or unnecessarily difficult, it is still of course desirable that the steady-state error should be small (what a small error is depends, of course, on the particular problem under study). If, however, the steady-state error becomes too large, or, even worse, diverges to infinity, then the closed-loop control system is not fulfilling its purpose.

Let us denote the steady-state error by e_{ss} . We know that we can find it using the final value theorem (Theorem 2.4):

$$e_{ss} = \lim_{t \rightarrow +\infty} e(t) = \lim_{s \rightarrow 0} s e(s) \quad (20.2)$$

We also know that

$$e(s) = r(s) - H(s)G(s)C(s)e(s) \Leftrightarrow e(s) = \frac{r(s)}{1 + C(s)G(s)H(s)} \quad (20.3)$$

and thus the steady-state error is given by

$$e_{ss} = \lim_{t \rightarrow +\infty} e(t) = \lim_{s \rightarrow 0} \frac{s r(s)}{1 + C(s)G(s)H(s)} \quad (20.4)$$

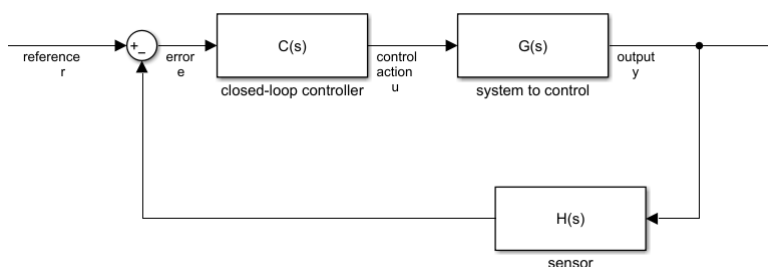


Figure 20.1: Closed-loop control system.

In this chapter we will see what this means for three different types of references,

- the step,
- the ramp, and
- the parabola.

References found in practice are usually of one of these types, or can at least be approximated in this way. We will see what consequences the results have for the choice of controller $C(s)$.

20.1 Steps as references

When the reference for the closed loop control system is a step with amplitude K , we say that e_{ss} is a **position steady-state error**. (The reason why is clear if output $y(t)$ is a position, and $r(t) = K$ is a reference for what that position should be. But we keep talking of a position steady-state error also when $y(t)$ is some other variable.) Then

$$r(t) = \begin{cases} 0, & t < 0 \\ K, & t \geq 0 \end{cases} \Rightarrow r(s) = \frac{K}{s} \quad (20.5)$$

then

$$e_{ss} = \lim_{t \rightarrow +\infty} e(t) = \lim_{s \rightarrow 0} \frac{s \frac{K}{s}}{1 + C(s)G(s)H(s)} = \frac{K}{1 + \lim_{s \rightarrow 0} C(s)G(s)H(s)} \quad (20.6)$$

We see that there are two different cases, depending on product $C(s)G(s)H(s)$, which we call the open-loop transfer function. Remembering Definition 11.3, we conclude the following:

- If $C(s)G(s)H(s)$ is of type 0, then, irrespective of its number of zeros,

$$\lim_{s \rightarrow 0} C(s)G(s)H(s) = \lim_{s \rightarrow 0} \frac{\overbrace{N(s)}^{\text{the number of zeros is irrelevant}}}{\underbrace{(s+p_1)(s+p_2)(s+p_3)\dots}_{\text{no poles at the origin: } p_1, p_2, p_3 \dots \neq 0}} = K_p \quad (20.7)$$

and thus

$$e_{ss} = \frac{K}{1 + \lim_{s \rightarrow 0} C(s)G(s)H(s)} = \frac{K}{1 + K_p} \quad (20.8)$$

- If $C(s)G(s)H(s)$ is of type 1, then numerator $N(s)$ has no zeros at the origin, because otherwise the pole at the origin would have been cancelled. So, irrespective of its number of zeros,

$$\lim_{s \rightarrow 0} C(s)G(s)H(s) = \lim_{s \rightarrow 0} \frac{\overbrace{N(s)}^{\text{the number of zeros is irrelevant}}}{\underbrace{s(s+p_1)(s+p_2)(s+p_3)\dots}_{\text{one pole at the origin: } p_1, p_2, p_3 \dots \neq 0}} = +\infty \quad (20.9)$$

and thus

$$e_{ss} = \frac{K}{1 + \lim_{s \rightarrow 0} C(s)G(s)H(s)} = 0 \quad (20.10)$$

- Likewise, if $C(s)G(s)H(s)$ is of type 2 or more, then, irrespective of its number of zeros (which cannot be at the origin),

$$\lim_{s \rightarrow 0} C(s)G(s)H(s) = \lim_{s \rightarrow 0} \frac{N(s)}{\underbrace{s^n(s+p_1)(s+p_2)(s+p_3)\dots}_{n \geq 2 \text{ poles at the origin}}} = +\infty \quad (20.11)$$

and thus $e_{ss} = 0$ just the same.

In other words, we need at least one pole at the origin in the open loop to ensure no steady state error. A good sensor should verify $H(s) \approx 1$, and have no poles at the origin; indeed, it makes little sense that the sensor should integrate $y(t)$. So,

plant is of type 0, the controller should be of type ensure no steady state

- if plant $G(s)$ has one pole at the origin, or more, e_{ss} will be zero;
- if plant $G(s)$ has no poles at the origin, controller $C(s)$ should have (at least) one pole at the origin so that e_{ss} will be zero (it can be e.g. a PI or PID controller);
- if neither plant $G(s)$ nor controller $C(s)$ have poles at the origin, there will be a constant steady state error.

Example 20.1. Figure 20.2 shows the unit step responses of a second order, type 0 system, controlled by a proportional (thus, type 0) controller and by a PI (thus, type 1) controller. The plots were drawn as follows:

```
% plant: type 0
s = tf('s');
G = 1 / (s^2 + 0.5*s + 1);
t = 0 : 0.1 : 120;
r = ones(size(t));
% controller: type 0
C = 1; F = feedback(C*G, 1);
y = lsim(F, r, t);
figure, plot(t,r, t,y), legend({'reference','output'}), ylim([0 1.1])
xlabel('time [s]'), ylabel('reference and output')
% controller: type 1
C = 1 + 0.1/s; F = feedback(C*G, 1);
y = lsim(F, r, t);
figure, plot(t,r, t,y), legend({'reference','output'}), ylim([0 1.1])
xlabel('time [s]'), ylabel('reference and output')
```

The steady state error e_{ss} for the proportional controller is, as expected,

$$K_p = \lim_{s \rightarrow 0} 1 \times \frac{1}{s^2 + 0.5s + 1} = 1 \quad (20.12)$$

$$e_{ss} = \frac{1}{1 + K_p} = \frac{1}{2} \quad (20.13)$$

A 50% steady state error is likely to be unacceptable in all situations. Suppose we could accept, at most, a 5% steady state error. We could, of course, use the PI controller, and get a 0% error, or settle for a proportional controller P given by

$$e_{ss} = 0.05 = \frac{1}{1 + K_p} \Leftrightarrow 1 + K_p = 20 \Leftrightarrow K_p = 19 \quad (20.14)$$

$$\Rightarrow K_p = 19 = \lim_{s \rightarrow 0} P \times \frac{1}{s^2 + 0.5s + 1} \Leftrightarrow P = 19 \quad (20.15)$$

While with this controller $P = 19$ the steady state error will now be low enough, as you can check, there will be a very high overshoot (and you can see why if you find where closed loop poles went to). That is the price to pay for keeping a simple controller (proportional, in this case). We will learn more about PID tuning in the next chapter. \square

20.2 Ramps as references

When the reference for the closed loop control system is a ramp with slope K , we say that e_{ss} is a **velocity steady-state error**. (Once more, the reason why is clear if output $y(t)$ is a position, and reference $r(t) = Kt$ corresponds to a constant velocity. Again, we keep talking of a velocity steady-state error even if the output is not a position.) Then

$$r(t) = \begin{cases} 0, & t < 0 \\ Kt, & t \geq 0 \end{cases} \Rightarrow r(s) = \frac{K}{s^2} \quad (20.16)$$

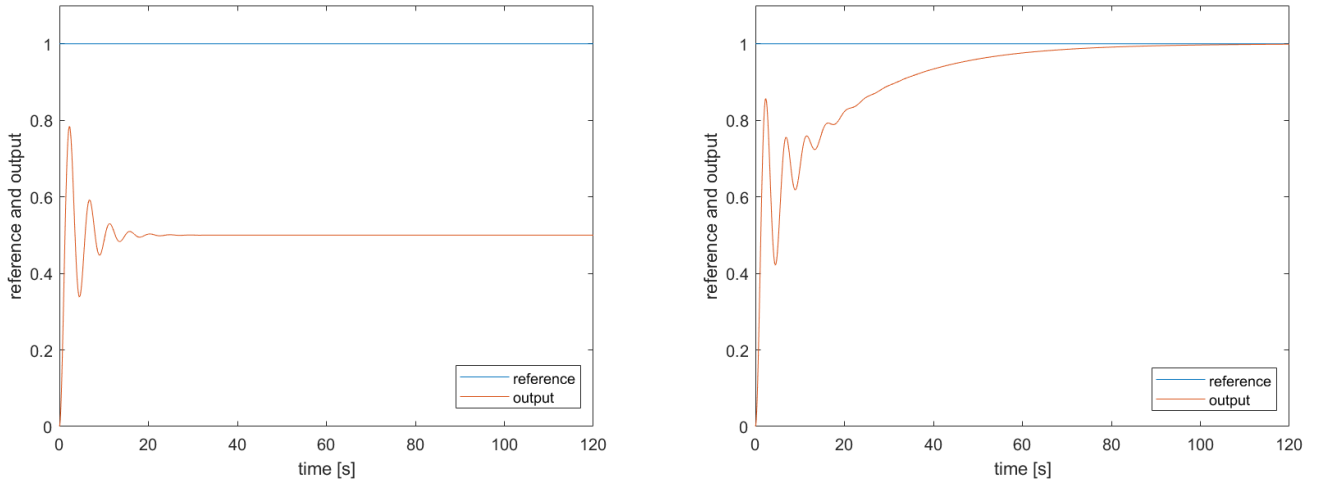


Figure 20.2: Unit step responses of the closed loop control of $\frac{1}{s^2+0.5s+1}$ from Example 20.1. Left: proportional control $P = 1$. Right: PI control $C(s) = 1 + \frac{0.1}{s}$.

then

$$\begin{aligned} e_{ss} &= \lim_{t \rightarrow +\infty} e(t) = \lim_{s \rightarrow 0} \frac{s \frac{K}{s^2}}{1 + C(s)G(s)H(s)} \\ &= \lim_{s \rightarrow 0} \frac{K}{s + sC(s)G(s)H(s)} = \frac{K}{\lim_{s \rightarrow 0} sC(s)G(s)H(s)} \end{aligned} \quad (20.17)$$

We see that there are now three different cases, depending on the open-loop transfer function:

- If $C(s)G(s)H(s)$ is of type 0, then, irrespective of its number of zeros,

$$\lim_{s \rightarrow 0} sC(s)G(s)H(s) = \lim_{s \rightarrow 0} \frac{\overbrace{s N(s)}^{\text{the number of zeros is irrelevant}}}{\underbrace{(s + p_1)(s + p_2)(s + p_3) \dots}_{\text{no poles at the origin: } p_1, p_2, p_3 \dots \neq 0}} = 0 \quad (20.18)$$

and thus

$$e_{ss} = \frac{K}{\lim_{s \rightarrow 0} sC(s)G(s)H(s)} = +\infty \quad (20.19)$$

- If $C(s)G(s)H(s)$ is of type 1, then numerator $N(s)$ has no zeros at the origin, because otherwise the pole at the origin would have been cancelled. So, irrespective of its number of zeros,

$$\begin{aligned} \lim_{s \rightarrow 0} sC(s)G(s)H(s) &= \lim_{s \rightarrow 0} \frac{\overbrace{s N(s)}^{\text{the number of zeros is irrelevant}}}{\underbrace{s(s + p_1)(s + p_2)(s + p_3) \dots}_{\text{one pole at the origin: } p_1, p_2, p_3 \dots \neq 0}} \\ &= \frac{N(0)}{p_1 p_2 p_3 \dots} = K_v \neq 0 \end{aligned} \quad (20.20)$$

and thus

$$e_{ss} = \frac{K}{\lim_{s \rightarrow 0} sC(s)G(s)H(s)} = \frac{K}{K_v} \quad (20.21)$$

- If $C(s)G(s)H(s)$ is of type 2, then, irrespective of its number of zeros (which cannot be at the origin),

$$\lim_{s \rightarrow 0} sC(s)G(s)H(s) = \lim_{s \rightarrow 0} \frac{s N(s)}{\underbrace{s^2(s + p_1)(s + p_2)(s + p_3) \dots}_{\text{2 poles at the origin}}} = +\infty \quad (20.22)$$

and thus

$$e_{ss} = \frac{K}{\lim_{s \rightarrow 0} sC(s)G(s)H(s)} = 0 \quad (20.23)$$

- Likewise, if $C(s)G(s)H(s)$ is of type 3 or more, then, irrespective of its number of zeros,

$$\lim_{s \rightarrow 0} sC(s)G(s)H(s) = \lim_{s \rightarrow 0} \frac{sN(s)}{\underbrace{s^n(s+p_1)(s+p_2)(s+p_3)\dots}_{n \geq 3 \text{ poles at the origin}}} = +\infty \quad (20.24)$$

and thus $e_{ss} = 0$ just the same.

In other words, we need at least two poles at the origin in the open loop to ensure no steady state error. As it makes no sense that poles at the origin are found in the sensor,

- if plant $G(s)$ has two or more poles at the origin, e_{ss} will be zero;
- if plant $G(s)$ has one pole at the origin, controller $C(s)$ should have (at least) one pole at the origin so that e_{ss} will be zero (it can be e.g. a PI controller), otherwise e_{ss} will be constant;
- if plant $G(s)$ has no poles at the origin, controller $C(s)$ should have (at least) two poles at the origin so that e_{ss} will be zero (it can be e.g. a PI² controller), while a controller with one pole at the origin will originate a constant e_{ss} ;
- if neither plant $G(s)$ nor controller $C(s)$ have poles at the origin, the steady state error will be infinite, i.e. the control system will not follow the reference.

Plant and controller should have 2 poles at the origin among them for $e_{ss} = 0$

Plant and controller with 1 poles at the origin among them have constant e_{ss}

If the plant is of type 0, the controller should be of type 2 to ensure no steady state error

No poles at the origin: a ramp reference cannot be followed

Example 20.2. Figure 20.3 shows the unit slope ramp responses of the type 0 system from Example 20.2, controlled by the proportional (thus, type 0) controller $P = 19$ and by the PI (thus, type 1) controller. The plots were drawn as follows:

```
% plant: type 0
s = tf('s');
G = 1 / (s^2 + 0.5*s + 1);
t = 0 : 0.1 : 400;
r = t;
% controller: type 0
C = 19; F = feedback(C*G, 1);
y = lsim(F, r, t);
figure, plot(t,r, t,y), legend({'reference','output'})
xlabel('time [s]'), ylabel('reference and output')
% controller: type 1
C = 1 + 0.1/s; F = feedback(C*G, 1);
y = lsim(F, r, t);
figure, plot(t,r, t,y), legend({'reference','output'})
xlabel('time [s]'), ylabel('reference and output')
```

The steady state error e_{ss} for the proportional controller keeps increasing with time, as expected, and becomes arbitrarily large. The PI controller achieves a constant steady state error, given, as expected, by

$$K_v = \lim_{s \rightarrow 0} s \left(1 + \frac{0.1}{s} \right) \frac{1}{s^2 + 0.5s + 1} = \lim_{s \rightarrow 0} \frac{s + 0.1}{s^2 + 0.5s + 1} = 0.1 \quad (20.25)$$

$$e_{ss} = \frac{1}{K_v} = 10 \quad (20.26)$$

□

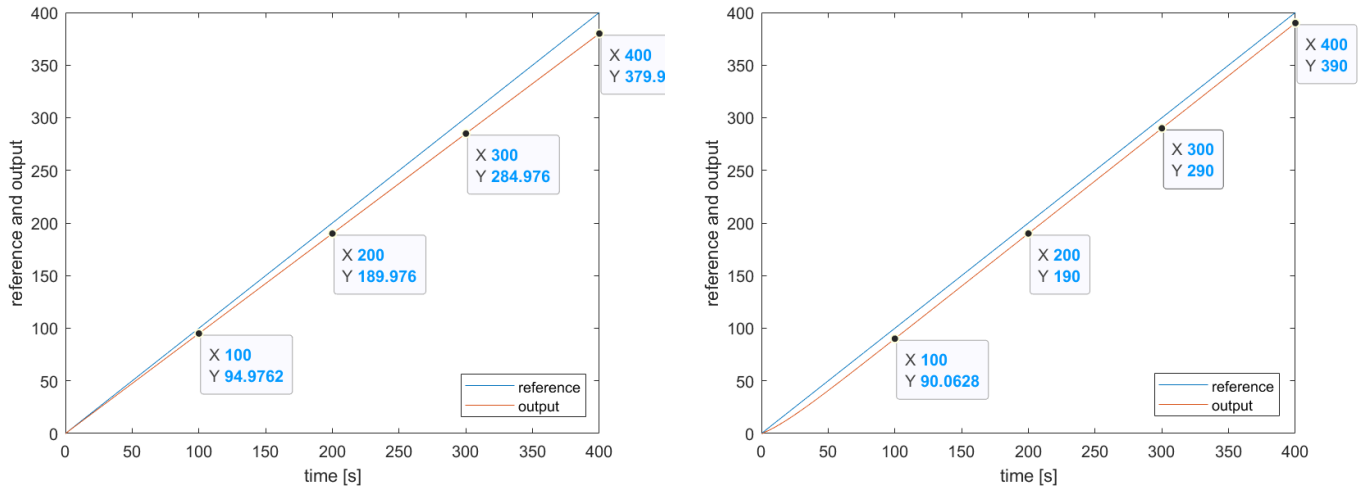


Figure 20.3: Unit slope ramp responses of the closed loop control of $\frac{1}{s^2+0.5s+1}$ from Example 20.2. Left: proportional control $P = 19$, obtained in Example 20.1. Right: PI control $C(s) = 1 + \frac{0.1}{s}$.

20.3 Parabolas as references

When the reference for the closed loop control system is a parabola given by $\frac{K}{2}t^2$, we say that e_{ss} is an **acceleration steady-state error**. (Again, this is because, if output $y(t)$ is a position, the reference corresponds to a constant acceleration.) Then

$$r(t) = \begin{cases} 0, & t < 0 \\ \frac{K}{2}t^2, & t \geq 0 \end{cases} \Rightarrow r(s) = \frac{K}{s^3} \quad (20.27)$$

then

$$\begin{aligned} e_{ss} &= \lim_{t \rightarrow +\infty} e(t) = \lim_{s \rightarrow 0} \frac{s \frac{K}{s^3}}{1 + C(s)G(s)H(s)} \\ &= \lim_{s \rightarrow 0} \frac{K}{s^2 + s^2 C(s)G(s)H(s)} = \frac{K}{\lim_{s \rightarrow 0} s^2 C(s)G(s)H(s)} \end{aligned} \quad (20.28)$$

We see that there are now three different cases, depending on the open-loop transfer function:

- If $C(s)G(s)H(s)$ is of type 0, then, irrespective of its number of zeros,

$$\lim_{s \rightarrow 0} s^2 C(s)G(s)H(s) = \lim_{s \rightarrow 0} \frac{\overbrace{s^2 N(s)}^{\text{the number of zeros is irrelevant}}}{\underbrace{(s+p_1)(s+p_2)(s+p_3)\dots}_{\text{no poles at the origin: } p_1, p_2, p_3 \dots \neq 0}} = 0 \quad (20.29)$$

and thus

$$e_{ss} = \frac{K}{\lim_{s \rightarrow 0} s C(s)G(s)H(s)} = +\infty \quad (20.30)$$

- If $C(s)G(s)H(s)$ is of type 1, then numerator $N(s)$ has no zeros at the origin, because otherwise the pole at the origin would have been cancelled. So, irrespective of its number of zeros,

$$\begin{aligned} \lim_{s \rightarrow 0} s^2 C(s)G(s)H(s) &= \lim_{s \rightarrow 0} \frac{\overbrace{s^2 N(s)}^{\text{the number of zeros is irrelevant}}}{\underbrace{s(s+p_1)(s+p_2)(s+p_3)\dots}_{\text{one pole at the origin: } p_1, p_2, p_3 \dots \neq 0}} \\ &= \frac{N(0)}{p_1 p_2 p_3 \dots} = 0 \end{aligned} \quad (20.31)$$

and thus $e_{ss} = +\infty$ just the same.

- If $C(s)G(s)H(s)$ is of type 2, then, irrespective of its number of zeros (which cannot be at the origin),

$$\begin{aligned} \lim_{s \rightarrow 0} s^2 C(s)G(s)H(s) &= \lim_{s \rightarrow 0} \frac{s^2 N(s)}{\underbrace{s^2 (s+p_1)(s+p_2)(s+p_3) \dots}_{2 \text{ poles at the origin}}} \\ &= \frac{N(0)}{p_1 p_2 p_3 \dots} = K_a \neq 0 \end{aligned} \quad (20.32)$$

and thus

$$e_{ss} = \frac{K}{\lim_{s \rightarrow 0} s C(s)G(s)H(s)} = \frac{1}{K_a} \quad (20.33)$$

- If $C(s)G(s)H(s)$ is of type 3 or more, then, irrespective of its number of zeros,

$$\lim_{s \rightarrow 0} s^2 C(s)G(s)H(s) = \lim_{s \rightarrow 0} \frac{s^2 N(s)}{\underbrace{s^n (s+p_1)(s+p_2)(s+p_3) \dots}_{n \geq 3 \text{ poles at the origin}}} = +\infty \quad (20.34)$$

and thus

$$e_{ss} = \frac{K}{\lim_{s \rightarrow 0} s C(s)G(s)H(s)} = 0 \quad (20.35)$$

In other words, we need at least three poles at the origin in the open loop to ensure no steady state error. As it makes no sense that poles at the origin are found in the sensor,

- if plant $G(s)$ has three or more poles at the origin, e_{ss} will be zero;
- if plant $G(s)$ has two poles at the origin, controller $C(s)$ should have (at least) one pole at the origin so that e_{ss} will be zero (it can be e.g. a PI controller), while a constant controller will originate a constant e_{ss} ;
- if plant $G(s)$ has one pole at the origin, controller $C(s)$ should have (at least) two poles at the origin so that e_{ss} will be zero (it can be e.g. a PI² controller), while a controller with one pole at the origin will originate a constant e_{ss} ;
- if plant $G(s)$ has no poles at the origin, controller $C(s)$ should have (at least) three poles at the origin so that e_{ss} will be zero, while a controller with two poles at the origin will originate a constant e_{ss} ;
- if plant $G(s)$ and controller $C(s)$ together have no poles at the origin, or only one, the steady state error will be infinite, i.e. the control system will not follow the reference.

Plant and controller should have 3 poles at the origin among them for $e_{ss} = 0$

Plant and controller with 2 poles at the origin among them have constant e_{ss}

Less than two poles at the origin: a parabolic reference cannot be followed

20.4 Summing up the results

Table 20.1 sums up the results in this chapter.

Glossary

At the next peg the Queen turned again, and this time she said, “Speak in French when you can’t think of the English for a thing—turn out your toes as you walk—and remember who you are!”

Lewis CARROLL (1832 — †1898), *Through the Looking-Glass, and what Alice found there* (1871), II

acceleration steady state error erro estacionário de aceleração

position steady state error erro estacionário de posição

velocity steady state error erro estacionário de velocidade

Table 20.1: Steady state errors of a closed loop control system (see Figure 20.1).

reference for the output when $t \geq 0$	type of the open loop $C(s)G(s)H(s)$				constant given by
	type 0	type 1	type 2	type 3 or higher	
step $e(t) = K$	$\frac{K}{1 + K_p}$	0	0	0	$K_p = \lim_{s \rightarrow 0} C(s)G(s)H(s)$
ramp $e(t) = Kt$	$+\infty$	$\frac{K}{K_v}$	0	0	$K_v = \lim_{s \rightarrow 0} sC(s)G(s)H(s)$
parabola $e(t) = \frac{K}{2}t^2$	$+\infty$	$+\infty$	$\frac{K}{K_a}$	0	$K_a = \lim_{s \rightarrow 0} s^2C(s)G(s)H(s)$

Exercises

1. Consider a unit feedback control system with transfer function in the direct loop

$$G(s) = 10 \frac{s + 2}{(s + 3)(s + 5)}. \quad (20.36)$$

For a unit step input signal $u(t) = 1$, $t \geq 0$, what is the steady-state error e_{ss} of the time response?

2. Consider a closed-loop system with unit feedback and transfer function in the direct loop

$$G(s) = \frac{Y(s)}{E(s)} = \frac{8}{(s + 2)(s + 4)}. \quad (20.37)$$

- (a) For the closed loop's unit step response, find:

- i. the peak time t_p
- ii. the maximum overshoot $y(t_p)$
- iii. the rise time t_r
- iv. the settling time t_s

- (b) For a unit ramp input signal $r(t) = t$, $t \geq 0$, what is this plant's steady-state error $e_{ss} = \lim_{t \rightarrow +\infty} r(t) - y(t)$?

3. Consider a system with transfer function $Y(s)/U(s) = G(s)$. The output time response $y(t)$ in closed loop for a step input in the reference $r(t) = 10$, $t \geq 0$, exhibits the steady-state error $e_{ss} = 1$ shown in Figure 20.4. For $m \leq n$, which of the following statements is true?

- A) $G(s) = 10 \frac{1 + b_1s + \cdots + b_ms^m}{1 + a_1s + \cdots + a_ns^n}$
- B) $G(s) = 9 \frac{1 + b_1s + \cdots + b_ms^m}{1 + a_1s + \cdots + a_ns^n}$
- C) $G(s) = 10 \frac{1 + b_1s + \cdots + b_ms^m}{s(1 + a_1s + \cdots + a_ns^n)}$
- D) $G(s) = 9 \frac{1 + b_1s + \cdots + b_ms^m}{s(1 + a_1s + \cdots + a_ns^n)}$

4. Figure 20.5 shows a block diagram that models the pen in a plotter and the corresponding control system, and Figure 20.6 shows its unit-step response.

- (a) Find from the step-response the system's damping coefficient ξ and natural frequency ω_n .
- (b) We want to improve the time response $y(t)$, so as to have an overshoot $M_p \leq 5\%$ and a 2% settling time $t_s < 0.1$ s. Find, if possible, a value of gain K for these specifications.

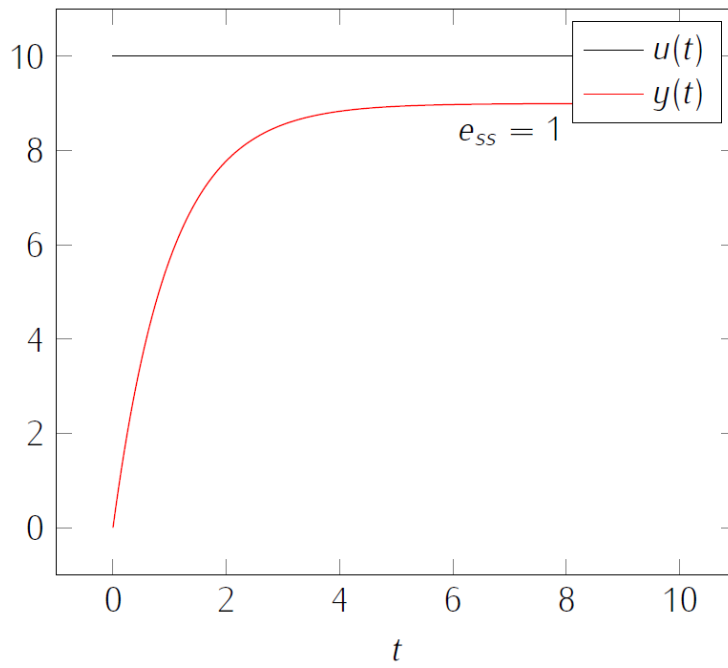


Figure 20.4: Step response from Exercise 3.

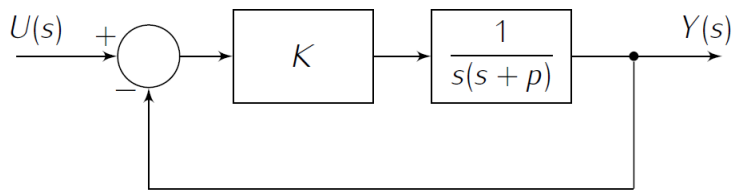


Figure 20.5: Model of a pen in a plotter printer from Exercise 4.

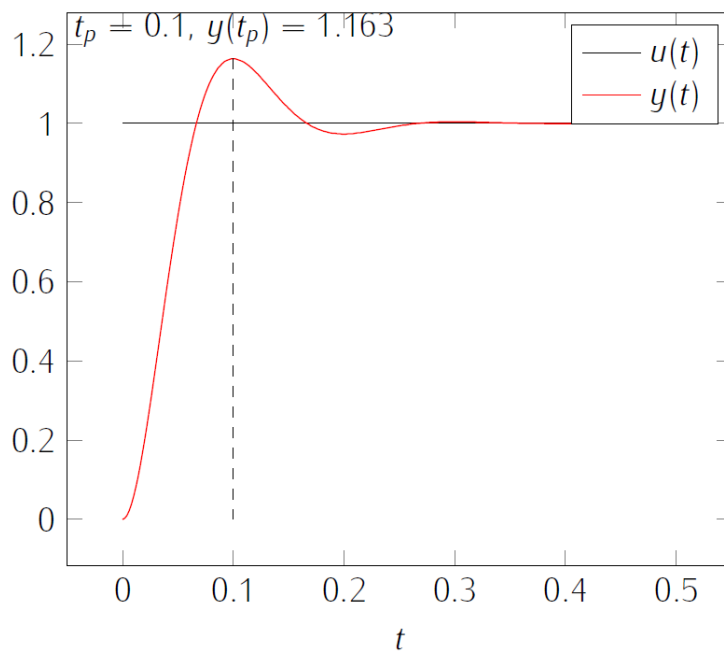


Figure 20.6: Step response of the model of a pen in a plotter printer from Exercise 4.

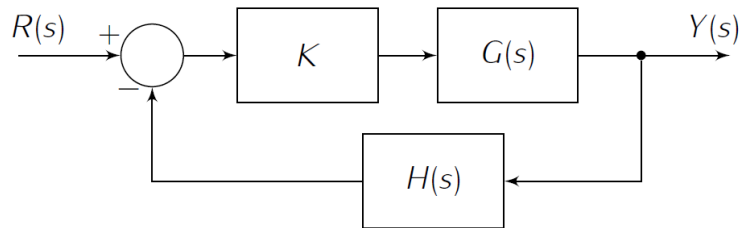


Figure 20.7: Control loop from Exercise 5.

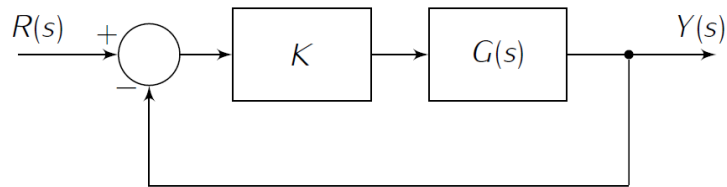


Figure 20.8: Control loop from Exercise 6.

5. In the closed-loop system of Figure 20.7,

$$G(s) = \frac{s+2}{s(s+1)(s+4)} \quad (20.38)$$

$$H(s) = 1 \quad (20.39)$$

$$K > 0 \quad (20.40)$$

- (a) Find the range of values of K for which the closed-loop is stable.
 - (b) Find K so that the the steady-state error for a unit ramp input is 10%.
6. Consider the plant in Figure 20.8, where

$$G(s) = \frac{1}{(s+1)(s+3)(s+8)}, \quad K > 0. \quad (20.41)$$

- (a) Plot its root-locus.
- (b) For which values of K is the system stable?
- (c) Can you find a 5% steady-state error e_{ss} for a unit-step input, solely by changing gain K ?
- (d) Plot the Bode diagram of $G(s)$ for $K = 240$.

Chapter 21

Design of PID controllers

In this paper, the three principal control effects found in present controllers are examined and practical names and units of measurement are proposed for each effect. (...) Formulas are given which enable the controller settings to be determined from the experimental or calculated values of the lag and unit reaction rate of the process to be controlled. These units form the basis of a quick method for adjusting a controller on the job. The effect of varying each controller setting is shown in a series of chart records. It is believed that the conceptions of control presented in this paper will be of assistance in the adjustment of existing controller applications and in the design of new installations.

J. G. ZIEGLER (1909 — †1997), N. B. NICHOLS (1914 — †1997), *Optimum settings for automatic controllers*, Transactions of the ASME, Nov. 1942

In this chapter we address design methods for controllers of the PID family. A generic PID controller corresponds to a transfer function usually written in one of three ways:

$$C(s) = P + \frac{I}{s} + Ds \quad (21.1)$$

$$C(s) = k_p \left(1 + \frac{1}{T_i s} + T_d s \right) \quad (21.2)$$

$$C(s) = K_P \underbrace{\left(1 + \frac{1}{T_I s} \right)}_{\text{integral part}} \underbrace{(1 + T_D s)}_{\text{derivative part}} \quad (21.3)$$

(21.1) is the same as (15.31), but this notation avoids confusions with (21.2) and (21.3). It is of course possible to rewrite a PID given in one of these three forms in either of the other two (see Exercise 1).

21.1 Root locus and Bode diagrams

The effects of the three components of the control action can be studied with the tools we now have:

- In Chapter 20 we learned how we could see if the integral part eliminates or reduces steady state errors.
- The derivative part adds a zero to the controller. We know from the root locus, studied in Chapter 16, that open loop zeros pull closed loop poles. The derivative part is used to pull the root locus to the zones of the plane corresponding to a fast response without excessive overshoot.
- The effect of the proportional part can also be studied with the root locus.

A PID controller has

- a pole at the origin,

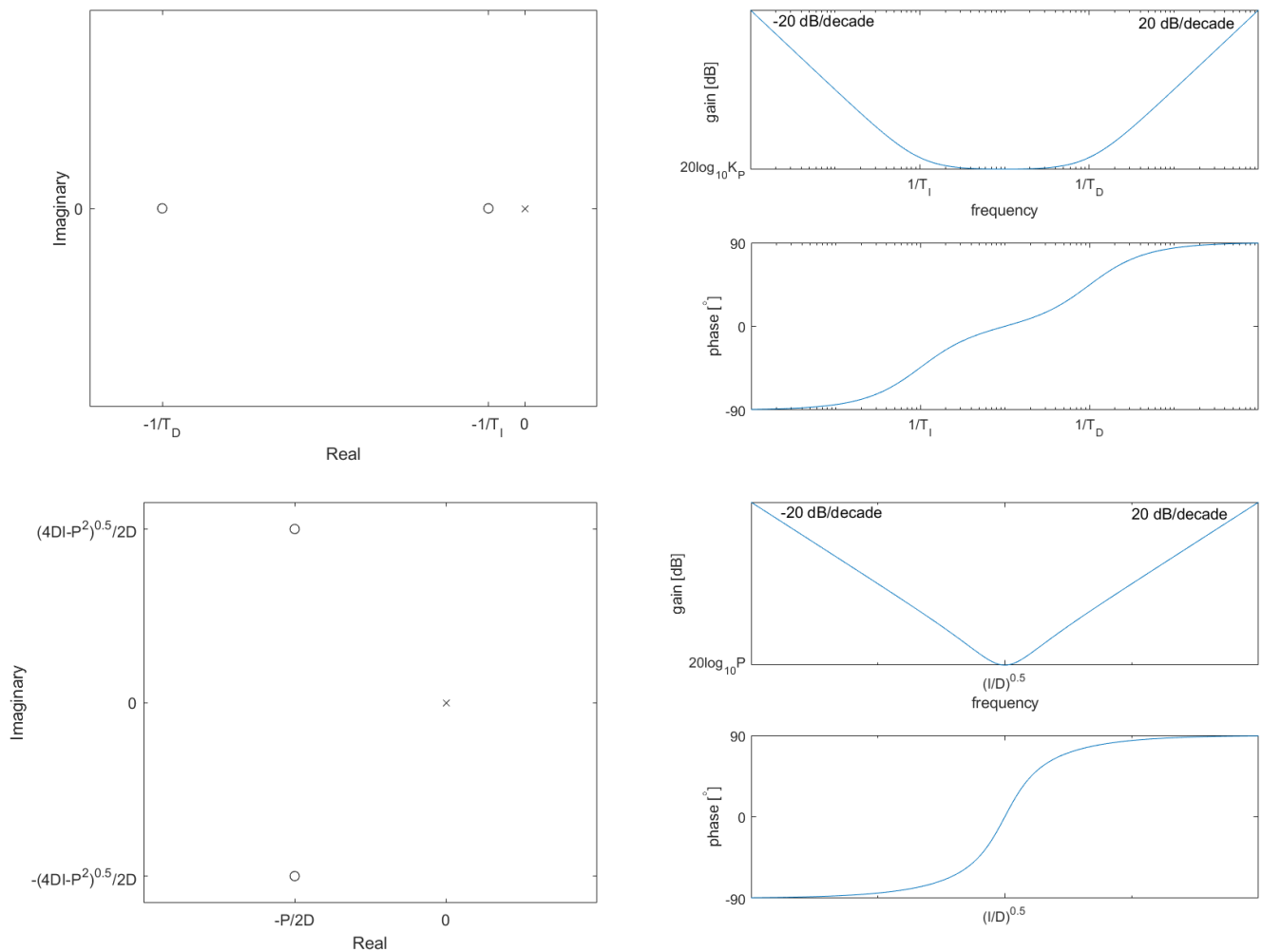


Figure 21.1: Poles and zeros in the complex plane and Bode diagram of a PID. Top: real zeros. The PID is given by (21.3) with $T_I > T_D$. If $T_D > T_I$, the zeros switch, and the lowest value of the gain is $20\log_{10} \frac{K_P T_D}{T_I}$. Bottom: complex conjugate zeros. The PID is given by (21.1), with $P^2 - 4DI < 0$. The steepness of the transition from a negative to a positive slope depends on how large the imaginary part of the zeros is (remember the resonance peak in Figure 11.19).

- two zeros, which may be real or complex conjugate. These zeros have negative real parts (it makes no sense at all to have non-minimum phase zeros, since they would pull closed loop poles to the right complex half plane, where they are unstable).

Figure 21.1 shows the location of the poles and zeros on the complex plane of a PID, and its Bode diagram. Notice that:

- At low frequencies, $\omega \approx 0$, the PID frequency response can be approximated by its integral part:

$$C(j\omega) = P + \frac{I}{j\omega} + Dj\omega \approx \frac{I}{j\omega} \quad (21.4)$$

It will thus have a -90° phase and a linear decreasing gain with a -20 dB/decade slope.

- At high frequencies, $\omega \rightarrow +\infty$, the PID frequency response can be approximated by its derivative part:

$$C(j\omega) = P + \frac{I}{j\omega} + Dj\omega \approx Dj\omega \quad (21.5)$$

It will thus have a $+90^\circ$ phase and a linear increasing gain with a $+20$ dB/decade slope.

Figure 21.2 shows the same for PI and PD controllers, that, lacking one of the control action components, can be easily obtained from the more general PID case.

Remember that, as noticed in Chapter 15, neither of these controllers are strictly proper. It is obviously impossible to have an ever increasing gain for high frequencies, as in Figure 21.1 for the PID or in Figure 21.2 for the PD. Even the not decreasing gain for high frequencies of the PD in Figure 21.2 cannot exist in reality (as we saw in Chapter 11). We will see in Chapter 29 how additional poles are used with PID control, when we will address implementation questions.

21.2 Tuning rules

The simplest way to design a controller is to use a **tuning rule**. Tuning rules provide the controller's parameters using simple calculations based on some few characteristics of the plant to be controlled; a complete model is not necessary. Tuning rules have been found heuristically (by trial and error after a long experience with many cases) or analytically (from calculations that apply to plants with some characteristics usually found in practice). While controllers obtained with all design methods can and should be fine tuned, tuning rules, because of their simplicity, are particularly apt to provide controller parameters that have to be fine tuned — i.e. slight variations of the parameters, by trial and error, should be carried out, checking if performance can still be improved.

While there are many tuning rules for controllers of the PID family, we will only study the two oldest rules (which were found heuristically, but can be justified analytically), which are among the simplest, but also among the most effective.

The Ziegler-Nichols reaction curve method can be applied to plants that have a step response shaped like an S, as seen in Figure 21.3. As you know by now, such a step response is found in plants which

- have all poles and zeros real and negative (and are thus stable, minimum-phase, and not underdamped),
- are at least of second order,
- have a number of poles which is at least equal to the number of zeros plus two (otherwise the response would begin with a slope, and have no inflection point),

What a tuning rule is

Fine tuning

Ziegler-Nichols reaction curve method
When there is an S-shaped step response

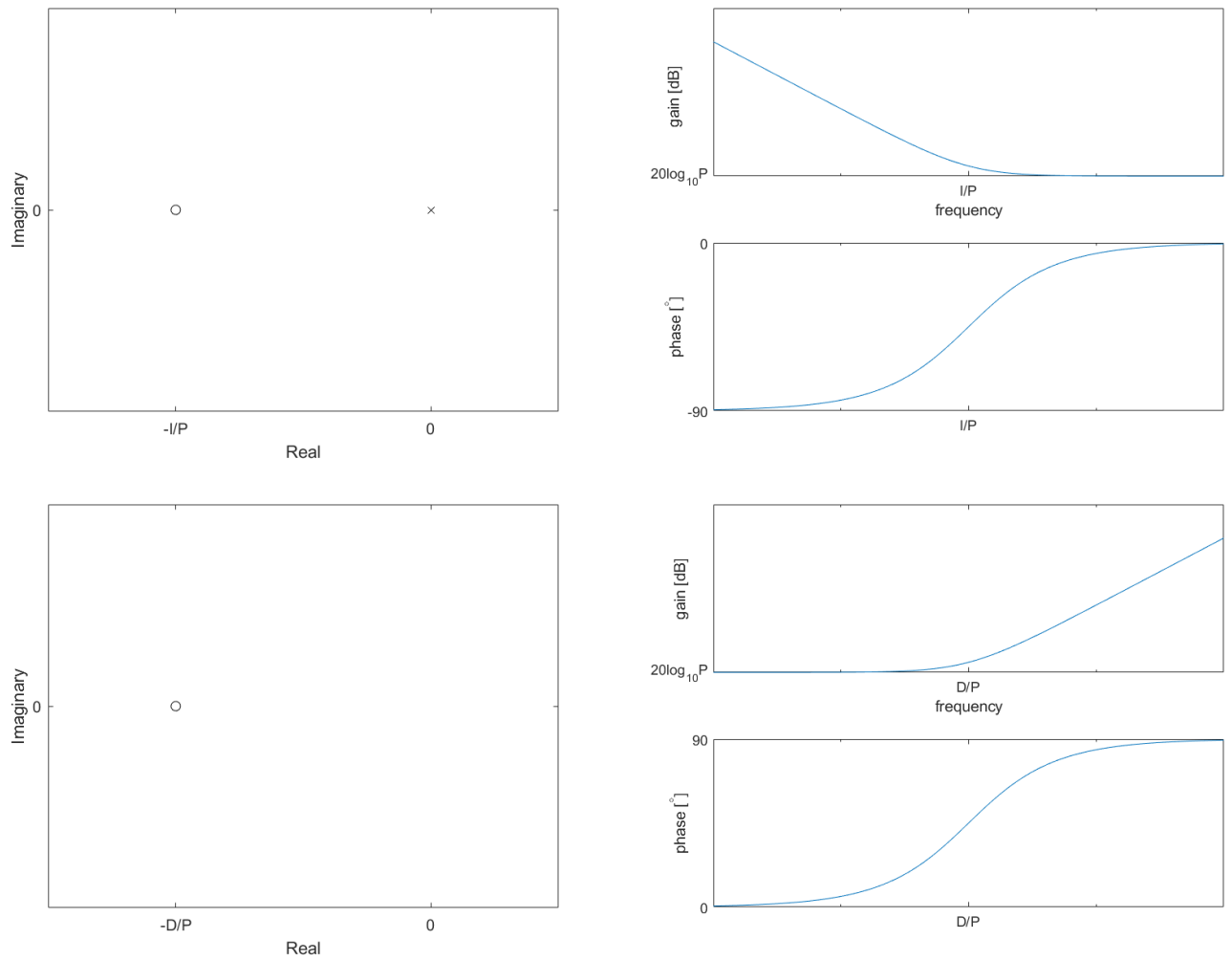


Figure 21.2: Poles and zeros in the complex plane and Bode diagram of a PI (top) and of a PD (bottom).

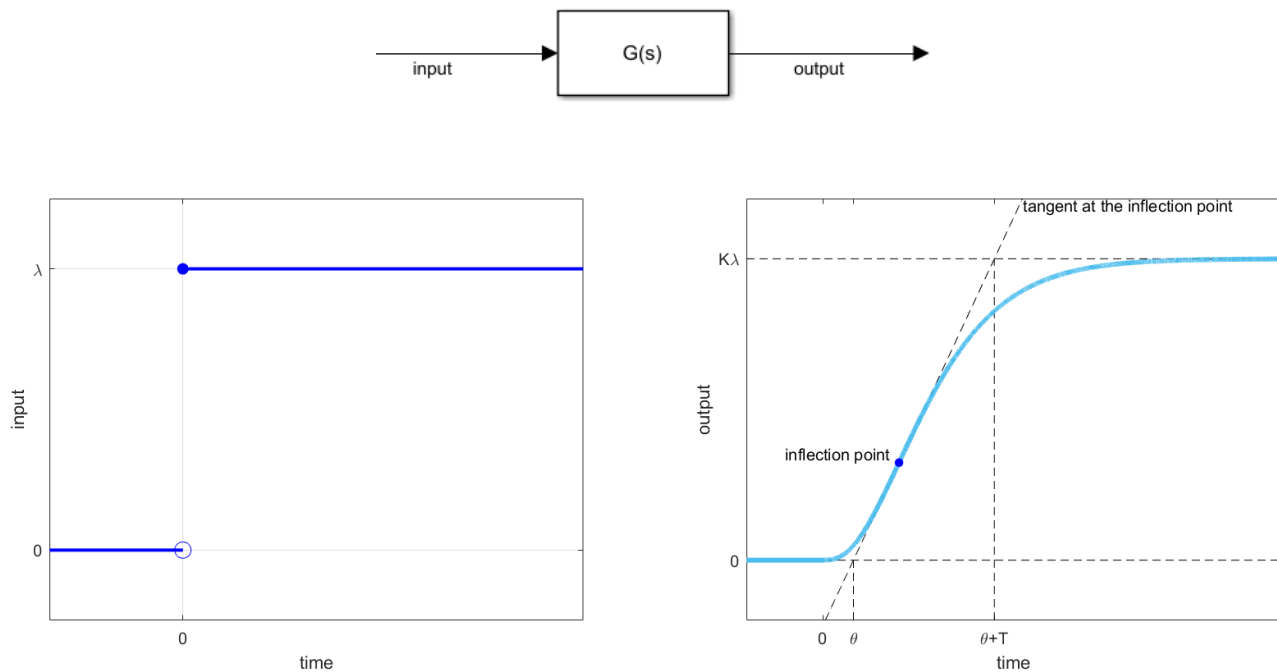


Figure 21.3: S-shaped step response of a plant $G(s)$ (in open loop), for the application of the Ziegler-Nichols reaction curve method.

Table 21.1: PID parameters according to the Ziegler-Nichols reaction curve method. See Figure 21.3

Type of controller	K_p	T_i	T_d
P	$\frac{T}{\theta K}$	—	—
PI	$\frac{0.9T}{\theta K}$	$\frac{\theta}{0.3}$	—
PID	$\frac{1.2T}{\theta K}$	2θ	0.5θ

though it is possible to find transfer functions with all the characteristics above which do *not* have an S-shaped step response. If the plant has a step response which is not really S-shaped, but is something close, you may still apply the rule, though, of course, results can be expected to be poorer. Some of the most common cases that may still work with this rule, illustrated in Figure 21.4, are step responses:

- slightly underdamped,
- which have no inflection point (e.g. first order plants), but have a delay, i.e. the step response does not begin at once when the input is applied, but only after some period of time, called delay (we will study delayed plants below in Chapter 24),
- of non-minimum phase plants.

When this rule can also be applied

The rule cannot be applied to plants with step responses which are clearly not S-shaped. Usual cases, illustrated in Figure 21.5, include:

- clearly underdamped step responses,
- responses with no inflection point (and no delay), such as first order transfer functions, or those with a number of poles that exceeds the number of zeros by only one.

When this rule cannot be applied

Once the characteristics of the step response shown in Figure 21.3 are known, controller parameters are found for (21.2) as shown in Table 21.1.

Notice that:

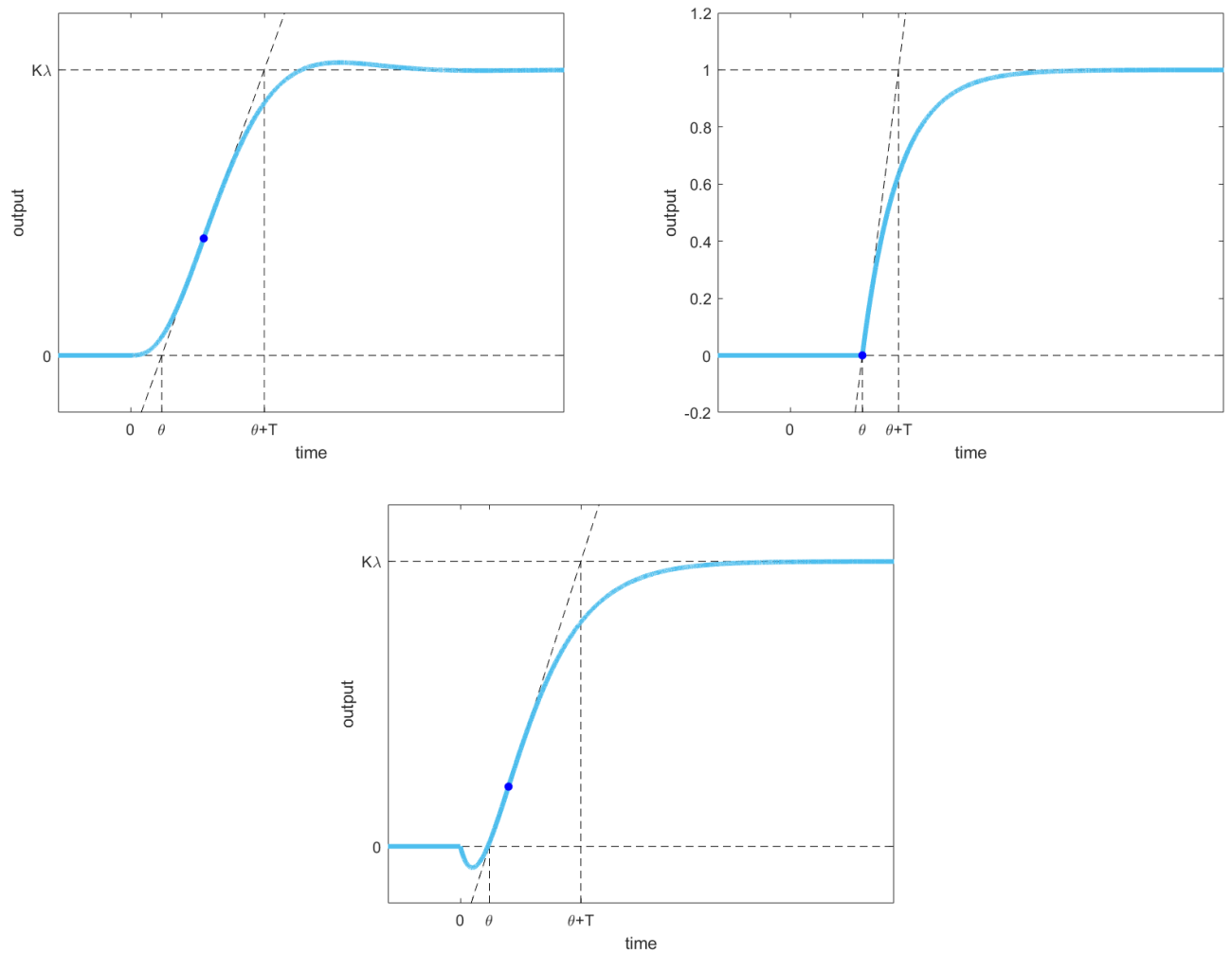


Figure 21.4: Step responses which are not S-shaped but are still close enough so that the Ziegler-Nichols reaction curve method can be applied. The farther the step response is from an S-shape, the poorer the results are likely to be.

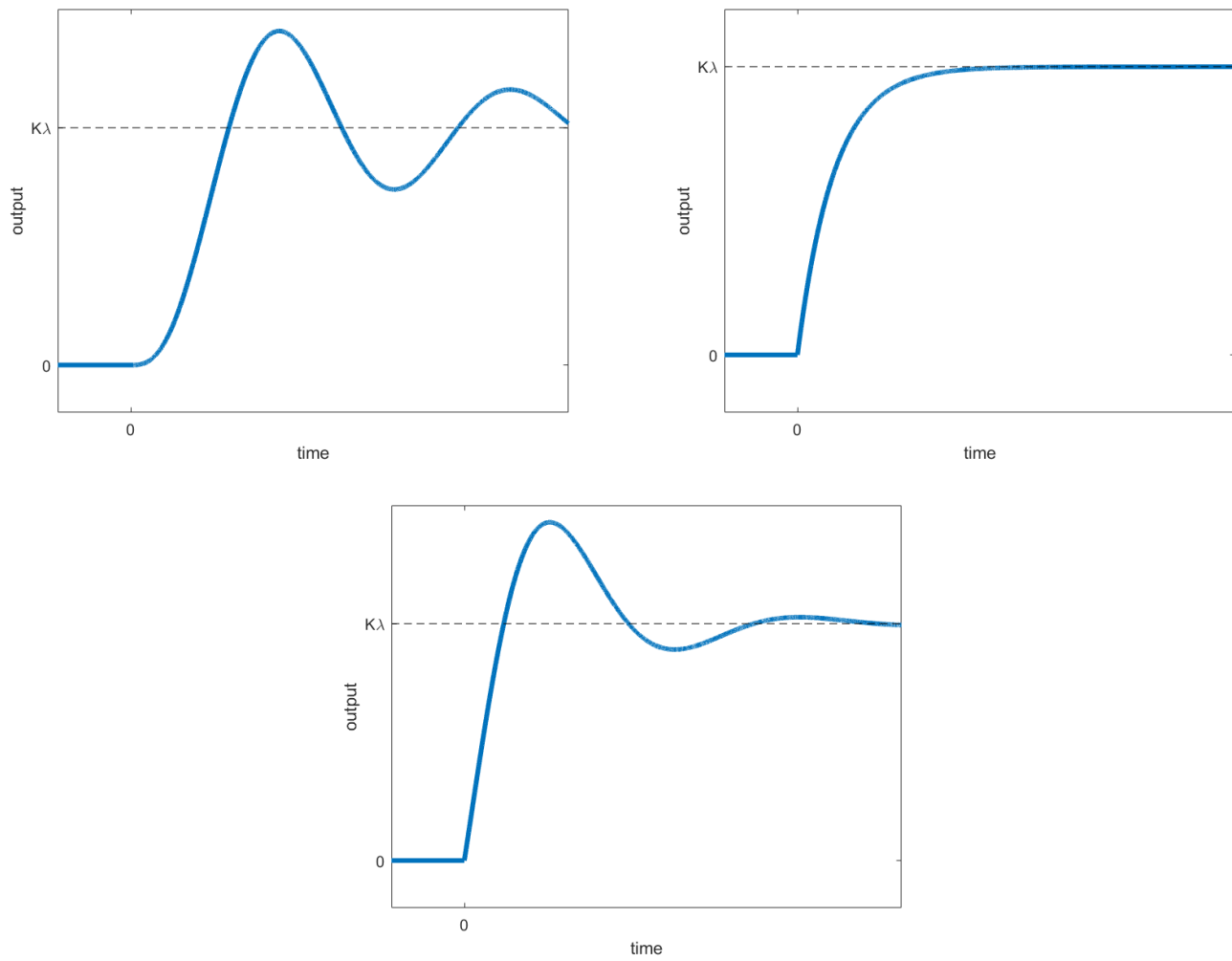


Figure 21.5: Step responses which are not S-shaped and for which the Ziegler-Nichols reaction curve method cannot be applied.

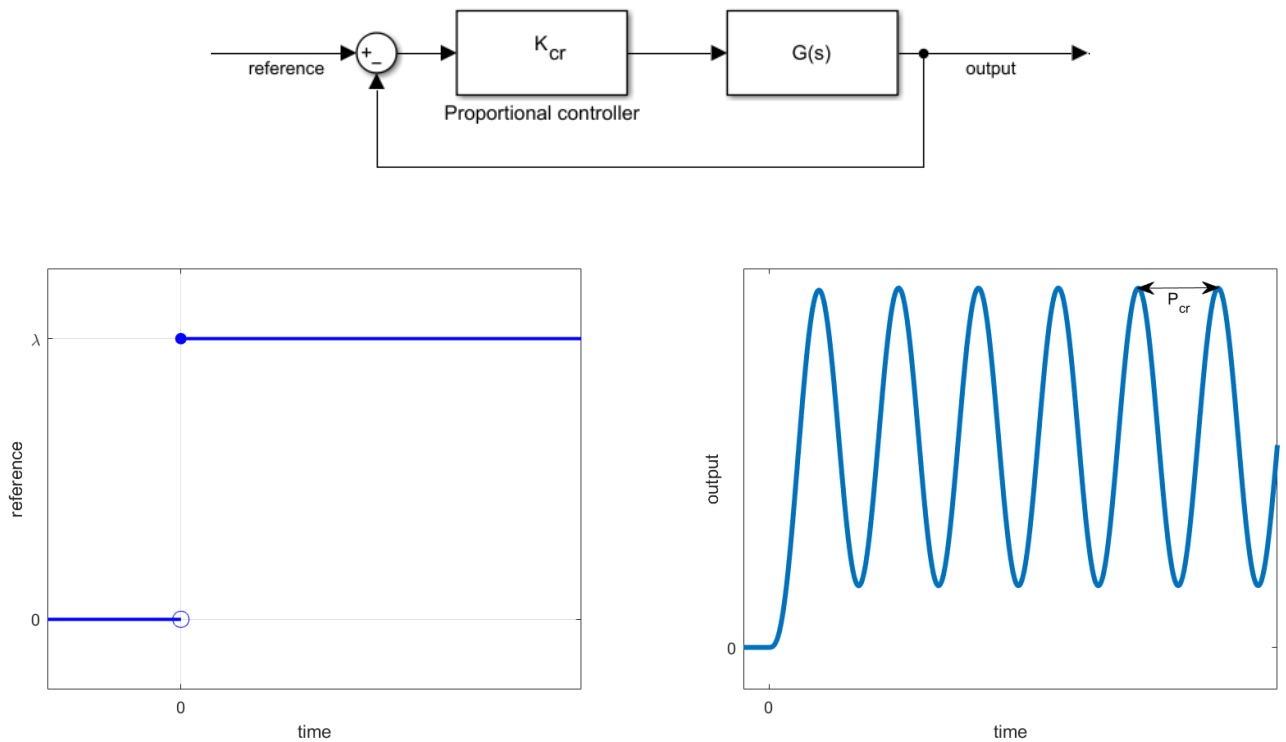


Figure 21.6: Marginally stable closed loop with proportional control, with critical oscillations in the step response, for the application of the Ziegler-Nichols critical gain method.

- The step response can be obtained experimentally. This is possible for most plants.
- The step response can also be obtained numerically from a model of the plant, though if there is a model better design methods can be used instead (such as those of Section 21.3 below).
- It makes little sense to obtain the step response analytically from a model of the plant, since the tuning rule is an approximative design method, and the result will likely have to be fine tuned, as mentioned above.

*Ziegler-Nichols
gain method*

Critical gain

Critical period

critical

The Ziegler-Nichols critical gain method can be applied to plants that, when controlled with proportional gain K in closed loop, have a **critical gain** K_{cr} , i.e. a gain that makes the closed loop marginally stable. The step response of this marginally stable closed loop has oscillations with a constant amplitude; the period of those oscillations is the **critical period** P_{cr} . The closed loop is illustrated in Figure 21.6. As you know from studying the root locus, a plant will have a critical gain if:

- it has a number of poles which exceeds the number of zeros by at least three,
- its poles are stable, so that the corresponding branches are on the left complex half plane for small values of K , and one or more cross the imaginary axis as K increases.

Still, there are some plants which have a critical gain but for which this rule does not apply (this usually happens when the root locus is particularly convoluted).

Once K_{cr} and P_{cr} are known, controller parameters are found for (21.2) as shown in Table 21.2.

Notice that:

- The critical gain and period may be found experimentally, but there are many plants for which this is not possible for safety reasons: it may be dangerous to bring a control loop to marginal stability (there are plants that can explode, break apart, melt down...).

Table 21.2: PID parameters according to the Ziegler-Nichols critical gain method. See Figure 21.3. It is not guaranteed that PIDs designed to have small or no overshoot will, in fact, meet that specification.

Type of controller	K_p	T_i	T_d
P	$0.5K_{cr}$	—	—
PI	$0.45K_{cr}$	$\frac{P_{cr}}{1.2}$	—
PID	$0.6K_{cr}$	$\frac{P_{cr}}{2}$	$\frac{P_{cr}}{8}$
PID with a small overshoot	$\frac{1}{3}K_{cr}$	$\frac{P_{cr}}{2}$	$\frac{P_{cr}}{3}$
PID with no overshoot	$0.2K_{cr}$	$\frac{P_{cr}}{2}$	$\frac{P_{cr}}{3}$

- K_{cr} and P_{cr} can be found numerically or analytically from a model of the plant (you can use the root locus and the Routh-Hurwitz criterion).

21.3 PID design by pole-placement

Controllers can be designed by placing closed loop poles in the regions of the complex plane corresponding to the desired specifications. As you know, responses also depend on zeros, which means that this design method, though analytical and usually leading to better performances than tuning rules, may still require fine tuning at the end.

Example 21.1. Suppose that we want to control plant

$$G(s) = \frac{1}{(s+1)(s+2)(s+10)} \quad (21.6)$$

so as to achieve

- a steady-state $e_{ss} = 0$ for a constant reference,
- $t_{s,2\%} \leq 0.8$ s,
- $M_p \leq 4.3\%$.

Concerning the first specification, since the plant has no poles at the origin, the first specification will require a controller with one pole at the origin, i.e. a PI or a PID.

Concerning the second specification, from what we saw in Section 16.4 we know that the real part $-a$ of the closed loop dominant poles $s = -a \pm b$ must verify

$$-a \leq -\frac{4}{0.8} = -5 \quad (21.7)$$

Concerning the third specification, we know that the dominant poles must verify

$$|\angle s| \geq \arctan \frac{\pi}{\log 0.043} = 135^\circ \quad (21.8)$$

Putting all this together, it is seen that a PI or PID is needed, that will put the closed loop dominant poles in the zone of the complex plane shown in Figure 21.7. We now use the root locus to check if a PI controller, that adds a pole at the origin and a negative real zero, is enough to bring the dominant poles into that zone. Figure 21.8 shows that to be impossible. The root locus diagrams shown correspond to particular values of the PI zero, but it is clear that, whatever the precise location of the zero in either of the three cases, specifications are always impossible to fulfil.

Thus, a PID is needed. It is expedient to use (21.3) and begin with the design of a PD; the integral part is added at the end. The transfer function $F(s)$ of the closed loop formed by plant (21.6) and a PD is given by

$$F(s) = \frac{K_P(1 + T_D s)G(s)}{1 + K_P(1 + T_D s)G(s)} \quad (21.9)$$

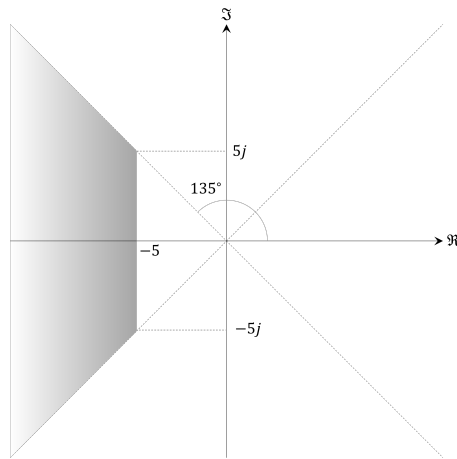


Figure 21.7: Zone of C where closed loop dominant poles must be according to the specifications of Example 21.1.

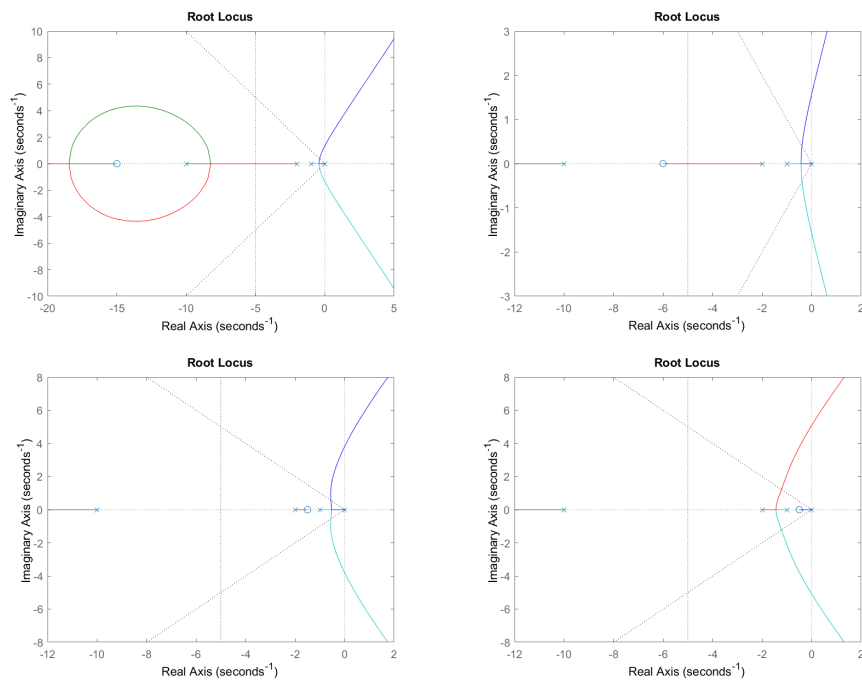


Figure 21.8: A PI cannot fulfil the specifications of Example 21.1, as the two rightmost branches never go to the left of -5 . Top left: the zero is to the left of the pole at -10 . Top right: the zero lies between the poles at -10 and -2 . Bottom left: the zero lies between the poles at -2 and -1 . Bottom right: the zero is to the right of the poles at -1 .

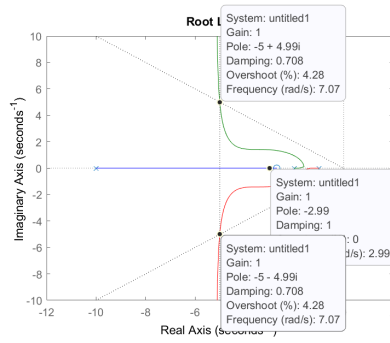


Figure 21.9: Root locus of (21.6) controlled with PD $C(s) = 129.67(1 + 0.37s)$, from Example 21.1.

We want the situation in Figure 21.9, with the leftmost pole going right towards the zero, and the rightmost poles being pulled to the left inside the desired zone. The limit case is that in which the branches pass through $-5 \pm 5j$, i.e.

$$\begin{aligned}
 & 1 + K_P(1 + T_D s)G(s) \Big|_{s=-5 \pm 5j} = 0 \\
 \Leftrightarrow & K_P(1 + T_D(-5 \pm 5j))G(-5 \pm 5j) = -1 \\
 \Leftrightarrow & \begin{cases} |K_P(1 + T_D(-5 \pm 5j))G(-5 \pm 5j)| = 1 \\ \arg[K_P(1 + T_D(-5 \pm 5j))G(-5 \pm 5j)] = \pm 180^\circ \end{cases} \quad (21.10)
 \end{aligned}$$

Let us choose for instance $s = -5 + 5j$ and the -180° phase for the second equation, that becomes

$$\begin{aligned}
 & \overbrace{\arg K_P}^0 + \arg(1 - 5T_D + 5T_D j) - \arg(-5 + 5j + 1) - \arg(-5 + 5j + 2) - \arg(-5 + 5j + 10) = -180^\circ \\
 \Leftrightarrow & \arctan \frac{5T_D}{1 - 5T_D} = -180^\circ + \arctan \frac{5}{-4} + \arctan \frac{5}{-3} + \arctan \frac{5}{5} \\
 \Leftrightarrow & \frac{5T_D}{1 - 5T_D} = \tan 114.6^\circ \\
 \Leftrightarrow & 5T_D = -2.18(1 - 5T_D) \Leftrightarrow T_D = 0.37 \quad (21.11)
 \end{aligned}$$

The first equation now becomes

$$\begin{aligned}
 & \frac{K_P |1 - 5 \times 0.37 + 5 \times 0.37j|}{|-5 + 5j + 1| |-5 + 5j + 2| |-5 + 5j + 10|} = 1 \\
 \Leftrightarrow & K_P \sqrt{(-0.85)^2 + 1.85^2} = \sqrt{16 + 25} \sqrt{9 + 25} \sqrt{25 + 25} \\
 \Leftrightarrow & K_P = 129.67 \quad (21.12)
 \end{aligned}$$

Figure 21.9 shows the root locus of $C(s)G(s)$ where $C(s) = 129.67(1 + 0.37s)$ is the PD just found. The two problems with PID design by pole placement should be clear by now:

- first, we had to solve a non-linear system of equations, which in this case was relatively easy, but often involves much more difficult calculations;
- then, the result turned out not to achieve the desired performance. In this case, even though we pulled the two complex conjugate poles as little to the left as possible, the real pole which is moving right ends up being dominant (it is about $s = -3$, clearly outside the desired zone), and so the settling time specification will not be fulfilled. Very often something like this happens, and the pole or poles which are being placed become dominated by another pole or by a zero, changing the expected behaviour of the controlled plant.

For this reason, it is usually expedient to use app `controlSystemDesigner` to design controllers by pole placement. Figure 21.10 shows that neither the settling time specification nor the maximum overshoot specification are, in fact, verified. Thus, the position of the PD zero should be adjusted until they are,

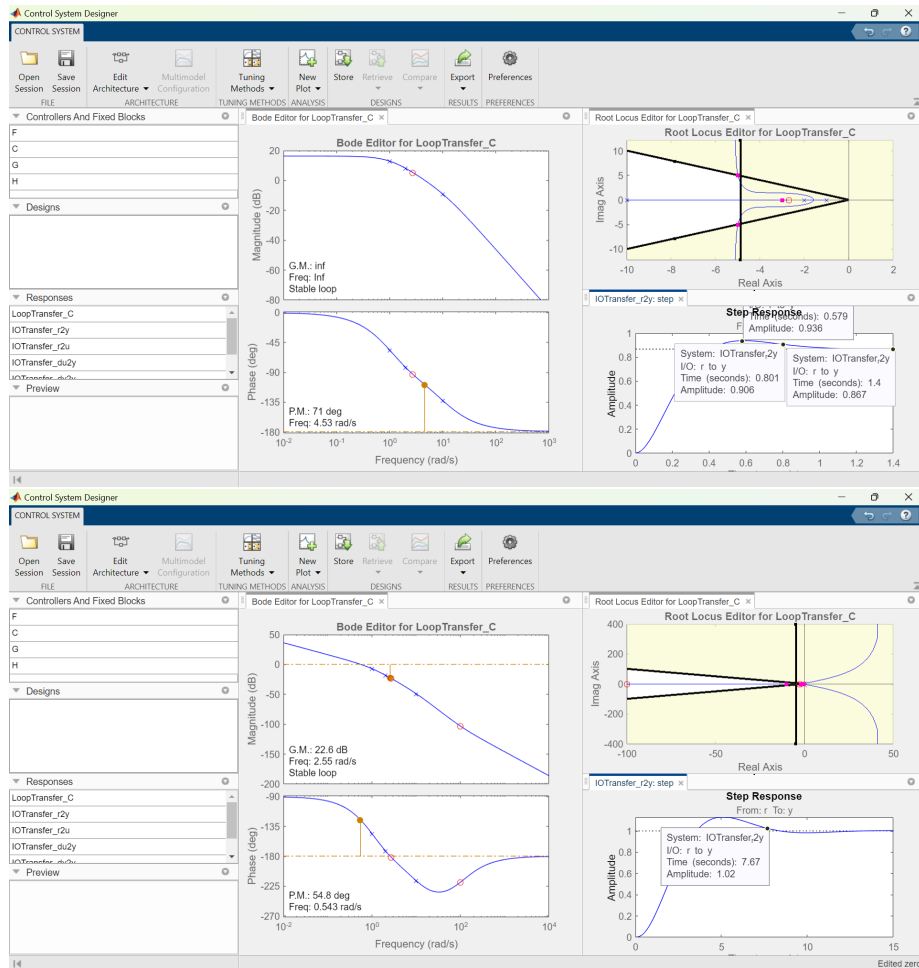


Figure 21.10: Top: controlSystemDesigner confirms that plant (21.6) controlled with PD $C(s) = 129.67(1 + 0.37s)$ does not fulfil the specifications of Example 21.1. Bottom: adding the PI part of the PID controller.

if this is at all possible. If, on the other hand, the resulting settling time could be accepted, then the PI is added, with a zero sufficiently fast to perturb as little as possible the time response achieved by the PD. This too is shown in Figure 21.10; notice how there is now no steady state error, but the settling time increases further to more than 7 s. \square

Glossary

Ἀκούσαντες δὲ ὅτι τῇ Ἑβραϊδὶ διαλέκτῳ προσεφώνει αὐτοὺς μᾶλλον παρέσχον ἡσυχίαν.

Saint LUKE the Evangelist (? — †84), *Acts of the Apostles*, xxii 2

critical gain ganho crítico
critical period período crítico
tuning rule regra de sintonia

Exercises

1. Show that:

(a) to rewrite (21.2) in the form of (21.1), we must do

$$P = k_p \quad (21.13)$$

$$I = \frac{k_p}{T_i} \quad (21.14)$$

$$D = k_p T_d \quad (21.15)$$

(b) to rewrite (21.3) in the form of (21.1), we must do

$$P = K_P \left(1 + \frac{T_D}{T_I} \right) \quad (21.16)$$

$$I = \frac{K_P}{T_I} \quad (21.17)$$

$$D = K_P T_D \quad (21.18)$$

(c) to rewrite (21.1) in the form of (21.2), we must do

$$k_p = P \quad (21.19)$$

$$T_i = \frac{P}{I} \quad (21.20)$$

$$T_d = \frac{D}{P} \quad (21.21)$$

(d) to rewrite (21.1) in the form of (21.3), we must do either

$$K_P = \frac{P + \sqrt{P^2 - 4DI}}{2} \quad (21.22)$$

$$T_I = \frac{P + \sqrt{P^2 - 4DI}}{2I} \quad (21.23)$$

$$T_D = \frac{2D}{P + \sqrt{P^2 - 4DI}} \quad (21.24)$$

or in alternative

$$K_P = \frac{P - \sqrt{P^2 - 4DI}}{2} \quad (21.25)$$

$$T_I = \frac{P - \sqrt{P^2 - 4DI}}{2I} \quad (21.26)$$

$$T_D = \frac{2D}{P - \sqrt{P^2 - 4DI}} \quad (21.27)$$

(e) to rewrite (21.3) in the form of (21.2), we must do

$$k_p = K_P \left(1 + \frac{T_D}{T_I} \right) \quad (21.28)$$

$$T_i = T_I + T_D \quad (21.29)$$

$$T_d = \frac{T_I T_D}{T_I + T_D} \quad (21.30)$$

(f) to rewrite (21.2) in the form of (21.3), we must do either

$$K_P = k_p \frac{1 + \sqrt{1 - 4\frac{T_d}{T_i}}}{2} \quad (21.31)$$

$$T_I = T_i \frac{1 + \sqrt{1 - 4\frac{T_d}{T_i}}}{2} \quad (21.32)$$

$$T_D = T_d \frac{2}{1 + \sqrt{1 - 4\frac{T_d}{T_i}}} \quad (21.33)$$

or in alternative

$$K_P = k_p \frac{1 - \sqrt{1 - 4\frac{T_d}{T_i}}}{2} \quad (21.34)$$

$$T_I = T_i \frac{1 - \sqrt{1 - 4\frac{T_d}{T_i}}}{2} \quad (21.35)$$

$$T_D = T_d \frac{2}{1 - \sqrt{1 - 4\frac{T_d}{T_i}}} \quad (21.36)$$

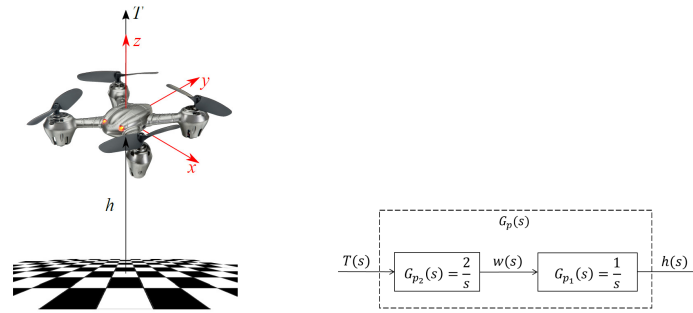


Figure 21.11: Left: multirotor drone; centre: block diagram with the control system of the drone.

2. Find a PI controller and a PID controller for plant $G(s) = \frac{-\frac{1}{2}s + 1}{s^2 + s + \frac{1}{3}}$, using the open-loop (reaction curve) Ziegler-Nichols method.
3. Find a PID controller for plant $G(s) = \frac{10}{s(s+1)^2}$, using the closed-loop (critical gain) Ziegler-Nichols method.
4. Consider the multirotor drone in Figure 21.11. The four rotors exert an upwards force F , due to which the drone hovers at height h , with a vertical velocity $w = \dot{h}$. The system can be represented by the block diagram in the same Figure and will be controlled in closed loop aiming at
 - a 2% settling time of 1 second or less,
 - a maximum overshoot of 4.3% or less, and
 - no steady-state error for a constant height reference.
 - (a) Find which specifications cannot be satisfied with a proportional controller $C(s) = 1$.
 - (b) Propose a controller structure, find its parameters, and verify that specifications are satisfied.
5. A torque T_f is applied to a ship to control its roll θ , as seen in Figure 21.12. The corresponding transfer function is $G_p(s) = \frac{\Theta(s)}{T_f(s)} = \frac{9}{s^2 + 1.2s + 9}$. The system will be controlled in closed loop aiming at
 - a 2% settling time of 4 second or less,
 - a maximum overshoot of 4.3% or less, and
 - a steady-state error for a 0° reference of 10% or less.

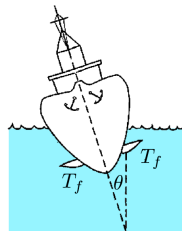


Figure 21.12: Ship from Exercise 5.

- (a) Find which specifications cannot be satisfied with a proportional controller $C(s) = 1$.
 - (b) Propose a controller structure, find its parameters, and verify that specifications are satisfied.
6. Find PID controllers for the following plants:

- (a) $G_1(s) = \frac{1}{s+1}$
- (b) $G_2(s) = \frac{1}{(s+1)^2}$
- (c) $G_3(s) = \frac{1}{(s+1)^3}$
- (d) $G_4(s) = \frac{1}{(s+1)^4}$
- (e) $G_5(s) = \frac{1}{(s+1)^8}$
- (f) $G_6(s) = \frac{1}{(s+1)(1+0.1s)(1+0.1^2s)(1+0.1^3s)}$
- (g) $G_7(s) = \frac{1}{(s+1)(1+0.2s)(1+0.2^2s)(1+0.2^3s)}$
- (h) $G_8(s) = \frac{1}{(s+1)(1+0.5s)(1+0.5^2s)(1+0.5^3s)}$
- (i) $G_9(s) = \frac{1-0.1s}{(s+1)^3}$
- (j) $G_{10}(s) = \frac{1-0.2s}{(s+1)^3}$
- (k) $G_{11}(s) = \frac{1-0.5s}{(s+1)^3}$
- (l) $G_{12}(s) = \frac{1-s}{(s+1)^3}$
- (m) $G_{13}(s) = \frac{1-2s}{(s+1)^3}$
- (n) $G_{14}(s) = \frac{1-5s}{(s+1)^3}$

7. Find PID controllers for the following plants:

- (a) $G_{15}(s) = \frac{100}{(s+10)^2} \left(\frac{1}{s+1} + \frac{0.5}{s+0.05} \right)$
- (b) $G_{16}(s) = \frac{(s+6)^2}{s(s+1)^2(s+36)}$
- (c) $G_{17}(s) = \frac{1}{(s+1)(s^2+0.2s+1)}$
- (d) $G_{18}(s) = \frac{2^2}{(s+1)(s^2+0.2 \times 2s+2^2)}$
- (e) $G_{19}(s) = \frac{5^2}{(s+1)(s^2+0.2 \times 5s+5^2)}$
- (f) $G_{20}(s) = \frac{10^2}{(s+1)(s^2+0.2 \times 10s+10^2)}$
- (g) $G_{21}(s) = \frac{1}{s^2-1}$
- (h) $G_{22}(s) = \frac{1}{s(s+1)}$
- (i) $G_{23}(s) = \frac{1}{s(s+1)^2}$
- (j) $G_{24}(s) = \frac{1}{s(s+1)^3}$
- (k) $G_{25}(s) = \frac{1}{s(s+1)^4}$
- (l) $G_{26}(s) = \frac{1}{s(s+1)^8}$
- (m) $G_{27}(s) = \frac{1}{s(s+1)(1+0.1s)(1+0.1^2s)(1+0.1^3s)}$

$$(n) G_{28}(s) = \frac{1}{s(s+1)(1+0.2s)(1+0.2^2s)(1+0.2^3s)}$$

$$(o) G_{29}(s) = \frac{1}{s(s+1)(1+0.5s)(1+0.5^2s)(1+0.5^3s)}$$

$$(p) G_{30}(s) = \frac{1-0.1s}{s(s+1)^3}$$

$$(q) G_{31}(s) = \frac{1-0.2s}{s(s+1)^3}$$

$$(r) G_{32}(s) = \frac{1-0.5s}{s(s+1)^3}$$

$$(s) G_{33}(s) = \frac{1-s}{s(s+1)^3}$$

$$(t) G_{34}(s) = \frac{1-2s}{s(s+1)^3}$$

$$(u) G_{35}(s) = \frac{1-5s}{s(s+1)^3}$$

Chapter 22

Design of lead-lag controllers

“Time lag—time lag! That idiot of a platform controller thought he was using a local radio circuit. But he’d been accidentally switched through a satellite—oh, maybe it wasn’t his fault, but he should have noticed. That’s a half-second time lag for the round trip. Even then it wouldn’t have mattered flying in calm air. It was the turbulence over the Grand Canyon that did it. When the platform tipped, and he corrected for that—it had already tipped the other way. Ever tried to drive a car over a bumpy road with a half-second delay in the steering?”

Arthur C. CLARKE (1917 — †2008), *A meeting with Medusa*, 2, Playboy, December 1971



This chapter is still being written.

In the picture: National Pantheon, or Church of Saint Engratia, Lisbon (source: <http://www.panteonacional.gov.pt/171-2/historia-2/>), some-when during its construction (1682–1966).

Glossary

“Nay, ye shall know the truth. We come from another world, though we are men such as ye; we come,” I went on, “from the biggest star that shines at night.”

“Oh! oh!” groaned the chorus of astonished aborigines.

“Yes,” I went on, “we do, indeed”; and again I smiled benignly, as I uttered that amazing lie. “We

come to stay with you a little while, and to bless you by our sojourn. Ye will see, O friends, that I have prepared myself for this visit by the learning of your language.”

“It is so, it is so,” said the chorus.

“Only, my lord,” put in the old gentleman, “thou hast learnt it very badly.”

I cast an indignant glance at him, and he quailed.

—Já que vos perdoei, porque sois ignorantes, condescendo tambem em vos dizer quem somos. Somos Espiritos! Vivemos além, por cima das nuvens, n’uma d’aquellas estrelas que vós vêdes de noite brilhar. E viemos visitar esta terra, mas em paz e para alegria de todos!

Entre os indigenas correram grandes ah! ah! Lentos e maravilhados.

Eu prosegui, mais grave:

—Nós conhecemos todos os reis e todas as gentes. E eu, que sou a voz dos outros, conheço todas as linguas.

—A nossa bem mal! Arriscou com timidez o velho guerreiro.

Dardejei-lhe um olhar chammejante que o estarreceu.

H. Rider HAGGARD (1856 — †1925), *King Solomon’s mines* (1885), VII (transl. Eça de QUEIROZ (1845 — †1900), *As minas de Salomão*, 1891, V)

word in English word in Portuguese
palavra em inglês palavra em português

Exercises

- Consider plant $G(s) = \frac{s + 5}{(s + 0.5)(s^2 + 0.6s + 1.09)}$.
 - Find a lead compensator $C(s)$ for this plant that fulfills the following specifications:
 - The gain margin must be infinite.
 - The phase margin must be $PM = 20^\circ$ with a $\pm 10\%$ tolerance.
 - The steady state error cannot be affected.
 - Find a lead compensator $C(s)$ for this plant that puts a pair of poles at $-4 + 10j$, and verify the 2%-settling time against the performance expected.
- The transfer function relating the voltage applied to the motor $u(t)$ with the azimuth angle of an antenna $\theta(t)$ is $G(s) = \frac{\Theta(s)}{U(s)} = \frac{1}{s(10s + 1)}$; see Figure 22.1. We want
 - an overshoot to a step in the reference angle of 16% or less;
 - a 2% settling time of 10 s or less;
 - no steady-state error for a constant reference.
 - Find the region of the s -plane where closed loop poles can lie.
 - Prove that you cannot fulfill all specifications using proportional control only.
 - Propose a structure for the controller.
 - Find the parameters of a controller of the lead-lag family and show that all specifications are thereby fulfilled.
 - Do the same for a controller of the PID family.

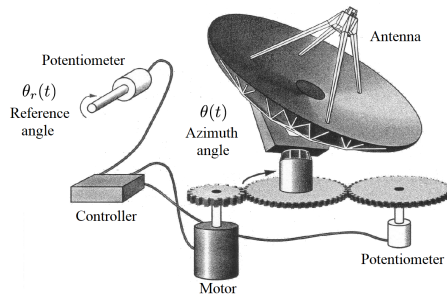


Figure 22.1: Antenna from Exercise 2.

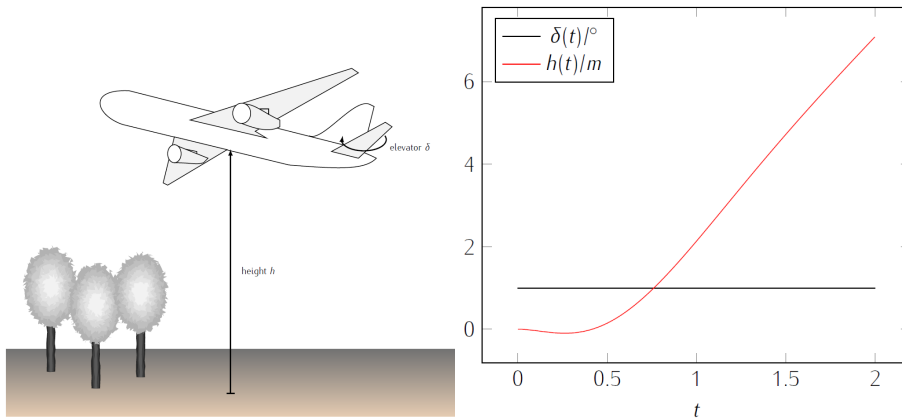


Figure 22.2: Airplane from Exercise 3 and the unit step response from its model.

3. Figure 22.2 shows a plane, together with the variation of its height $h(t)$ in metres when a 1° -step is applied to angle of the elevators $\delta(t)$.
- Which of the plots A—H in Figure 22.3 corresponds to the root locus of transfer function $\frac{H(s)}{\Delta(s)}$ when $K > 0$?
 - Which of the plots A—H in Figure 22.3 corresponds to the root locus of transfer function $\frac{H(s)}{\Delta(s)}$ when $K < 0$?
 - Is a feedback loop with proportional control enough to ensure an accurate tracking of a constant vertical velocity $\dot{h}(t)$?
 - Knowing that a feedback loop with a proportional controller equal to 1 leads to a steady state error of 0.9 when tracking a constant vertical velocity $\dot{h}(t) = 1$ m/s, design a controller reducing this steady state error to 0.3, without significantly deteriorating the transient response.
4. The transfer function relating the depth of facing performed by a lathe (output) with the control action of the motor (input) is

$$G_p(s) = \frac{2}{s(s+1)(s+5)} \quad (22.1)$$

The desired specifications are:

- maximum overshoot of 10% or less;
- steady-state error of 0.125 m or less for a constant velocity reference.

- Show that not all specifications can be satisfied with a proportional controller $K = 1$.
 - Propose a structure for the controller, find its parameters, and verify that specifications are fulfilled.
5. The horizontal position $x(t)$ of a robot is given by transfer function

$$G_p(s) = \frac{X(s)}{U(s)} = \frac{1.6}{(s+1)^2} \quad (22.2)$$

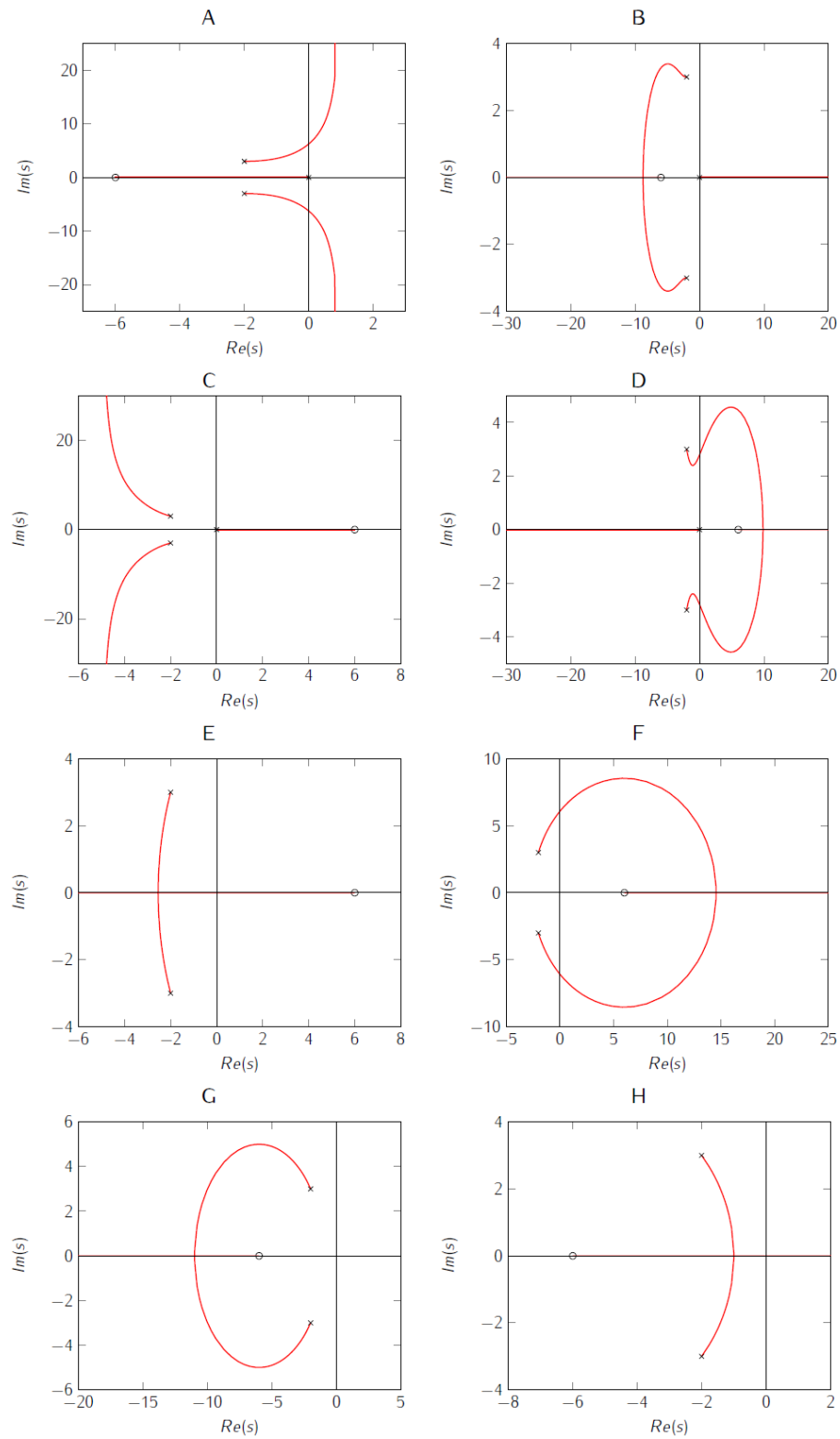


Figure 22.3: Possible root locus plots for the airplane from Exercise 3.

where $u(t)$ is the control action. The desired specifications are:

- maximum overshoot of 10% or less;
 - phase margin of 50° or more;
 - steady-state error of 10% or less for a constant position reference.
- (a) Show that not all specifications can be satisfied with a proportional controller $K = 1$.
- (b) Propose a structure for the controller, find its parameters, and verify that specifications are fulfilled.

Chapter 23

Internal Model Control

Sir Humphrey: The Civil Service was united in its desire to make sure that the Common Market didn't work. That's why we went into it.

Hacker: What are you talking about?

Sir Humphrey: Minister, Britain has had the same foreign policy objective for at least the last 500 years: to create a disunited Europe. In that cause we have fought with the Dutch against the Spanish, with the Germans against the French, with the French and Italians against the Germans, and with the French against the Germans and Italians. Divide and rule, you see. Why should we change now, when it's worked so well?

Hacker: That's all ancient history, surely?

Sir Humphrey: Yes, and current policy. We had to break the whole thing up, so we had to get inside. We tried to break it up from the outside, but that wouldn't work. Now that we're inside we can make a complete pig's breakfast of the whole thing: set the Germans against the French, the French against the Italians, the Italians against the Dutch... The Foreign Office is terribly pleased; it's just like old times.

Antony JAY (1930 — †2016), Jonathan LYNN (1943 — ...), *Yes, Minister*, I 5
(The writing on the wall, 1980)

Internal Model Control (IMC) may be seen as a variation of closed-loop control or as a design method for closed-loop controllers. It is the object of Exercise 3 from Chapter 9, and is developed in this Chapter.

23.1 IMC as a variation of closed-loop control

IMC can be used when there is a good model of the system and a good inverse model of the system (i.e. a model that receives the system's output $y(t)$ as input, and delivers the system's input $u(t)$ as output). These are used in the configuration shown in Figure 23.1, where

- $G(s)$ is the actual system to control,
- $G^*(s)$ is the model of the plant,
- $G^{-1}(s)$ is the inverse model of the plant.

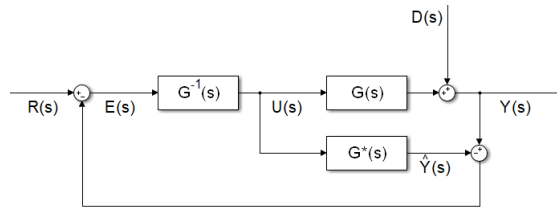


Figure 23.1: Block diagram of IMC.

In this configuration,

$$\begin{aligned}
 E &= R - (Y - \hat{Y}) \\
 &= R - D - GU + G^*U \\
 &= R - D - GG^{-1}E + G^*G^{-1}E \quad (23.1) \\
 \Rightarrow E(1 + GG^{-1} - G^*G^{-1}) &= R - D \\
 \Rightarrow E &= R \frac{1}{(1 + GG^{-1} - G^*G^{-1})} - D \frac{1}{(1 + GG^{-1} - G^*G^{-1})} \\
 Y &= GG^{-1}E + D = R \frac{GG^{-1}}{1 + GG^{-1} - G^*G^{-1}} - D \frac{GG^{-1}}{1 + GG^{-1} - G^*G^{-1}} + D \\
 &= R \frac{GG^{-1}}{1 + GG^{-1} - G^*G^{-1}} + D \frac{1 - G^*G^{-1}}{1 + GG^{-1} - G^*G^{-1}} \quad (23.2)
 \end{aligned}$$

Notice that if the model is perfect, i.e. if $G^*(s) = G(s)$, then the error (23.1) is

$$E(s) = R(s) - D(s) \underbrace{-G(s)U(s) + G^*(s)U(s)}_0 = R(s) - D(s). \quad (23.3)$$

If the inverse model is perfect, then

$$Y(s) = D(s) + \underbrace{G(s)G^{-1}(s)}_1 E(s) = D(s) + R(s) - D(s) = R(s) \quad (23.4)$$

In other words, with perfect models, IMC achieves perfect disturbance rejection, i.e. perfect robustness to disturbances. In practice, models are never perfect, but IMC often works well enough if models are fairly good.

IMC is suitable when black-box models (direct and inverse) of the plant are available. If models are transfer functions, then a proper $G^*(s)$ implies an inverse model $G^{-1}(s)$ which is not proper. Consequently, additional poles will have to be included — just as for PIDs and lead-lag controllers. To be more precise, if $G^*(s)$ has n poles and m zeros and is strictly proper, then $n - m + 1$ poles have to be added to $G^{-1}(s)$ so that it will become strictly proper.

Example 23.1. Consider the following case:

$$G(s) = \frac{s + 2}{(s^2 + 0.18s + 1)(s + 0.1)} \quad (23.5)$$

$$G^*(s) = \frac{s + 2}{(s^2 + 0.2s + 1)(s + 0.1)} \quad (23.6)$$

$$G^{-1}(s) = \frac{10^2(s^2 + 0.2s + 1)(s + 0.1)}{(s + 2)(s + 10)^2} \quad (23.7)$$

Notice that:

- the dominant pole of the model is -0.1 ;
- the model also has two non-dominant complex conjugate poles with $\xi = 0.1$ and $\omega_n = 1$ rad/s;
- we are assuming that the damping coefficient ξ of the complex poles was identified with a 10% error;

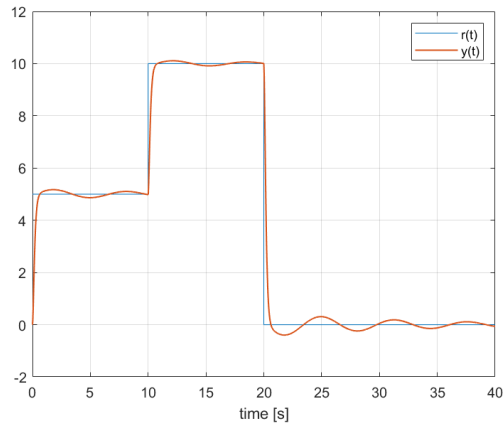


Figure 23.2: Results of Example 23.1.

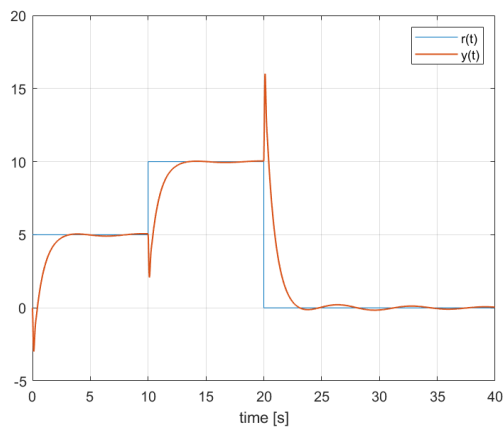


Figure 23.3: Results of Example 23.2.

- two poles, one decade above ω_n , were added to the inverse model, so that it will be proper.

The output of the IMC system in Figure 23.1 in this case is shown in Figure 23.2 for a reference consisting of three steps. \square

Example 23.2. Consider the following case:

$$G(s) = \frac{-s + 2}{(s^2 + 0.18 * s + 1)(s + 0.1)} \quad (23.8)$$

$$G^*(s) = \frac{2}{(s^2 + 0.2 * s + 1)(s + 0.1)} \quad (23.9)$$

$$G^{-1}(s) = \frac{10^3(s^2 + 0.2 * s + 1)(s + 0.1)}{2(s + 10)^3} \quad (23.10)$$

The difference from Example 23.1 is that there is a non-minimum phase zero. As a consequence, there would be an unstable pole in the inverse model. The way around this is to replace that pole by its steady-state gain; a third additional pole has to be added for a proper inverse model. And, because this approximation was used in the inverse model G^{-1} , it is better to use in G^* as well, otherwise results would be worst. The output of the IMC system in Figure 23.1 in this case is shown in Figure 23.3 for the same reference of Example 23.1. Notice how the effect of the non-minimum phase zero which was not properly modelled is now clearly felt. We will see in Example 23.3 how to improve upon this. \square

23.2 IMC as a design method for closed-loop controllers

Suppose that G^* and G^{-1} are known as transfer functions. Then IMC is equivalent to a usual closed-loop (see Figure 9.13) if

$$C(s) = \frac{G^{-1}(s)}{1 - G^{-1}(s)G^*(s)}. \quad (23.11)$$

In fact, in that case,

$$\begin{aligned} \frac{Y}{R} &= \frac{CG}{1 + CG} \\ &= \frac{\frac{G^{-1}G}{1 - G^{-1}G^*}}{1 + \frac{G^{-1}G}{1 - G^{-1}G^*}} \\ &= \frac{G^{-1}G}{1 - G^{-1}G^* + G^{-1}G} \end{aligned} \quad (23.12)$$

which is the same as (23.2).

Of course, if the inverse model $G^{-1}(s)$ is the exact inverse of the model $G^*(s)$ of the plant, then (23.11) becomes

$$C(s) = \frac{G^{-1}}{\underbrace{1 - \underbrace{G^{-1}G^*}_1}_0} \quad (23.13)$$

which makes no sense, and at best means that the control action $U = CE = \lim_{D \rightarrow 0} \frac{G^{-1}E}{D}$ should be very large (likely saturating the actuator in practice). Still, fairly good models may be used with (23.11) resulting in acceptable controllers.

Example 23.3. Plant (23.8) from Example 23.2 can be controlled in closed loop using (23.11) found from (23.9)–(23.9):

$$\begin{aligned} C(s) &= \frac{\frac{10^3(s^2+0.2*s+1)(s+0.1)}{2(s+10)^3}}{1 - \frac{10^3(s^2+0.2*s+1)(s+0.1)}{2(s+10)^3} \frac{2}{(s^2+0.2*s+1)(s+0.1)}} \\ &= \frac{\frac{10^3(s^2+0.2*s+1)(s+0.1)}{2(s+10)^3}}{1 - \frac{10^3}{(s+10)^3}} \\ &= \frac{10^3(s^2 + 0.2 * s + 1)(s + 0.1)}{(s + 10)^3 - 10^3} \\ &= \frac{10^3(s^2 + 0.2 * s + 1)(s + 0.1)}{s^3 + 30s^2 + 300s} \end{aligned} \quad (23.14)$$

This closed-loop controller actually performs better than the IMC from Example 23.2, as you may see simulating both cases with MATLAB. This may seem surprising since we just proved equivalence for the general case. But cancelled poles and zeros in (23.14) allow much better results numerical simulations — and in practice (23.11) also works better than Figure 23.1 *if*, of course, models are known as transfer functions. \square

Glossary

I got off the plane in Recife (the Brazilian government was going to pay the part from Recife to Rio) and was met by the father-in-law of César Lattes, who was the director of the Center for Physical Research in Rio, his wife, and another man. As the men were off getting my luggage, the lady started talking to me in Portuguese: “You speak Portuguese? How nice! How was it that you learned Portuguese?”

I replied slowly, with great effort. “First, I started to learn Spanish... then I discovered I was going to Brazil...” Now I wanted to say, “So, I learned Portuguese,” but I couldn’t think of the word for “so.” I knew how to make BIG words, though, so I finished the sentence like this: “CONSEQUENTEMENTE, *aprendi Português!*”

When the two men came back with the baggage, she said, “Oh, he speaks Portuguese! And with such wonderful words: CONSEQUENTEMENTE!”

Richard P. FEYNMAN (1918 — †1988), Ralph LEIGHTON (1949 — ...), “*Surely you’re joking, Mr. Feynman!*” — *Adventures of a curious character* (1985), 4

Internal Model Control controlro por modelo interno

Exercises

1. Find a closed-loop controller for the plant of Example 23.1 using (23.11). Compare its performance with the one shown in Figure 23.2.
2. Apply IMC to the plants of Exercises 6 and 7 from Chapter 21. Simulate the results using SIMULINK for the references mentioned in each Exercise.
 - (a) Implement the simulation using the block diagram of Figure 23.1.
 - (b) Implement the simulation using (23.11) and a closed loop.

Chapter 24

Delay systems

Abrupt and nerve-shattering the telephone rang. Ashley leapt up to answer it. Captain Granforte barred his way.

“Let him go, Captain,” said George Harlequin. “Let him do what he wants.”

Granforte stepped aside and Ashley stood with the receiver in his hands, listening to the crackling, impersonal voices saying “*Pronto! Pronto! Pronto!*” all the way up to Rome, and looking down at the dead face of Vittorio d’Orgagna and the blood that spread out over his white shirt-front. The ‘*prontos*’ started again in descending scale—Rome, Terracina, Naples, Castellammare, Sorrento—and finally, Hansen came on.

“*Pronto!* Hansen speaking.”

“This is Ashley... Sorrento.”

“Great to hear you, Ashley boy! Great, great! What’s news?”

Morris WEST (1916 — †1999), *The big story* (1957), 13

This chapter is still being written. In the picture below: National Pantheon, or Church of Saint Engratia, Lisbon (source: <http://www.panteaonacional.gov.pt/171-2/historia-2/>), somewhen during its construction (1682–1966).

24.1 Pure delays

Origin. Laplace transform. Frequency response. Bode, Nyquist and Nichols diagrams. The example of hot water in the shower. Margins.

Theorem 24.1. The Laplace transform of delay θ is the negative exponential $e^{-\theta s}$, i.e. given a function $f(t)$ with Laplace transform $F(s)$,

$$\mathcal{L}[f(t - \theta)] = F(s)e^{-\theta s} \quad (24.1)$$

Proof. Since we are using the unilateral Laplace transform, and thus the integral in

$$\mathcal{L}[f(t - \theta)] = \int_0^{+\infty} f(t - \theta)e^{-st} dt \quad (24.2)$$

begins at $t = 0$, it is clear from Figure 24.1 that we can multiply $f(t - \theta)$ by a delayed Heaviside function, without changing the result:

$$\mathcal{L}[f(t - \theta)] = \int_0^{+\infty} f(t - \theta)H(t - \theta)e^{-st} dt \quad (24.3)$$

Using variable change

$$\tau = t - \theta \quad (24.4)$$

$$t = 0 \Rightarrow \tau = -\theta \quad (24.5)$$

$$t = +\infty \Rightarrow \tau = +\infty \quad (24.6)$$

$$d\tau = dt \quad (24.7)$$

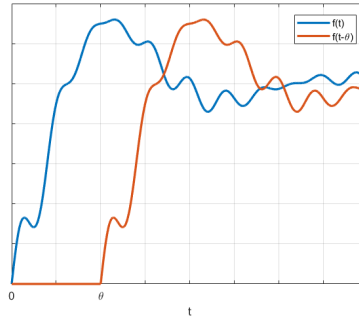


Figure 24.1: A function of time, and the same function delayed.

this becomes

$$\mathcal{L}[f(t-\theta)] = \int_{-\theta}^{+\infty} f(\tau)H(\tau)e^{-s(t-\theta)} d\tau \quad (24.8)$$

and because $H(t) = 0$ for $t < 0$ this integral can begin in 0:

$$\begin{aligned} \mathcal{L}[f(t-\theta)] &= e^{-s\theta} \int_0^{+\infty} f(\tau) \overbrace{H(\tau)}^1 e^{-s\tau} d\tau \\ &= e^{-s\theta} \underbrace{\int_0^{+\infty} f(\tau)e^{-s\tau} d\tau}_{\mathcal{L}[f(t)]} = F(s) e^{-s\theta} \square \end{aligned} \quad (24.9)$$

24.2 Padé approximations

Origin. Poles and zeros. Margins (approximated). Use in controller design.

24.3 Smith predictor

Variation of IMC. What it does and does not.

24.4 Control of systems similar to delay systems

Systems with non-minimum phase. Control design.

Systems with two significantly different time constants. Cascade control has already been mentioned in Exercise 4 from Chapter 9.



This chapter is still being written. In the picture: National Pantheon, or Church of Saint Engratia, Lisbon (source: <http://www.panteaonacional.gov.pt/171-2/historia-2/>), somewhen during its construction (1682–1966).

Glossary

«Padre nosso que estaes nos céus:—dizia Engracia Ripa, deixando correr um dos bugalhos de umas contas da terra sancta que tinha nas mãos.—Ora essa!—Sanctificado seja o vosso nome.—Forte tractante!—Venha a nós o vosso reino.—E uma pessoa com a sua áquella de que era um home como se quer!—Seja feita a vossa vontade.—Safa!—Assim na terra como nos céus.—Com que então setenta?»

«Entregadinhas!—*Ave Maria, gracia plena*:—respondeu a tia Jeronyma, que latinisava furiosamente á força de viver com o prior.—Como lh’o hei-de dizer?—*Domisteco*.—Foi o demo que o tentou.—*Beneditês tu...*»

Alexandre HERCULANO (1810 — †1877), *O Parocho da aldeia* (1843), VIII

cascade control controlo em cascata
(pure) delay atraso (puro)
Smith predictor preditor de Smith

Exercises

1. Consider the irrigation canal in Figure 24.2. The water from a reservoir enters the canal through a gate at height $y \in [0 \text{ m}, 1 \text{ m}]$. The gate is actuated by a motor controller by voltage $V \in [-10 \text{ V}, +10 \text{ V}]$. The velocity of the gate is given by

$$\dot{y} = \varphi V \quad (24.10)$$

and maximum velocity of the gate achieved by the motor is $\pm 0.1 \text{ m/s}$. We know from Fluid Mechanics that water dynamics in the canal is given by Saint-Venant equations (a simplified version of the Navier-Stokes equations for this case), but this can be approximated by a first order transfer function with delay, relating water height h with gate height y . In this canal,

$$\frac{h(s)}{y(s)} = \frac{100}{100s + 1} e^{-20s} \quad (24.11)$$

- (a) Draw the block diagram for this system, with input V and output h .
- (b) Show that a proportional controller K for gate height y leads to a closed loop transfer function given by

$$\frac{y(s)}{y_{\text{ref}}(s)} = \frac{0.01K}{s + 0.01K} \quad (24.12)$$

- (c) Find the value of K achieving the smallest possible settling time for a unit step reference, without saturation of the control action.
- (d) Let $K = 10$. Show that $h(s)$ and $y_{\text{ref}}(s)$ are related by transfer function

$$G(s) = \frac{h(s)}{y_{\text{ref}}(s)} = \frac{0.1 \times 100}{(s + 0.1)(100s + 1)} e^{-20s} \quad (24.13)$$

- (e) Show that, using a (1,1) Padé approximation, $G(s)$ above can be approximated by

$$\tilde{G}(s) = \frac{h(s)}{y_{\text{ref}}(s)} = \frac{100(1 - 10s)}{(10s + 1)^2(100s + 1)} \quad (24.14)$$

- (f) Plot the Bode diagram of $\tilde{G}(s)$. (Take into account that this is a non-minimum phase system.)
- (g) Since $G(s)$ is a system with delay a Smith predictor is actually a better control architecture. The block diagram is given in Figure 24.3, where κ is a controller. Identify signals 1, 2, 3, 4 and 5, and transfer functions 6, 7, 8, 9 and 10.

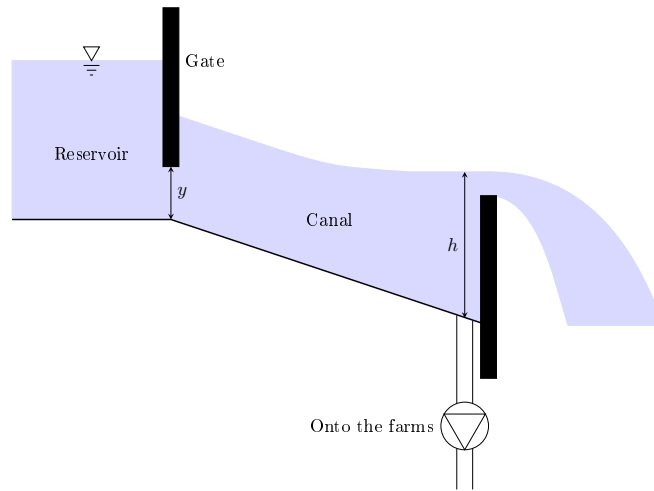


Figure 24.2: Irrigation canal.

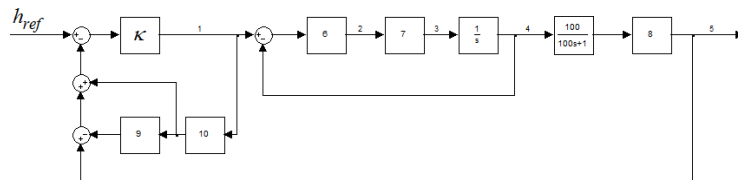


Figure 24.3: Smith predictor.

2. The angular velocity of the motor of a drone is $\omega(t)$ and depends on a control action $v(t)$:

$$G(s) = \frac{1}{0.04s + 1} e^{-0.35s} \quad (24.15)$$

- (a) Comparing plant (24.15) with plant

$$\tilde{G}(s) = \frac{1}{0.04s + 1}, \quad (24.16)$$

what does the term $e^{-0.35}$ change: only the gain, only the phase, both gain and phase, or neither gain nor phase?

- (b) Show that using a Padé approximation it is possible to approximate (24.15) by

$$G(s) \approx \frac{-s + 5.714}{0.04s^2 + 1.229s + 5.714} \quad (24.17)$$

- (c) Use approximation (24.17) and the Routh-Hurwitz criterion to find the values that a proportional controller K can take, ensuring the stability of the closed loop.
- (d) Explain if the real range of values of K that stabilise plant (24.15) is larger or smaller than the one you got using Padé approximation (24.17).
3. Explain why a non-minimum phase zero makes a plant harder to control.
4. Consider a plant with the transfer function $G(s) = \frac{0.5}{s(s+1)}e^{-2s}$. Find its gain and phase margins.
5. Find PID controllers for the following plants both without, and with, a Smith predictor, and compare performance.

(a) $G_1(s) = e^{-s}$

(b) $G_2(s) = \frac{1}{1 + 0.1s} e^{-s}$

(c) $G_3(s) = \frac{1}{1 + 0.2s} e^{-s}$

(d) $G_4(s) = \frac{1}{1 + 0.5s} e^{-s}$

(e) $G_5(s) = \frac{1}{1 + 2s} e^{-s}$

(f) $G_6(s) = \frac{1}{1 + 5s} e^{-s}$

(g) $G_7(s) = \frac{1}{1 + 10s} e^{-s}$

(h) $G_8(s) = \frac{1}{(1 + 0.1s)^2} e^{-s}$

(i) $G_9(s) = \frac{1}{(1 + 0.2s)^2} e^{-s}$

(j) $G_{10}(s) = \frac{1}{(1 + 0.5s)^2} e^{-s}$

(k) $G_{11}(s) = \frac{1}{(1 + 2s)^2} e^{-s}$

(l) $G_{12}(s) = \frac{1}{(1 + 5s)^2} e^{-s}$

(m) $G_{13}(s) = \frac{1}{(1 + 10s)^2} e^{-s}$

Part V

Practical implementation of control systems

Who would be satisfied with a navigator or engineer, who had no practice or experience whereby to carry on his scientific conclusions out of their native abstract into the concrete and the real?

Saint John Henry NEWMAN (1801 — †1890), *An essay in aid of a grammar of assent* (1870), VIII 1, 2

In this part of the lecture notes:

Chapter 25 is an introduction to digital signals and systems.

Chapter 26 shows how to find digital approximations of previously designed controllers.

Chapter 27 presents tools for the study of digital closed loop control systems, and how to use them to design digital controllers directly, rather than approximating a previous design.

Chapter 28 is about the effects of non-linearities in control systems.

Chapter 29 presents an overview of several practical issues of implementation.

Here is what you need to know beforehand to follow these chapters:

- The Laplace and Fourier transforms, from Chapter 2;
- Transfer functions, from Sections 4.1 and 4.2 of Chapter 4;
- System theory, from Part II;
- Filters, from Sections 12.2 and 12.3 of Chapter 12;
- Control theory, from Part IV (though you may have skipped Chapters 19 and 23).

Chapter 25

Digital signals and systems

Τοῦ δὲ ποσοῦ τὸ μὲν ἔστι διωρισμένον, τὸ δὲ συνεξέες.

ARISTOTLE (384 BC — †322 BC), *Kathegoriai*, VI

Controllers designed with the methods from Part IV can be implemented in many ways. All controllers until the mid-20th century used only mechanical, pneumatic or electrical components, and can be found today in simple cases or for special applications (remember Example 3.32).

Example 25.1. In the circuit of Figure 25.1,

$$V_A = -\frac{R_1}{R_1}e = -e \quad (25.1)$$

$$V_B = -\frac{\frac{1}{Cs}}{R_2}e = -\frac{1}{R_2Cs}e \quad (25.2)$$

$$u = -\frac{R_4}{R_3}V_A - \frac{R_4}{R_3}V_B = \underbrace{\frac{R_4}{R_3}}_{K_p} \left(1 + \underbrace{\frac{1}{R_2Cs}}_{T_i}\right)e \quad (25.3)$$

This is, consequently, an implementation of a PI controller. If R_4 and R_2 are potentiometers, the two parameters of the controller can be tuned sliding or rotating two buttons. \square

Because (as we saw in Chapter 3) most systems nowadays include digital signals, and because controllers are usually implemented using microprocessors, computers (which of course have microprocessors), or Programmable Logic Controllers (which are computers designed specifically to implement controllers in harsh environments, resisting to dust, vibrations, extreme temperatures, etc.), which only handle digital signals, it is important to study the consequences digital signals have for control performance. Remember that a digital signal

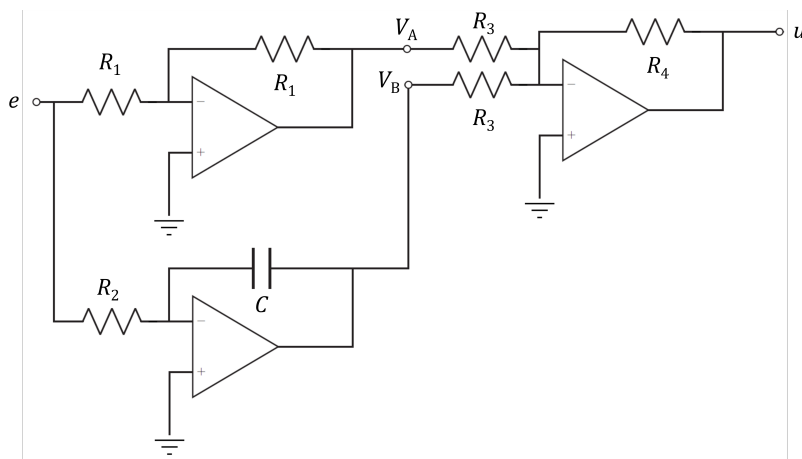


Figure 25.1: A PI controller implemented with OpAmps, resistors, and a capacitor.

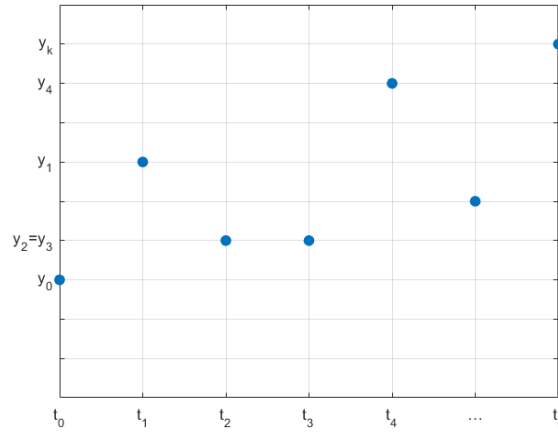


Figure 25.2: A digital signal.

- only exists in some time instants, i.e. is **discrete** in time;
- only assumes values from a discrete set, i.e. is **quantised** in amplitude.

We will concentrate on the effects of discretisation (and thus speak mostly of discrete signals and discrete systems, rather than digital signals and digital systems) because

- when a signal only assumes values in some time instants, derivatives no longer exist;
- whatever happens inbetween two sampling instants will only be known at the next sampling instant. Thus, the effect of discretisation is similar to that of a delay, and we know from Chapter 24 the problems this brings. Notice, however, that, even if the sampling time T_s is constant, the delay τ with which an event will be known will be variable. All we know is that $0 \leq \tau < T_s$.

When time is discrete, time instants and signal values are numbered as in Figure 25.2. Notice that

- if there are $n + 1$ time instants, they are numbered from 0 to t_n ;
- y_k is a shorthand for $y(t_k)$;
- if the sampling time T_s is constant, then

$$t_0 = 0 \quad (25.4)$$

$$t_1 = T_s \quad (25.5)$$

$$t_2 = 2T_s \quad (25.6)$$

$$t_3 = 3T_s \quad (25.7)$$

$$\vdots \quad (25.8)$$

$$t_k = kT_s \quad (25.9)$$

and

$$y_0 = y(t_0) = y(0) \quad (25.10)$$

$$y_1 = y(t_1) = y(T_s) \quad (25.11)$$

$$y_2 = y(t_2) = y(2T_s) \quad (25.12)$$

$$y_3 = y(t_3) = y(3T_s) \quad (25.13)$$

$$\vdots \quad (25.14)$$

$$y_k = y(t_k) = y(kT_s) \quad (25.15)$$

It is because of (25.4) and (25.10) that it is reasonable to begin numbering time instants at 0.

constant in practice

In practice, T_s is almost always constant, since this is much easier to implement — and it is fortunate that it be so, since the mathematical treatment is much simpler assuming a constant T_s , as we will always do in what follows.

25.1 The \mathcal{Z} transform

We will now find the Laplace transform of a discrete signal $y(t)$, such as (25.10)–(25.15), that only assumes a value if $t = k T_s$, $k \in \mathbb{Z}_0^+$. This signal is zero almost everywhere. Remembering property (10.21) of the Dirac delta function, we write

$$\begin{aligned} y(t) &= y(0)\delta(t) + y(T_s)\delta(t - T_s) + y(2T_s)\delta(t - 2T_s) + y(3T_s)\delta(t - 3T_s) + \dots + y(kT_s)\delta(t - kT_s) + \dots \\ &= y_0\delta(t) + y_1\delta(t - T_s) + y_2\delta(t - 2T_s) + y_3\delta(t - 3T_s) + \dots + y_k\delta(t - kT_s) + \dots \\ &= \sum_{k=0}^{+\infty} y_k\delta(t - kT_s) \end{aligned} \tag{25.16}$$

In this way, if we integrate $y(t)$

- and t is not a multiple of T_s (more precisely, $\sim \exists k \in \mathbb{Z}_0^+ : t = kT_s$), the integral is 0;
- and $t = 0$, we get back $y(0) = y_0$;
- and $t = T_s$, we get back $y(T_s) = y_1$;
- and $t = 2T_s$, we get back $y(2T_s) = y_2$;
- and, in the general case, $t = kT_s$, $k \in \mathbb{N}$, we get back $y(kT_s) = y_k$.

Theorem 25.1. The Laplace transform of a discrete signal y_k is

$$\mathcal{L}[y(t)] = \mathcal{L}\left[\sum_{k=0}^{+\infty} y_k\delta(t - kT_s)\right] = \sum_{k=0}^{+\infty} y_k (e^{-T_s s})^k \tag{25.17}$$

Proof. Applying (2.1) to $y(t)$, and then (24.1),

$$\begin{aligned} \mathcal{L}[y(t)] &= \int_0^{+\infty} \sum_{k=0}^{+\infty} y_k\delta(t - kT_s)e^{-st} dt \\ &= \sum_{k=0}^{+\infty} y_k \underbrace{\int_0^{+\infty} \delta(t - kT_s)e^{-st} dt}_{\mathcal{L}[\delta(t - kT_s)]} \\ &= \sum_{k=0}^{+\infty} y_k \underbrace{\mathcal{L}[\delta(t)]}_1 e^{-kT_s s} \square \end{aligned} \tag{25.18}$$

This Laplace transform justifies the following definitions:

Definition 25.1. The **delay operator** z^{-1} is given by

Delay operator z^{-1}

$$z^{-1} = e^{-T_s s} \tag{25.19}$$

Using this delay operator z^{-1} , we can write the Laplace transform (25.17) of a discrete signal y_k in a more simple way:

Definition 25.2. The \mathcal{Z} transform of a discrete signal y_k is its Laplace transform written with the delay operator z^{-1} :

$$\mathcal{Z}[y_k] = \sum_{k=0}^{+\infty} y_k z^{-k} \quad \square \tag{25.20}$$

Here are some consequences of the definition of z^{-1} :

- the delay operator $z^{-1} = e^{-T_s s}$ is a delay of one sampling time T_s ;
- $z^{-k} = e^{-kT_s s}$ is a delay of k sampling times T_s ;

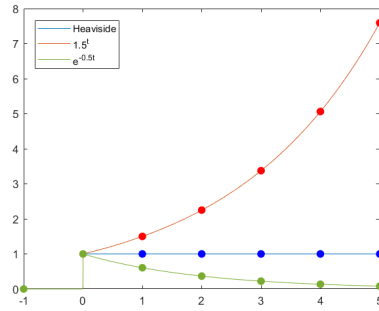


Figure 25.3: Three discrete time signals from Examples 25.2, 25.3, and 25.4, together with the continuous time functions that they discretise.

- the **forward operator** $z = e^{T_s s}$ is a forward shift in time of one sampling time T_s ;
- $z^k = e^{k T_s s}$ is a forward shift in time of k sampling times T_s ;
- because z and z^{-1} are defined with an exponential, neither is ever 0. Thus, it is always possible to divide by z or by z^{-1} .

Example 25.2. The \mathcal{Z} transform of the Heaviside function (see Figure 25.3)

$$h_0 = 1 \quad (25.21)$$

$$h_1 = 1 \quad (25.22)$$

$$h_2 = 1 \quad (25.23)$$

$$h_3 = 1 \quad (25.24)$$

$$\vdots \quad (25.25)$$

$$h_k = 1, \forall k \in \mathbb{Z}_0^+ \quad (25.26)$$

is the sum of a geometric series:

$$\mathcal{Z}[h_k] = \sum_{k=0}^{+\infty} z^{-k} = \sum_{k=0}^{+\infty} (z^{-1})^k = z^0 + z^{-1} + z^{-2} + z^{-3} + \dots \quad (25.27)$$

The first term is $z^0 = 1$, and the ratio is z^{-1} ; thus, the sum is

$$\mathcal{Z}[h_k] = H(z) = \frac{1}{1 - z^{-1}} = \frac{z}{z - 1} \quad (25.28)$$

Example 25.3. The \mathcal{Z} transform of the power function with base $a \neq 0$

$$x_0 = a^0 = 1 \quad (25.29)$$

$$x_1 = a^1 = a \quad (25.30)$$

$$x_2 = a^2 \quad (25.31)$$

$$x_3 = a^3 \quad (25.32)$$

$$\vdots \quad (25.33)$$

$$x_k = a^k, \forall k \in \mathbb{Z}_0^+ \quad (25.34)$$

is also the sum of a geometric series:

$$\mathcal{Z}[x_k] = \sum_{k=0}^{+\infty} a^k z^{-k} = \sum_{k=0}^{+\infty} \left(\frac{a}{z}\right)^k \quad (25.35)$$

The first term is $\left(\frac{a}{z}\right)^0 = 1$, and the ratio is $\frac{a}{z}$; thus, the sum is

$$\mathcal{Z}[x_k] = X(z) = \frac{1}{1 - \frac{a}{z}} = \frac{z}{z - a} = \frac{1}{1 - a z^{-1}} \quad (25.36)$$

Notice that (25.28) is a particular case of (25.36), for $a = 1$. □

Example 25.4. Function e^{-at} can be discretised as $x_k = e^{-akT_s}$, and its \mathcal{Z} transform is

$$\mathcal{Z}[x_k] = \sum_{k=0}^{+\infty} e^{-akT_s} z^{-k} = \sum_{k=0}^{+\infty} (z^{-1} e^{-aT_s})^k \quad (25.37)$$

The first term is 1, and the ratio is $z^{-1} e^{-aT_s}$; thus, the sum is

$$\mathcal{Z}[x_k] = X(z) = \frac{1}{1 - z^{-1} e^{-aT_s}} = \frac{z}{z - e^{-aT_s}} \quad \square \quad (25.38)$$

The \mathcal{Z} transform, being a Laplace transform, enjoys the linearity property \mathcal{Z} is linear of \mathcal{L} . This can also be shown from definition:

Theorem 25.2. The \mathcal{Z} transform is a linear operator:

$$\mathcal{Z}[a x_k] = \sum_{k=0}^{+\infty} a x_k z^{-k} = a \mathcal{Z}[x_k] \quad (25.39)$$

$$\begin{aligned} \mathcal{Z}[x_k + y_k] &= \sum_{k=0}^{+\infty} (x_k + y_k) z^{-k} \\ &= \sum_{k=0}^{+\infty} x_k z^{-k} + \sum_{k=0}^{+\infty} y_k z^{-k} = \sum_{k=0}^{+\infty} x_k z^{-k} + \sum_{k=0}^{+\infty} y_k z^{-k} \\ &= \mathcal{Z}[x_k] + \mathcal{Z}[y_k] \quad \square \end{aligned} \quad (25.40)$$

Theorem 25.3. The initial and final values of a discrete signal x_k are retrieved from its \mathcal{Z} transform $X(z)$ as *Initial value theorem*
Final value theorem

$$x_0 = \lim_{z \rightarrow +\infty} X(z) \quad (25.41)$$

$$\lim_{k \rightarrow +\infty} x_k = \lim_{z \rightarrow 1} (1 - z^{-1}) X(z) \quad (25.42)$$

Proof. (25.41) is an obvious consequence of the definition:

$$\lim_{z \rightarrow +\infty} X(z) = \lim_{z \rightarrow +\infty} x_0 + x_1 z^{-1} + x_2 z^{-2} + x_2 z^{-2} + x_3 z^{-3} + \dots = x_0 \quad (25.43)$$

To prove (25.42), we first notice that since z^{-1} is a delay then $z^{-1} X(z)$ is found as follows:

$$X(z) = \mathcal{Z}[x_k] = \sum_{k=0}^{+\infty} x_k z^{-k} \quad (25.44)$$

$$\Rightarrow z^{-1} X(z) = \mathcal{Z}[x_{k-1}] = \sum_{k=0}^{+\infty} x_{k-1} z^{-k} \quad (25.45)$$

We now subtract (25.45) from (25.44):

$$\begin{aligned} X(z) - z^{-1} X(z) &= \sum_{k=0}^{+\infty} x_k z^{-k} - \sum_{k=0}^{+\infty} x_{k-1} z^{-k} \\ \Leftrightarrow X(z) (1 - z^{-1}) &= \sum_{k=0}^{+\infty} (x_k - x_{k-1}) z^{-k} \\ \Rightarrow \lim_{z \rightarrow 1} X(z) (1 - z^{-1}) &= \lim_{z \rightarrow 1} \sum_{k=0}^{+\infty} (x_k - x_{k-1}) z^{-k} = \lim_{n \rightarrow +\infty} \sum_{k=0}^n (x_k - x_{k-1}) \\ &= \lim_{n \rightarrow +\infty} x_0 - x_{-1} + x_1 - x_0 + x_2 - x_1 + x_3 - x_2 + x_4 - x_3 + \dots + x_{n-1} - x_{n-2} + x_n - x_{n-1} \\ &= \lim_{n \rightarrow +\infty} x_n - \underbrace{x_{-1}}_0 \end{aligned} \quad (25.46)$$

We assume $x_{-1} = 0$ because the first sample of the signal is x_0 . \square

25.2 Discrete transfer functions

Because the time is discrete, there are no derivatives; it is impossible to calculate

No derivatives in discrete time

$$f'(t) = \lim_{h \rightarrow 0} \frac{f(t) - f(t-h)}{h} \quad (25.47)$$

since h cannot be smaller than T_s , or $f(t)$ will no longer be defined. That is why, as we saw in Chapter 3, digital systems are not described by differential equations, but rather by difference equations, that relate the successive values of the input and the output. Operator z^{-1} is very useful to write difference equations.

Example 25.5. The model

$$\frac{Y(s)}{U(s)} = \frac{1}{2+3s} \Rightarrow y(t) = \frac{1}{2}u(t) - \frac{3}{2}y'(t) \quad (25.48)$$

can be approximated as

$$y(t) = \frac{1}{2}u(t) - \frac{3}{2} \frac{y(t) - y(t-T_s)}{T_s} \Rightarrow y_k = \frac{1}{2}u_k - \frac{3}{2} \frac{y_k - y_{k-1}}{T_s} \quad (25.49)$$

using the smallest possible time interval. We will see below in Section 25.3 and Chapter 26 that this is not the only possible approximation, or even the best; but it suffices to show how difference equation (25.49) can be written using z^{-1} :

$$\begin{aligned} y_k &= \frac{1}{2}u_k - \frac{3}{2} \frac{y_k - y_{k-1}}{T_s} \\ \Leftrightarrow Y(z) &= \frac{1}{2}U(z) - \frac{3}{2} \frac{Y(z)(1-z^{-1})}{T_s} \\ \Leftrightarrow 2T_s Y(z) &= T_s U(z) - 3Y(z)(1-z^{-1}) \\ \Leftrightarrow Y(z)(2T_s + 3 - 3z^{-1}) &= U(z)T_s \\ \Leftrightarrow \frac{Y(z)}{U(z)} &= \frac{T_s}{2T_s + 3 - 3z^{-1}} \end{aligned} \quad (25.50)$$

Discrete transfer function (25.50) is a differential equation put under the form of a **discrete transfer function**. While a discrete transfer function relates the \mathcal{Z} transform of the output (in the numerator) and the input (in the denominator), just like a continuous time transfer function, is not unusual to abuse notation and write instead

$$\frac{y_k}{u_k} = \frac{T_s}{2T_s + 3 - 3z^{-1}} \quad (25.51)$$

There are several ways of creating a discrete transfer function with MATLAB:

- `tf` will create a discrete transfer function in z if, after the vectors with the coefficients of the numerator and the denominator, the sampling time is given;
- `tf` will create a discrete transfer function in z^{-1} if this is requested by option `'Variable'`;
- `z = tf('z')` creates the forward operator z , for which the sampling time may or may not be specified;
- `z_1 = tf([0 1],1,'Variable','z^-1')` creates the delay operator z^{-1} .

Simulink also has a block for discrete transfer functions.

Example 25.6. Transfer function

$$G(z) = \frac{z+2}{3z^2+4z+5} = \frac{z^{-1}+2z^{-2}}{3+4z^{-1}+5z^{-2}} \quad (25.52)$$

can be created as a function of z either as

```
>> tf([1 2],[3 4 5],0.5)
```

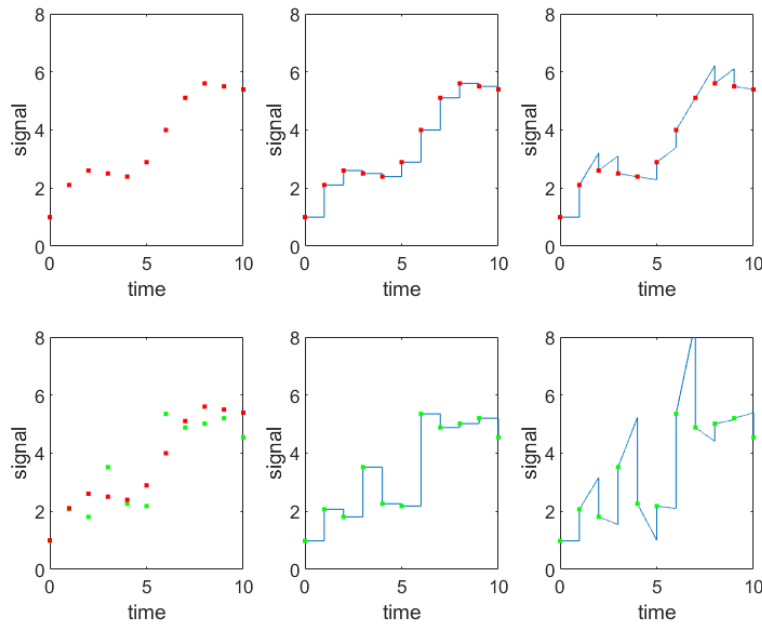


Figure 25.4: Top row: a signal, discrete in time, and the continuous signals obtained with a ZOH and a FOH. It is clear that the FOH easily over or underestimates the signal. Bottom row: the same signal, corrupted by noise. The ZOH is affected; the FOH provides wild outputs.

or as

```
>> z = tf('z',0.5);
>> (z+2)/(3*z^2+4*z+5)
```

As a function of z^{-1} , we either make (notice the leading zero)

```
>> tf([0 1 2],[3 4 5],0.5,'Variable','z^-1')
```

or

```
>> z_1 = tf([0 1],[1],0.5,'Variable','z^-1');
>> (z_1+2*z_1^2)/(3+4*z_1+5*z_1^2)
```

□

25.3 Zero order hold

The usual way of converting a discrete signal into a continuous one is the **zero order hold (ZOH)**, which we already met in Figure 3.13. It keeps the last value of the discrete signal constant until there is a new value. The output of a ZOH is a sequence of steps; i.e. a sequence of zero-order polynomials, that is to say, a zero-order spline (hence the name).

A possible alternative would be a first-order hold, that extrapolates the value from the last two previous samples, as seen in Figure 25.4. That Figure also makes clear the reason why this is seldom done: any noise will make the extrapolation far poorer than the result of a ZOH (and the electronics are of course harder to implement).

Control loops with a digital controller and a continuous plant are often found in practice. In that case, there must be a ZOH converting the digital control action u_k provided by the controller into a continuous signal $u(t)$, as seen in Figure 25.5. In that case, when discretising the plant, we have to find $\frac{y_k}{u_k}$, and thus the effects of the ZOH must be accounted for. The resulting discrete transfer function is usually denoted by $HG(z) = \frac{y_k}{u_k}$, to stress that it is not just $G(s)$ that is being discretised, but the ZOH too; the H stands for hold.

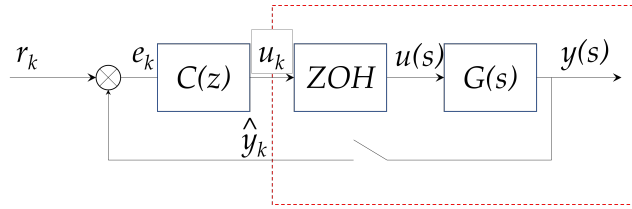


Figure 25.5: Control loop with a digital controller and a continuous plant. Notice that the reference r_k is discrete in time, just as the measured output \hat{y}_k and the closed loop error e_k . The plant output $y(t)$ is continuous, and the control action provided by the controller must be converted from a digital signal u_k into a continuous signal $u(t)$ by a ZOH. The part of the loop within the dashed line is what Theorem 25.4 is concerned with.

Theorem 25.4. A continuous plant $G(s) = \frac{y(s)}{u(s)}$ with a ZOH at its input (as shown in Figure 25.5) has a discrete equivalent $HG(z) = \frac{y_k}{u_k}$ given by

\mathcal{L} transform when the a ZOH

$$HG(z) = \frac{z-1}{z} \mathcal{Z} \left[\frac{G(s)}{s} \right] = (1-z^{-1}) \mathcal{Z} \left[\frac{G(s)}{s} \right] \quad (25.53)$$

Proof. The output $u(t)$ of the ZOH is given by a sequence of steps:

$$u(t) = u_k, \quad kT_s \leq t < (k+1)T_s \quad (25.54)$$

This can be written as

$$u(t) = \sum_{k=0}^{+\infty} u_k \left(H(t - kT_s) - H(t - (k+1)T_s) \right) \quad (25.55)$$

Thus

$$\begin{aligned} u(s) = \mathcal{L}[u(t)] &= \sum_{k=0}^{+\infty} u_k \left(\frac{1}{s} e^{-kT_s s} - \frac{1}{s} e^{-(k+1)T_s s} \right) \\ &= \underbrace{\sum_{k=0}^{+\infty} u_k e^{-kT_s s}}_{\mathcal{Z}[u_k]} \underbrace{\frac{1 - e^{-T_s s}}{s}}_{\text{due to the ZOH}} \end{aligned} \quad (25.56)$$

and

$$\frac{y_k}{u_k} = HG(z) = \mathcal{Z} \left[G(s) \frac{1-z^{-1}}{s} \right] = (1-z^{-1}) \mathcal{Z} \left[\frac{G(s)}{s} \right] \square \quad (25.57)$$

25.4 Choosing the sampling time

The sampling time of a system must be appropriate to the plants and signals involved. It cannot be too large, and should not be too short.

- If the sampling time is too large, the behaviour of the system will not be known. For instance, to study sea waves a sampling time of 10 s would not do, since many waves have shorter periods. These would never be measured.
- If the sampling time is too short, there will be useless measurements, and huge senseless data files. Furthermore, since experimental data is always corrupted with some noise, short sample times mean that more high frequency noise is being acquired. For instance, to study the tides a sampling time of 10 ms would not do, since the tides take hours to rise and fall (with periods of either 12 h or 24 h, depending on the zone of Earth). Most of what would be measured would be high frequency noise of the sensor.

Definition 25.3. If the sampling frequency is $\omega_f = \frac{2\pi}{T_s}$, the **Nyquist frequency** is $\frac{\omega_f}{2} = \frac{\pi}{T_s}$. \square

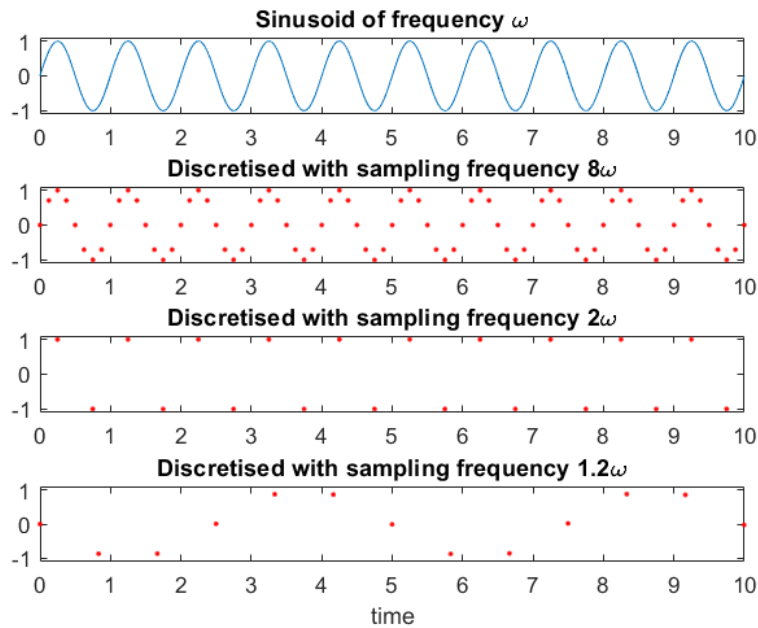


Figure 25.6: A sinusoid, sampled with different sampling periods. Notice how, if the sampling frequency is too small, the frequency of the sinusoid is unrecognisable. It seems to be lower than it is: this phenomenon is called aliasing.

Theorem 25.5. Let $x(t) = \sin \omega t$ be a sinusoid, and x_t be a discrete signal obtained sampling $x(t)$ with sampling time T_s . If the frequency of the sinusoid is above the Nyquist frequency, $\omega > \frac{\pi}{T_s}$, it cannot be found from the discrete signal x_t . \square

We will not present a formal proof, but this result is rather intuitive. It tells us that we need at least two points per period to sample a sinusoid so that its frequency will be recognisable. Otherwise, a phenomenon called *aliasing* will occur, illustrated in Figure 25.6.

Minimum sampling frequency
Aliasing

However, this bare minimum of two sampling instants per period is most undesirable: the amplitude of a sinusoid sampled in this way will surely be wrong. Only with an exceptional luck would the crest and the trough of the sinusoid be sampled, as seen in the first case of Figure 25.7. It would be very likely that the sample instants would match phases other than $\pm 90^\circ$, and thus the amplitude of the sampled sinusoid will be some random value, below the true one, as in the second and third cases of Figure 25.7. And, with an exceptional lack of luck, the last situation of Figure 25.7, in which only the zero crossings are hit, might arise.

In practice it is expedient to have 10 to 20 points per period. So, if the sinusoid has period T and frequency $\omega = \frac{2\pi}{T}$, the sampling time T_s and the sampling frequency $\omega_s = \frac{2\pi}{T_s}$ should verify

Rule of thumb for T_s

$$\frac{T}{20} \leq T_s \leq \frac{T}{10} \Leftrightarrow \frac{2\pi}{20\omega} \leq T_s \leq \frac{2\pi}{10\omega} \Leftrightarrow 20\omega \geq \omega_s \geq 10\omega \quad (25.58)$$

This is a heuristic rule; the lower limit for T_s ensures that the amplitude is never underestimated by more than 5%, as seen in Figure 25.8, which depicts the most unfavourable case possible. It may sometimes be convenient to have an even higher sampling frequency.

Most signals, of course, are not sinusoidal. So rule (25.58) must be applied to the fastest frequency (i.e. the smallest period) found in the signal — or at least the fastest frequency we may be interested in.

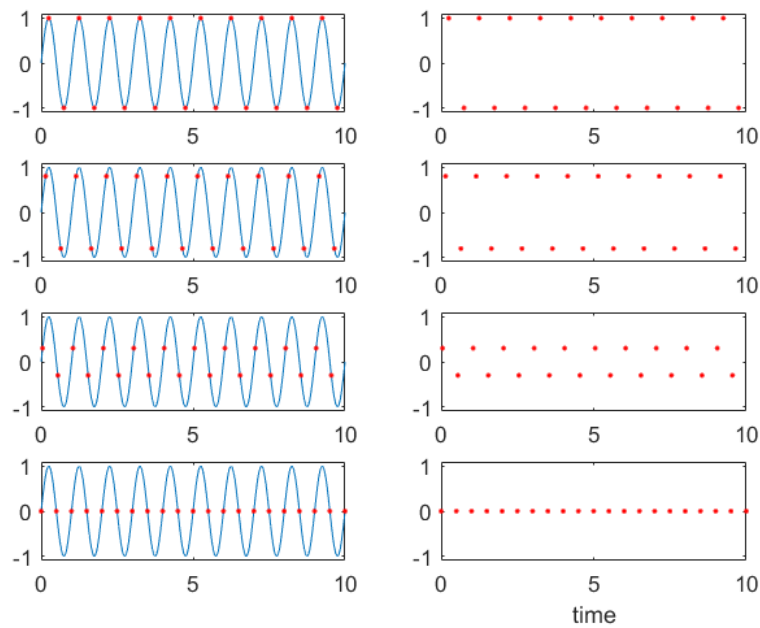


Figure 25.7: A sinusoid of frequency ω is discretised using sampling frequency $\frac{\omega}{2\pi}$. The first sampling time can fall at any phase of the sinusoid, leading to very different discrete signals. Top row: very favourable situation, only possible with exceptional luck. Middle rows: usual situations. Bottom row: very unfavourable situation, only possible with exceptional lack of luck.

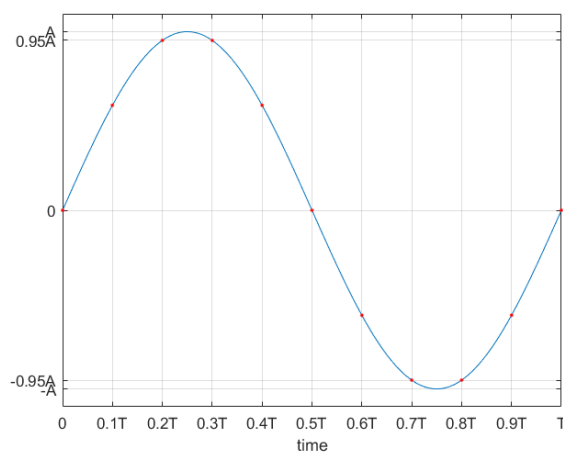


Figure 25.8: A sampling time $T_s = \frac{T}{10}$ ensures that the amplitude A of a sinusoid with period T (and frequency $\frac{2\pi}{T}$) can be found from the discretised signal with an error which never exceeds 5%. The situation depicted is the most unfavourable one as to the phases that sampling instants fall on.

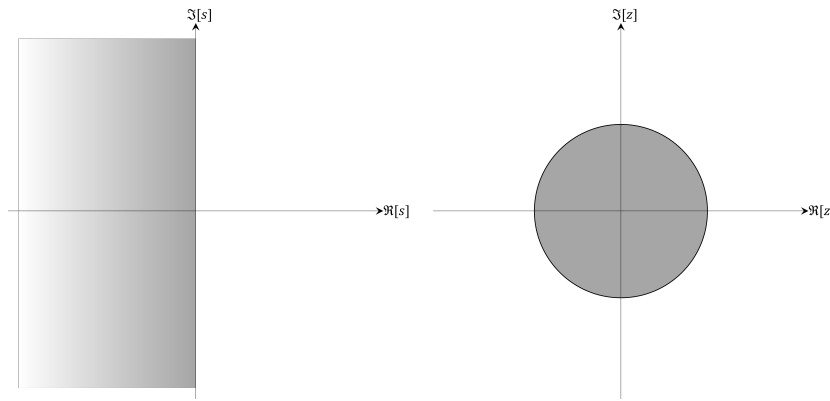


Figure 25.9: Zones where poles are stable. Left: poles in s . Right: poles in z .

25.5 Stability and causality of discrete transfer functions

Let us consider a plant with a pole given by $s = a + jb$. Because

$$z = e^{T_s s} \Leftrightarrow s = \frac{1}{T_s} \log z \tag{25.59}$$

the pole will be mapped to

$$z = e^{T_s a + jT_s b} = \underbrace{e^{T_s a}}_{\text{magnitude}} \underbrace{e^{jT_s b}}_{\text{phase}} \tag{25.60}$$

We see that the pole in z will have

- a magnitude that depends only on the real part of s ,
- a phase that depends only on the imaginary part of s .

Also,

- if pole s is stable, i.e. if $a < 0$, pole z will have a magnitude $e^{T_s a} < 1$, i.e. *Stable poles in z are inside the unit radius circle*
- if pole s is on the imaginary axis, i.e. if $a = 0$, pole z will have a magnitude $e^0 = 1$, i.e. z will be on the unit radius circle;
- if pole s is unstable, i.e. if $a > 0$, pole z will have a magnitude $e^{T_s a} > 1$, i.e. z will be outside the unit radius circle. *Outside the unit radius circle, poles in z are unstable*

In other words,

- poles in s are stable to the left of the imaginary axis, and unstable to the right;
- poles in z are stable inside the unit radius circle, and unstable outside. *Stable poles have $|z| < 1$*

Additionally,

- simple poles in s on the imaginary axis are marginally stable: the same applies to simple poles in z on the unit radius circle;
- multiple poles in s on the imaginary axis are unstable: the same applies to multiple poles in z on the unit radius circle. *Unstable poles have $|z| > 1$*

See Figure 25.9.

Example 25.7. Transfer function $G_1(z) = \frac{1}{z+2}$ has a pole at $z = -2$; since $|z| = 2 > 1$, this is an unstable pole, and $G_1(z)$ is unstable.

Transfer function $G_2(z) = \frac{1}{z-\frac{1}{2}}$ has a pole at $z = \frac{1}{2}$; since $|z| = \frac{1}{2} < 1$, this is an stable pole, and $G_2(z)$ is stable.

Figure 25.10 shows the unit step responses of $G_1(z)$ and $G_2(z)$, by which it can be confirmed that the first is stable and the second is not. \square

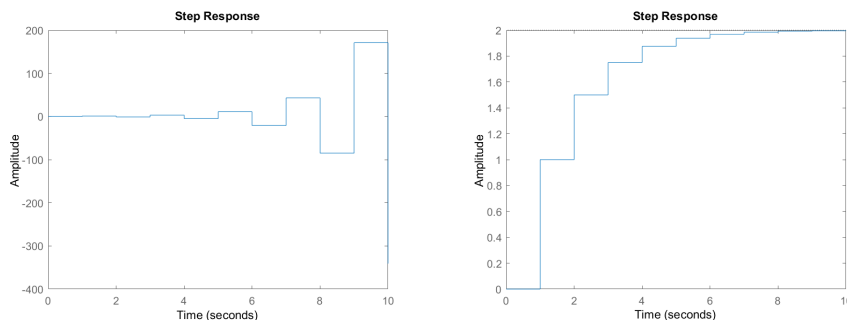


Figure 25.10: Unit step responses of $G_1(z)$ (left) and $G_2(z)$ (right) from Example 25.7. The sample time (which is irrelevant for stability) was assumed to be 1 s.

Notice that what matters are poles in z ; even if a discrete transfer function is given as a ratio of two polynomials in z^{-1} , what matters for stability are the roots of the denominator given as values of z .

Find poles in z , not in z^{-1} .

Example 25.8. The roots of the denominator of $G(z) = \frac{z^{-1}+6}{z^{-2}+5z^{-1}+6}$ are $z^{-1} = 2$ and $z^{-1} = 3$. Because this corresponds to $z = \frac{1}{2}$ and $z = \frac{1}{3}$, we conclude that $G(z)$ is stable. \square

We saw in Chapter 11 that only those transfer functions having more poles than zeros, i.e. which are proper, can be physically possible. The same happens in z ; it is in fact even much clearer in z .

Example 25.9. Let

$$G(z) = \frac{b_2z^2 + b_1z + b_0}{a_1z + 1} \quad (25.61)$$

This transfer function is not proper, since it has two zeros but only one pole. The corresponding difference equation is

$$\begin{aligned} \frac{b_2z^2 + b_1z + b_0}{a_1z + 1} &= \frac{y_k}{u_k} \Leftrightarrow \\ (b_2z^2 + b_1z + b_0)u_k &= (a_1z + 1)y_k \Leftrightarrow \\ b_2u_{k+2} + b_1u_{k+1} + b_0u_k &= a_1y_{k+1} + y_k \Leftrightarrow \\ y_{k+1} &= \frac{b_2}{a_1}u_{k+2} + \frac{b_1}{a_1}u_{k+1} + \frac{b_0}{a_1}u_k - \frac{1}{a_1}y_k \Leftrightarrow \\ y_k &= \frac{b_2}{a_1}u_{k+1} + \frac{b_1}{a_1}u_k + \frac{b_0}{a_1}u_{k-1} - \frac{1}{a_1}y_{k-1} \end{aligned} \quad (25.62)$$

The output y_k depends on a future input, u_{k+1} . \square

This Example illustrates the following result:

Theorem 25.6. If a discrete transfer function has m zeros and n poles, such that $m > n$, then the output y_k depends from future inputs $u_{k+1}, \dots, u_{k+m-n}$. \square

Causality

A system is

- **causal**, if its output does not depend from future inputs;
- **non-causal**, if its output depends from future inputs.

So, discrete transfer functions that are not proper are non-causal. By extension, continuous transfer functions that are not proper are called non-causal too.

Remark 25.1. The number of zeros and poles that matters is found in z , not in z^{-1} . We might be tempted to say that $H(z) = \frac{\frac{1}{3}z^{-2} + \frac{1}{3}z^{-1} + \frac{1}{3}}{1}$ has a polynomial of order 2 in the numerator and a polynomial of order 0 in the denominator;

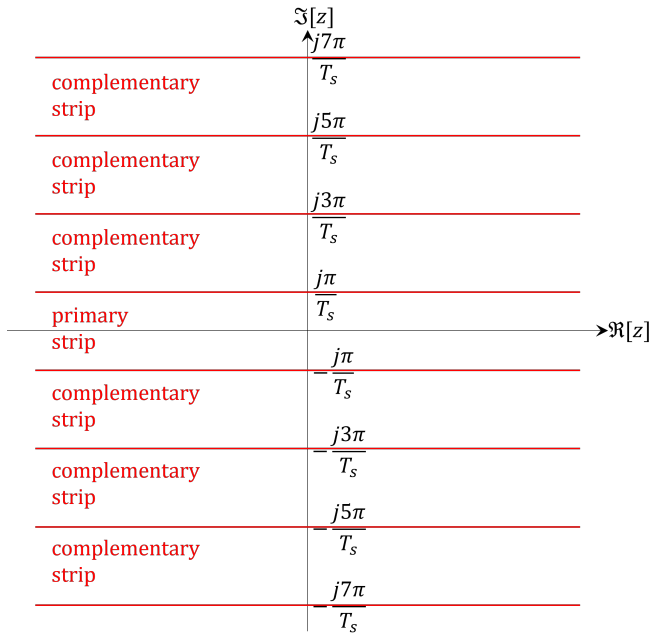


Figure 25.11: Strips of the complex plane where s is represented that are mapped onto the entire complex plane where z is represented.

of course, with 2 zeros and 0 poles it would not be causal. But its difference equation is

$$\frac{1}{3}z^{-2} + \frac{1}{3}z^{-1} + \frac{1}{3} = \frac{y_k}{u_k} \Rightarrow y_k = \frac{1}{3}u_k + \frac{1}{3}u_{k-1} + \frac{1}{3}u_{k-2} \quad (25.63)$$

and it is clear that the output depends on the current and on past inputs, not on anything future. We should have considered $H(z) = \frac{\frac{1}{3} + \frac{1}{3}z + \frac{1}{3}z^2}{z^2}$ and noticed that the transfer function has in fact 2 poles and 2 poles, and thus no problems of causality. \square

25.6 Primary and complementary strips in s

Since $s = a + jb$ is mapped to $z = e^{T_s a} e^{jT_s b} = e^{T_s a} \cos(T_s b) + j e^{T_s a} \sin(T_s b)$, and the cosine and sine functions are periodical, it is clear that all points of the form

$$s = a + j \left(b + \frac{2k\pi}{T_s} \right), \quad k \in \mathbb{Z} \quad (25.64)$$

will be mapped to very same point

$$\begin{aligned} z &= e^{T_s a} \cos \left(T_s \left(b + \frac{2k\pi}{T_s} \right) \right) + j e^{T_s a} \sin \left(T_s \left(b + \frac{2k\pi}{T_s} \right) \right) \\ &= e^{T_s a} \cos (T_s b) + j e^{T_s a} \sin (T_s b) \end{aligned} \quad (25.65)$$

This means that the complex plane where s is represented can be split into an infinite number of horizontal strips, of width $\frac{2\pi}{T_s}$, as seen in Figure 25.11. Each of them is mapped onto the entire complex plane where z is represented.

Each strip in s mapped to the entire plan in z

The existence of infinite points s mapped to the same value of z is a consequence of Theorem 25.5. Many sinusoids, when sampled in time, result in the same discrete signal, as seen in Figure 25.12. Of course, only one will be properly sampled: the one with a frequency below the Nyquist frequency, i.e. the one found inside the strip of the complex plane centred on the real axis, from $-j\frac{\pi}{T_s}$ to $+j\frac{\pi}{T_s}$. For this reason, this strip is called **primary strip**, and the others, where sinusoids have frequencies that cannot be properly sampled, are called **complementary strips**.

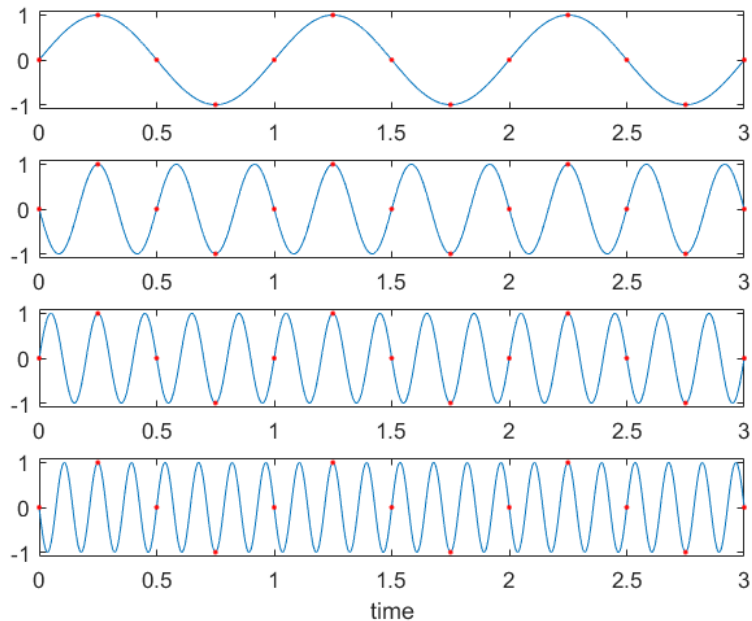


Figure 25.12: Different sinusoids sampled with the same sampling time can result in the same discrete signal.

Glossary

They were really getting quite fond of their strange pet and hoped that Aslan would allow them to keep it. The cleverer ones were quite sure by now that at least some of the noises which came out of his mouth had a meaning. They christened him Brandy because he made that noise so often.

C. S. LEWIS (1868 — †1963), *The Magician's nephew* (1955), 14

complementary strip faixa complementar

delay operator operador de atraso

forward operator operador de avanço

primary strip faixa primária

Programmable Logic Controller (PLC) autómato programável (AP)

zero order hold retentor de ordem zero

Exercises

- In a closed loop control system, the error e and the control action u are electrical signals. Design analog circuits to implement:
 - A proportional controller $\frac{u}{e} = 10$, using one OpAmp and two resistors.
 - A proportional controller $\frac{u}{e} = 0.5$, using two OpAmps and four resistors. Why is it that in this case one OpAmp is not enough?
 - A PD controller $\frac{u}{e} = 5(1 + 0.1s)$.
 - A PID controller $\frac{u}{e} = 5\left(1 + \frac{1}{2s} + 0.1s\right)$.
 - A generic PID controller $\frac{u}{e} = K_p\left(1 + \frac{1}{T_i s} + T_d s\right)$.
 - A lead controller $\frac{u}{e} = \frac{50s+0.1}{s+0.1}$.
 - A lag controller $\frac{u}{e} = \frac{s+10}{s+0.5}$.
 - A generic lead-lag controller $\frac{u}{e} = K\frac{s+b}{s+p}$.
- Discuss if a 100 ms sampling time is appropriate for the following systems, and propose a different value when it is not:

- (a) A system with a 100 rad/s open-loop bandwidth.
 (b) A system with a 5 rad/s open-loop bandwidth.
 (c) A system with a 0.3 rad/s open-loop bandwidth.
 (d) A tidal energy converter.
 (e) The controller of the opening of the valves in the engine of a car.
 (f) The controller of the opening of the valves in the diesel engine of a ship running at 80 rpm.
3. Knowing that $G(z) = \frac{z^2-3}{(z-1)^2}$ and that the sampling time is 100 ms, find $x(t)$ for $t = 0$ s, 0.1 s, 0.2 s, ... 2 s, when
- (a) the input is a unit impulse;
 (b) the input is a unit step;
 (c) the input is $u_0 = 1$, $u_1 = 2$, and 0 otherwise.
4. Consider a signal with \mathcal{Z} transform given by $\mathcal{Z}\{x(t)\} = X(z) = \frac{z}{(z+1)(z-1)}$. Let T_s be the sampling time. Therefore:
- A)** $x(t) = 1 - (-1)^{nT_s}$, $n = 0, 1, 2, \dots$
B) $x(t) = \frac{1}{2}[1 - (-1)^{nT_s}]$, $n = 0, 1, 2, \dots$
C) $x(t) = 2[1 - (-1)^{nT_s}]$, $n = 0, 1, 2, \dots$
D) None of the above.
5. Consider signal $x(t)$, such that $X(s) = \mathcal{L}\{x(t)\} = \frac{s+2}{(s+1)^2(s+3)}$.
- (a) Obtain $X(z) = \mathcal{Z}\{x(t)\}$ for a generic sampling time h .
 (b) Do the same for $h = 1$ s.
 (c) Do the same for a sampling frequency of 10 Hz.
6. Consider a signal with \mathcal{Z} transform given by $\mathcal{Z}\{x(t)\} = \frac{2z(z - \frac{5}{12})}{(z - \frac{1}{2})(z - \frac{1}{3})}$.
- (a) Its inverse transform is:
- A)** $x(t) = \left(\frac{1}{2}\right)^k + \left(\frac{1}{3}\right)^k$, $k = 0, 1, 2, \dots$
B) $x(t) = \left(\frac{1}{2}\right)^k + \left(\frac{1}{3}\right)^k - \left(\frac{5}{12}\right)^k$, $k = 0, 1, 2, \dots$
C) $x(t) = \left(\frac{5}{12}\right)^k - \left(\frac{1}{2}\right)^k - \left(\frac{1}{3}\right)^k$, $k = 0, 1, 2, \dots$
D) None of the above.
- (b) Let h be the sampling period. From the transform definition, it can be concluded that:
- E)** $X(z) = 2 + \frac{5}{6}z^{-1} + \frac{13}{36}z^{-2} + \dots \Rightarrow x(0) = 2$, $x(h) = \frac{5}{6}$, $x(2h) = \frac{13}{36}, \dots$
F) $X(z) = 1 + \frac{1}{2}z^{-1} + \frac{1}{3}z^{-2} + \dots \Rightarrow x(0) = 1$, $x(h) = \frac{1}{2}$, $x(2h) = \frac{1}{3}, \dots$
G) $X(z) = \frac{5}{12} - \frac{1}{2}z^{-1} - \frac{1}{3}z^{-2} + \dots \Rightarrow x(0) = \frac{5}{12}$, $x(h) = -\frac{1}{2}$, $x(2h) = -\frac{1}{3}, \dots$
H) None of the above.
7. Given $X(z) = \frac{z}{(z-1)^2(z-2)}$ determine $x(kh)$, $k = 0, 1, 2, \dots$, where h is the sampling period.
8. Present a formal proof of Theorem 25.6, following the reasoning in Example 25.9.

Chapter 26

Digital approximations of continuous systems

In this chapter

26.1 Using the \mathcal{Z} transform

\mathcal{Z}

26.2 Mapping poles and zeros

map

26.3 Backward first order approximation

back

26.4 Forward first order approximation

forward

26.5 The Tustin approximation

Tustin

26.6 Approximating controllers

This chapter is still being written. In the picture: National Pantheon, or Church of Saint Engratia, Lisbon (source: <http://www.panteaonacional.gov.pt/171-2/historia-2/>), somewhen during its construction (1682–1966).



Glossary

Pirene leaned over the table to get a better view and Hardin continued: “The message from Anacreon was a simple problem, naturally, for the men who wrote it were men of action rather than men of words. It boils down easily and straightforwardly to the unqualified statement, when in symbols is what you see, and which in words, roughly translated, is, ‘You give us what we want in a week, or we take it by force.’ ”

Isaac ASIMOV (1920 — †1992), *Foundation*, II 5 (The Encyclopedists, *Astounding Science-Fiction*, May 1942)

word in English word in Portuguese
palavra em inglês palavra em português

Exercises

1. The PD controller $C(s) = 8s + 32$ has to be implemented with a sampling time of 0.02 s. Find the difference equation that should be implemented in a microprocessor.
2. The roll of the ship from Exercise 5 of Chapter 21 is going to be controlled with a PID controller $C(s) = 2s + 11 + \frac{1}{s}$, implemented with a sampling time of 10 ms. Find the different digital implementations of the controller that result from using different approximations, implement them using Matlab commands, and verify which approximation is the best. The system to be controlled is given by $G(s) = \frac{9}{s^2 + 1.2s + 9}$.

Chapter 27

Study and control of digital systems

Weston was pale and haggard from a night of calculations intricate enough to tax any mathematician even if his life did not hang on them.

C. S. LEWIS (1868 — †1963), *Out of the silent planet* (1938), 21

This chapter is still being written.

27.1 Block diagrams

In the picture: National Pantheon, or Church of Saint Engratia, Lisbon (source: <http://www.panteaonacional.gov.pt/171-2/historia-2/>), somewhen during its construction (1682–1966).

27.2 Pole placement

This chapter is still being written.

27.3 Steady state errors



27.4 Root locus

In the picture: National Pantheon, or Church of Saint Engratia, Lisbon (source: <http://www.panteaonacional.gov.pt/171-2/historia-2/>), somewhen during its construction (1682–1966).

27.5 Jury and Routh-Hurwitz criteria

This chapter is still being written.

27.6 Frequency responses

In the picture: National Pantheon, or Church of Saint Engratia, Lisbon (source: <http://www.panteaonacional.gov.pt/171-2/historia-2/>), somewhen during its construction (1682–1966).

27.7 Studying stability from frequency responses



Glossary

“But I think I’ve learnt to manage these donkeys, Fred.” I leant forward and quietly gave the order to turn left into Peter’s large hairy ear. “*Zur linken Zeite, mein Freund.*”

Peter veered to the left, barging into Paul and very nearly unseating Fred. We would have left the track altogether and started across the Forest had I not quickly murmured to him in German to go right again and then straight on.

“It’s quite simple,” I explained to Fred. “Mr. Mattock said that Aunt Lot was always muttering in German. That’s how she talked to the donkeys. We’ll be all right from now on.”

“You mean *you’ll* be all right,” grumbled Fred. “My subject’s History, not German. You’ll have to talk for us both.”

John PUDNEY (1909 — †1977), *Spring adventure* (1961), 1

word in English word in Portuguese
palavra em inglês palavra em português

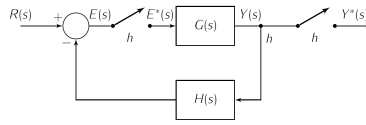


Figure 27.1: Block diagram of Exercise 1.

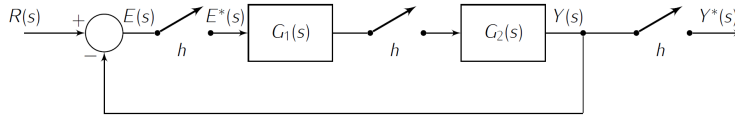


Figure 27.2: Block diagram of Exercise 2.

Exercises

1. Consider the system shown in Figure 27.1 with sampling period h . Is the closed-loop transfer function given by $\frac{Y(z)}{R(z)} = \frac{G(z)}{1 + G(z)H(z)}$, by $\frac{Y(z)}{R(z)} = \frac{G(z)}{1 + GH(z)}$, or by neither of these transfer functions?
2. Consider the system shown in Figure 27.2 with sampling period h . Which of the following transfer functions corresponds to this closed-loop?
 - (a) $\frac{Y(z)}{R(z)} = \frac{G_1G_2(z)}{1 + G_1G_2(z)}$
 - (b) $\frac{Y(z)}{R(z)} = \frac{G_1(z)G_2(z)}{1 + G_1G_2(z)}$
 - (c) $\frac{Y(z)}{R(z)} = \frac{G_1G_2(z)}{1 + G_1(z)G_2(z)}$
 - (d) $\frac{Y(z)}{R(z)} = \frac{G_1(z)G_2(z)}{1 + G_1(z)G_2(z)}$
3. The discretised transfer function relating the voltage $u(t)$ applied to the motor with the azimuth angle $\theta(t)$ of the antenna in Exercise 2 of Chapter 22 is $HG(z) = \frac{0.003z + 0.003}{z^2 - 1.975z + 0.975}$, with a sampling time of 0.25 s. Figure 27.3 shows the Bode diagram of the open loop formed by the plant and by controller $C(z) = \frac{30.361(z - 0.882)}{z - 0.286}$.
 - (a) Is the closed loop stable?
 - (b) Can this controller follow constant angle references?
4. The discretised transfer function relating the depth of facing (output) performed by the lathe of Exercise 4 from Chapter 22 with the control

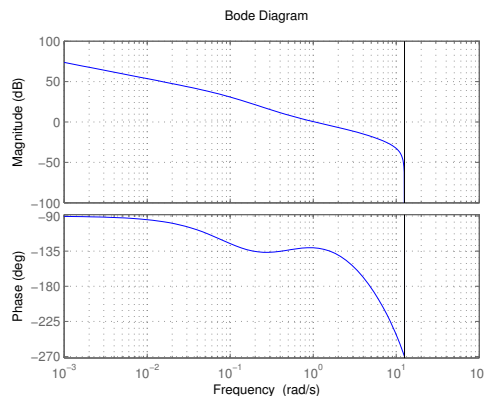


Figure 27.3: Bode diagram of Exercise 3.

action of the motor (input) is

$$G_p(z) = \frac{0.0477(z + 1.65)(z + 0.0788)}{(z - 1)(z - 0.4966)(z - 0.0302)} \quad (27.1)$$

with a 0.7 s sampling time.

- (a) Knowing that the bandwidth of the closed loop was estimated as 0.622 rad/s, explain if the sampling time is reasonable or not.
 - (b) Find the values of a proportional controller that stabilises this plant.
5. A helium-filled spherical balloon is to be designed so that its weight is compensated by its hydrostatic lift, and thus its height $h(t)$ will be given by

$$G(s) = \frac{h(s)}{u(s)} = \frac{1}{s(10s + 1)} \quad (27.2)$$

where $u(t)$ is the command action of the vertical thrust of its propellers.

- (a) Find the differential equation that models the plant, considering only a mass m , a buoyancy B , a weight W , a drag D and a propulsion T .
 - (b) Let $m = 300$ kg, let the drag coefficient be $\beta = 30$ N s/m, and let the densities of the fluids be $\rho_{\text{air}} = 1$ kg/m³ and $\rho_{\text{He}} = 0.2$ kg/m³. Find the radius of the balloon.
 - (c) Let $T(t) = 30u(t)$. Prove $G(s)$ and find its discrete equivalent, assuming a 0.8 s sampling time and a zero-order hold.
 - (d) Study the stability of the discrete closed-loop using the root-locus method.
 - (e) Find a digital controller that ensures closed-loop poles at 0.9 rad/s and $0.266 \pm 0.33j$ rad/s.
 - (f) Check if this controller achieves or not a 10 dB attenuation for a 0.02 rad/s disturbance.
 - (g) Study the closed loop's stability using the Nyquist diagram.
6. The transfer function of a robot is

$$G(s) = \frac{1.6}{(s + 1)^2} \quad (27.3)$$

- (a) Find the values of proportional controllers that stabilise the plant.
 - (b) Show that if there is a zero-order hold with 0.1 s sampling time in the control loop then the transfer function of the plant is $HG(z) = \frac{0.0075z + 0.007}{z^2 - 1.81z + 0.8187}$.
 - (c) Plot the root locus of $HG(z)$.
 - (d) Find the values of proportional controllers that stabilise the discrete plant.
 - (e) Compare the values of the proportional controller that can be used in both the continuous and the discrete cases.
7. Consider an inverted pendulum, that deviates from the vertical by an angle $\phi(t)$, and is controlled by a control action $u(t)$ that moves the base of the pendulum sideways (see Figure 27.4). The corresponding (linearised) transfer function is

$$G(s) = \frac{\Phi(s)}{U(s)} = \frac{5s}{(s - 5)(s + 5)(s + 0.2)} = \frac{5s}{s^3 + 0.2s^2 - 25s - 5} \quad (27.4)$$

- (a) Show that the discrete equivalent of (27.4), when a zero-order hold with a 200 Hz sampling frequency is applied at the input of the plant, is

$$HG(z) = \frac{6.2482 \times 10^{-5}(z - 1)(z + 1)}{(z - 1.025)(z - 0.999)(z - 0.9753)} \quad (27.5)$$

- (b) Figure 27.5 shows the Nyquist plot of (27.5). Show that (27.5) cannot be stabilised with a proportional controller.



Figure 27.4: Inverted pendulum in the Laboratory of Control and Robotics; its parameters are not those of transfer function (27.4).

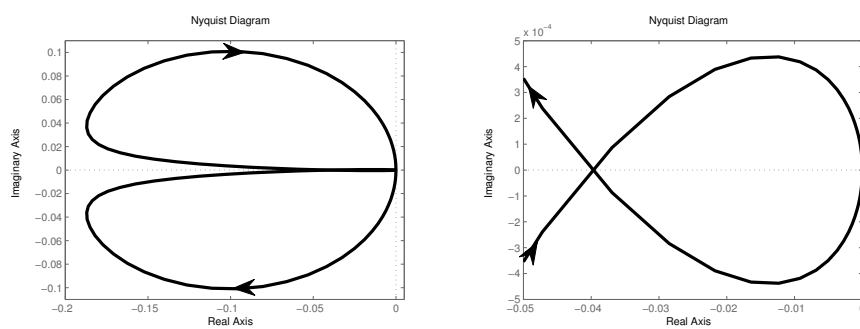


Figure 27.5: Nyquist diagram of (?); right plot: detail of the left plot.

Chapter 28

Non-linearities in control systems

This chapter concerns the effects of non-linearities in control systems. Non-linearities can stem:

- **From the plant.** We saw in Chapter 8 that soft non-linearities can be linearised, but of course that is only an approximation.
- **From the actuator.** While the actuator is often considered as part of the plant, it is better to consider it a separate model not only if it can be easily replaced but also if it is the source of non-linear behaviour. One particular non-linearity, saturation, is inevitable in any actuator, since no actuator can provide an arbitrarily large control action (this would always require infinite energy). At most, this non-linearity can be neglected if it is certain that in no case the controller will ever require a control action too large for the actuator. In mechatronic systems, mechanical parts often originate non-linearities such as dead zones (remember Figure 8.5) or backlash (remember Figure 4.16).
- **From the sensor.** As we saw in Chapter 13, there are sensors with non-linear transductions, which can be accounted for (in which case it is as if they did not exist) or linearised.
- **From the control law itself.** Design strategies of non-linear control laws fall outside the scope of these Lecture Notes; only one particular case of one particular strategy (that of reset control) will be presented below in Chapter 29.

The study of their effect is easier in continuous time (i.e. using transfer functions in s , not in z); in any case, the effects of non-linearities are similar in discrete time as well.

28.1 Non-linearities and responses in time

The effect of non-linearities in time responses of open-loop control systems is fairly easy to determine. In closed-loop, because the effect of the non-linearity is fed back, this is not that easy, but can be done using the **equivalent gain** K_{eq} , *Equivalent gain* defined as the ratio between the output y and the input x of the non-linearity (having hence no dimensions):

$$K_{eq} = \frac{y}{x} \quad (28.1)$$

Example 28.1. Consider the non-linearity described in Figure 28.1, combining a dead-zone with saturation:

$$y = \begin{cases} 0, & \text{if } -1 < x < 1 \\ x, & \text{if } -3 \leq x \leq -1 \vee 1 \leq x \leq 3 \\ -3, & \text{if } x < -3 \\ 3, & \text{if } x > 3 \end{cases} \quad (28.2)$$

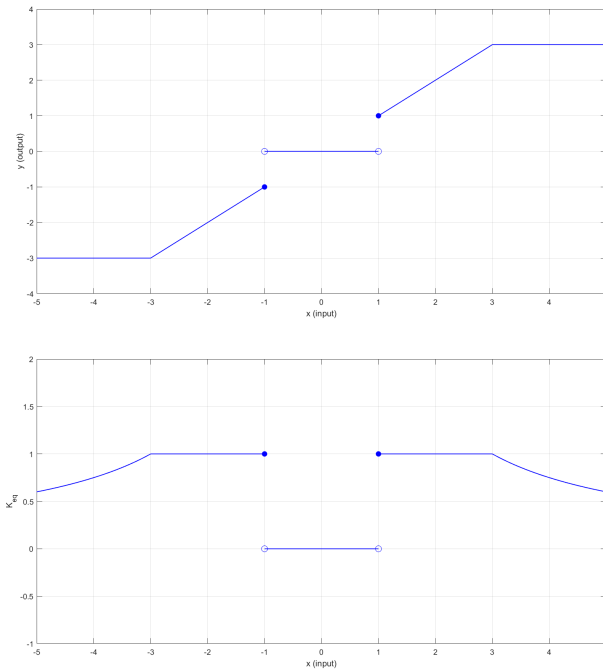


Figure 28.1: Non-linearity (left) and its equivalent gain (right).

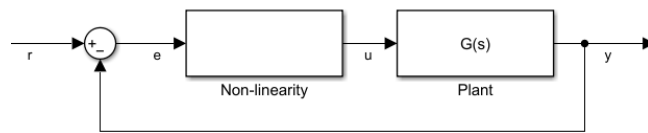


Figure 28.2: Control loop with one non-linearity.

The equivalent gain, also shown in the Figure, is

$$y = \begin{cases} \frac{0}{x}, & \text{if } -1 < x_1 \\ \frac{x}{x}, & \text{if } -3 \leq x \leq -1 \vee 1 \leq x \leq 3 \\ \frac{-3}{x}, & \text{if } x < -3 \\ \frac{3}{x}, & \text{if } x > 3 \end{cases} = \begin{cases} 0, & \text{if } -1 < x_1 \\ 1, & \text{if } -3 \leq x \leq -1 \vee 1 \leq x \leq 3 \\ \frac{3}{|x|}, & \text{if } x < -3 \vee x > 3 \end{cases} \quad (28.3)$$

Notice that, for input $x = 0$, the equivalent gain was, of course, extended by continuity. \square

Equivalent gain and root-locus

If a closed-loop configuration such as the one seen in Figure 28.2 can be found, in which the non-linearity affects the error, then the equivalent gain must be combined with a means of studying the influence of the gain in closed-loop stability, such as the root-locus diagram and the Routh-Hurwitz criterion, to assess the influence of the non-linearity in the control loop.

Example 28.2. In the control system of Figure 28.3,

$$G(s) = \frac{s+5}{s^2} \quad (28.4)$$

The root-locus of (28.4) is given in Figure 28.4. We now consider two cases, both shown in Figure 28.5:

- When the reference is $r = 1$, the initial value of the error is $e(0) = 1$ (and then the error decreases). The control action never saturates, so we always have $K_{eq} = 1$. The response is oscillatory, with an overshoot only slightly larger than what might be expected from the position of the closed loop poles. (Remember that the discrepancy is not surprising since the closed loop is *not* a second order system without zeros.)

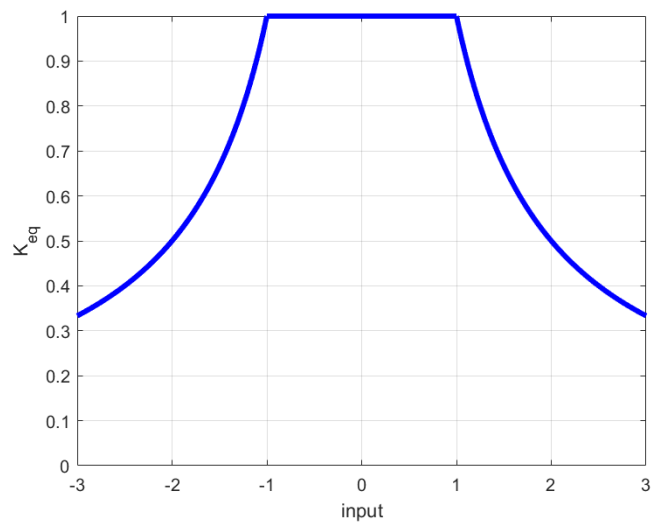
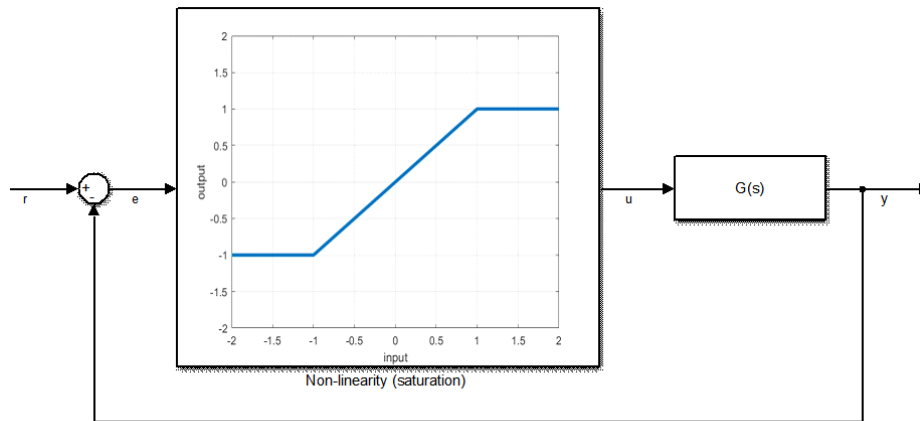


Figure 28.3: Control loop of Examples 28.2 and 28.3 (top), and equivalent gain of the saturation (bottom).

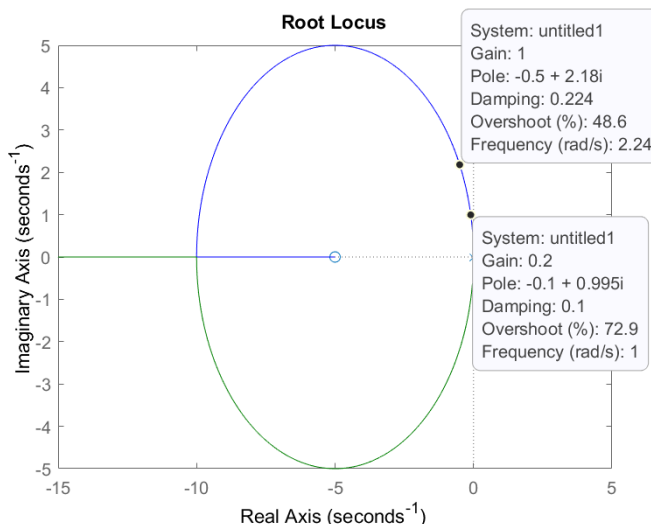


Figure 28.4: Root locus of (28.4) from Example 28.2.

- When the reference is $r = 5$, the initial value of the error is $e(0) = 5$, the control action saturates at $u(0) = 1$, and so $K_{eq} = \frac{1}{5} = 0.2$. In Figure 28.4 we see that this corresponds to a more oscillatory behaviour, as expected from the poles closer to the imaginary axis; the overshoot is somewhat lower than, but in line with, the expected value. \square

Example 28.3. If in the control system of Figure 28.3 the system is rather

$$G(s) = \frac{s + 5}{s(s^2 + 0.25s + 1)} \quad (28.5)$$

with the root-locus given in Figure 28.6, the values of the gain ensuring closed-loop stability are found as follows:

$$\begin{aligned} \frac{y(t)}{r(t)} &= \frac{K \frac{s+5}{s^3+0.25s^2+s}}{1 + K \frac{s+5}{s^3+0.25s^2+s}} \\ &= \frac{Ks + 5K}{s^3 + 0.25s^2 + s(K+1) + 5K} \end{aligned} \quad (28.6)$$

$$\begin{array}{c|cc} s^3 & 1 & K+1 \\ s^2 & 0.25 & 5K \\ s & K+1 - \frac{5K}{0.25} & \\ 1 & 5K & \end{array} \quad (28.7)$$

$$\begin{cases} K+1 - 20K > 0 \\ 5K > 0 \end{cases} \Rightarrow \begin{cases} K < \frac{1}{19} = 0.0526 \\ K > 0 \end{cases} \quad (28.8)$$

We now consider two cases as well, both shown in Figure 28.7:

Limit cycle

- When the reference is $r = 1$, the initial value of the error is $e(0) = 1$. The control action is limited to 1, so the equivalent gain is $\frac{1}{1}$, and thus the system becomes unstable. However, in the Figure we see that the amplitude of the oscillations does not grow indefinitely, as might be expected. This is because of the saturation. As the error grows, since the control action is limited to $[-1, 1]$, the equivalent gain becomes smaller, and the loop eventually falls back into the zone of stability. As the error decreases, the equivalent gain increases, and the control loop becomes unstable again. Because of this switching between a stable and an unstable situation, around the limit of stability, the oscillations end up with a constant amplitude. This is called a **limit cycle**.

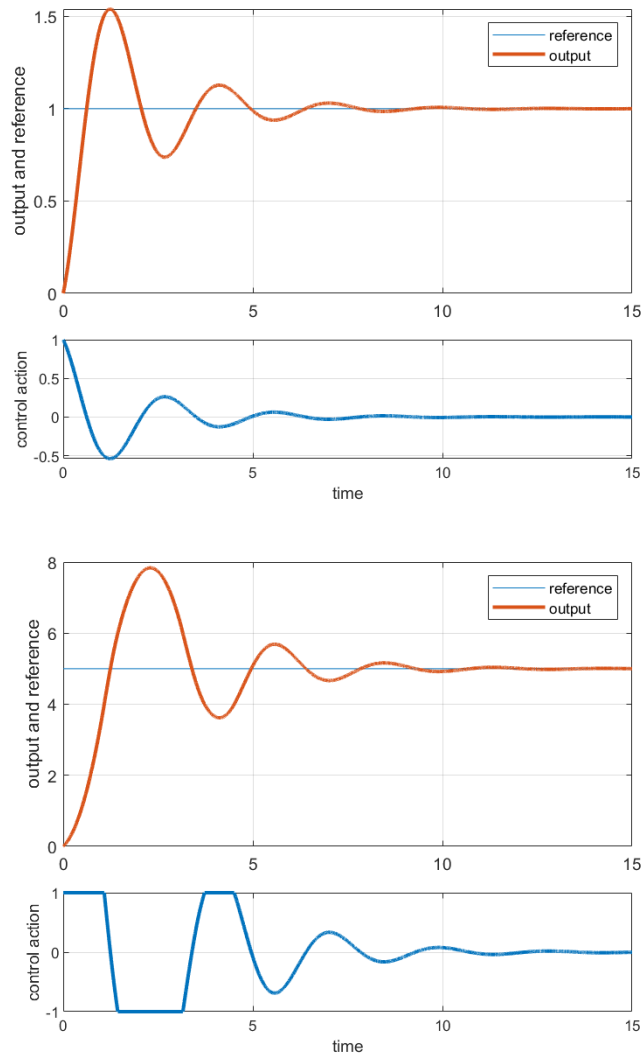


Figure 28.5: Example 28.2: reference 1 (top) and reference 5 (bottom).

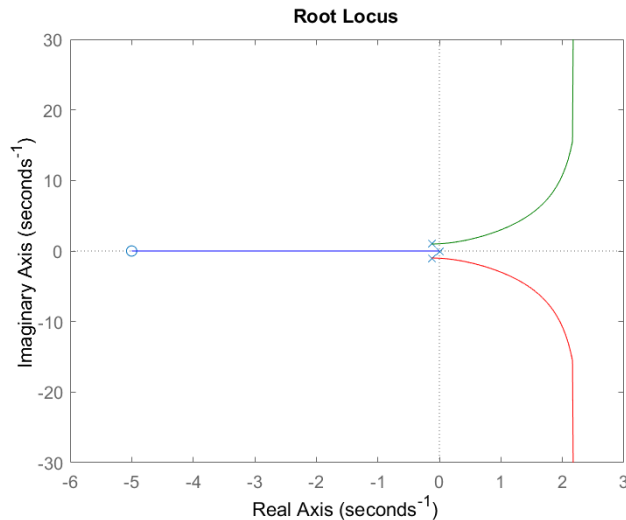


Figure 28.6: Root locus of (28.5) from Example 28.3.

- When the reference is $r = 100$, the initial value of the error is $e(0) = 100$. The control action is limited to 1, so the equivalent gain is $\frac{1}{100}$, and the system is stable. As the error becomes smaller, the equivalent gain increases, the system becomes unstable, and we have a limit cycle again.

Notice that the amplitude of the oscillations is the same in both cases of Figure 28.7. We will see below how it can be found. But there is a significant difference between the two cases. When $r = 1$, the amplitude of the oscillations of the limit cycle is so large in relation to r that the situation is in practice as bad as if the loop were unstable. When $r = 100$, the amplitude of the oscillations, not being neglectable, may in some cases be tolerable. Hence the importance of estimating limit cycles, and their amplitudes, caused by non-linearities. \square

Remark 28.1. Complex conjugate poles on the imaginary axis can also originate oscillations with a constant amplitude, but a limit cycle is caused by a non-linearity instead. \square

28.2 Describing function

To better study limit cycles, which originate oscillations, it is necessary to study the influence of non-linearities in the frequency response of control loops. This is at first sight impossible: if there is a non-linearity, the output of the control system for a sinusoidal input (i.e. reference) is no longer a sinusoid with the same period of the input. Hence it is impossible to speak about a gain and a phase, as when linear systems are studied. But it is possible to find an approximate gain and an approximate phase, using a Fourier series expansion, if the non-linear output $y(t)$ for a sinusoidal input $u(t) = U \sin(\omega t)$ (hence with period $T = \frac{2\pi}{\omega}$) is

A describing function has an approximate gain and an approximate phase

- periodic, with period T :

$$T = \inf \left\{ \tilde{T} : y(t + \tilde{T}) = y(t), \forall t \right\} \quad (28.9)$$

(in other words, T is shortest period of time after which $y(t)$ repeats itself);

- even, in the sense that

$$y\left(t + \frac{T}{2}\right) = -y(t), \forall t \quad (28.10)$$

Fourier series expansion

Theorem 28.1. Let $f(x)$ be a limited, periodic function, with period 2π , and a finite number of maxima, minima and discontinuities. Then the Fourier series

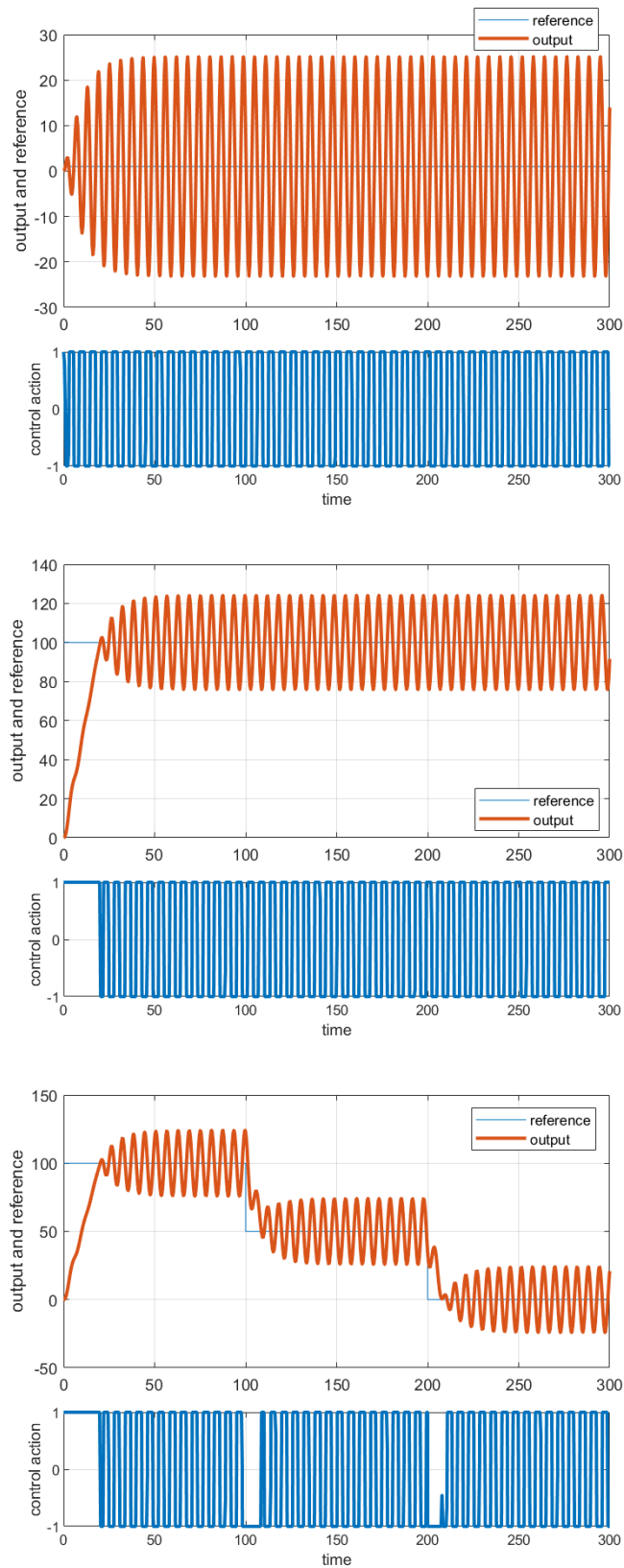


Figure 28.7: Example 28.3: reference 1 (top), reference 100 (centre), reference consisting of three steps (bottom).

$f(x)$ given by

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{+\infty} a_k \cos kx + b_k \sin kx \quad (28.11)$$

$$a_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos kx \, dx \quad (28.12)$$

$$b_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin kx \, dx \quad (28.13)$$

converges in the domain of f and verifies

$$f(t) = \frac{\lim_{\tau \rightarrow x^-} f(\tau) + \lim_{\tau \rightarrow x^+} f(\tau)}{2} \quad \square \quad (28.14)$$

Proof. The proof can be found in any textbook of Calculus. \square

Corollary 28.1. Suppose that $f(x)$ has an arbitrary period T , and that $\omega = \frac{2\pi}{T} \Leftrightarrow T = \frac{2\pi}{\omega}$. Then function $g(x) = f\left(\frac{T}{2\pi}x\right) = f\left(\frac{x}{\omega}\right)$ has period 2π , and we can write

$$g(x) = f\left(\frac{x}{\omega}\right) = \frac{a_0}{2} + \sum_{k=1}^{+\infty} a_k \cos kx + b_k \sin kx \quad (28.15)$$

$$a_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos kx \, dx \quad (28.16)$$

$$b_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin kx \, dx \quad (28.17)$$

Making $t = \frac{x}{\omega} \Leftrightarrow x = \omega t \Rightarrow dx = \omega dt$, we get

$$f(t) = \frac{a_0}{2} + \sum_{k=1}^{+\infty} a_k \cos k\omega x + b_k \sin k\omega x \quad (28.18)$$

$$a_k = \frac{\omega}{\pi} \int_0^{\frac{2\pi}{\omega}} f(t) \cos k\omega t \, dt \quad (28.19)$$

$$b_k = \frac{\omega}{\pi} \int_0^{\frac{2\pi}{\omega}} f(t) \sin k\omega t \, dt \quad \square \quad (28.20)$$

In our case, from (28.10), (28.19) and the definition of ω comes

$$\begin{aligned} a_0 &= \frac{2}{T} \int_0^T y(t) \, dt \\ &= \frac{2}{T} \int_0^{\frac{T}{2}} y(t) \, dt + \frac{2}{T} \underbrace{\int_{\frac{T}{2}}^T y(t) \, dt}_{-\int_0^{T/2} y(t) \, dt} = 0 \end{aligned} \quad (28.21)$$

and so we make

$$y(t) \approx a_1 \cos \omega t + b_1 \sin \omega t \quad (28.22)$$

We now make use of the following lemma:

Lemma 28.1.

$$A \cos(x) + B \sin(x) = \sqrt{A^2 + B^2} \sin\left(x + \arctan \frac{A}{B}\right) \quad (28.23)$$

Proof.

$$\begin{aligned} &\sqrt{A^2 + B^2} \sin\left(x + \arctan \frac{A}{B}\right) \\ &= \sqrt{A^2 + B^2} \sin x \cos \arctan \frac{A}{B} + \sqrt{A^2 + B^2} \cos x \sin \arctan \frac{A}{B} \\ &= \sqrt{A^2 + B^2} \sin x \frac{B}{\sqrt{A^2 + B^2}} + \sqrt{A^2 + B^2} \cos x \frac{A}{\sqrt{A^2 + B^2}} \quad \square \end{aligned} \quad (28.24)$$

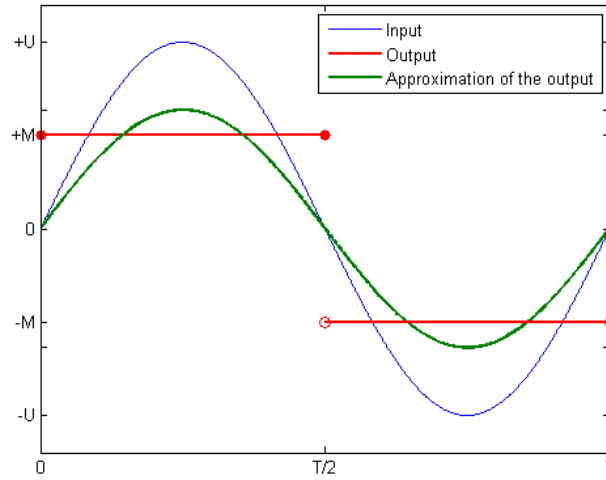


Figure 28.8: On-off non-linearity.

From (28.22) and (28.23), we conclude that, if we let

$$a = \frac{\omega}{\pi} \int_0^{\frac{2\pi}{\omega}} y(t) \cos \omega t \, dt = \frac{2}{T} \int_0^T y(t) \cos \frac{2\pi t}{T} \, dt \quad (28.25)$$

$$b = \frac{\omega}{\pi} \int_0^{\frac{2\pi}{\omega}} y(t) \sin \omega t \, dt = \frac{2}{T} \int_0^T y(t) \sin \frac{2\pi t}{T} \, dt \quad (28.26)$$

then the approximate gain of the non-linearity is given by

$$g_{df} = \frac{\sqrt{a^2 + b^2}}{U} \quad (28.27)$$

and the approximate phase is given by

$$\varphi_{df} = \arctan \frac{a}{b} \quad (28.28)$$

These approximate gain and phase depend on ω and so define a complex function of input amplitude U and frequency ω which is called describing function:

$$\begin{aligned} N(U, \omega) &= g_{df}(\omega) e^{j\varphi_{df}(\omega)} \\ &= \frac{\sqrt{a^2 + b^2}}{U} \left(\underbrace{\cos \arctan \frac{a}{b}}_{\frac{b}{\sqrt{a^2 + b^2}}} + j \underbrace{\sin \arctan \frac{a}{b}}_{\frac{a}{\sqrt{a^2 + b^2}}} \right) \\ &= \frac{b + ja}{U} \end{aligned} \quad (28.29)$$

Example 28.4. Let us find the describing function of the static two-valued non-linearity represented in Figure 28.8, known as on-off (the higher output standing for *on*, the lower for *off*):

$$y = \begin{cases} M, & \text{if } x \geq 0 \\ -M, & \text{if } x < 0 \end{cases} \quad (28.30)$$

When the input is (considering only one period)

$$x(t) = U \sin \left(\frac{2\pi t}{T} \right), \quad 0 \leq t \leq 2\pi \quad (28.31)$$

the output is

$$y(t) = \begin{cases} M, & 0 \leq t \leq \frac{T}{2} \\ -M, & \frac{T}{2} < t < T \end{cases} \quad (28.32)$$

Replacing (28.31)–(28.32) in (28.25)–(28.26),

$$\begin{aligned} a &= \frac{2}{T} \left(\int_0^{\frac{T}{2}} M \cos \frac{2\pi t}{T} dt + \int_{\frac{T}{2}}^T -M \cos \frac{2\pi t}{T} dt \right) \\ &= \frac{2M}{T} \left(\underbrace{\left[\frac{T}{2\pi} \sin \frac{2\pi t}{T} \right]_0^{\frac{T}{2}}}_{0-0} - \underbrace{\left[\frac{T}{2\pi} \sin \frac{2\pi t}{T} \right]_{\frac{T}{2}}^T}_{0-0} \right) = 0 \end{aligned} \quad (28.33)$$

$$\begin{aligned} b &= \frac{2}{T} \left(\int_0^{\frac{T}{2}} M \sin \frac{2\pi t}{T} dt + \int_{\frac{T}{2}}^T -M \sin \frac{2\pi t}{T} dt \right) \\ &= \frac{2M}{T} \left(\left[-\frac{T}{2\pi} \cos \frac{2\pi t}{T} \right]_0^{\frac{T}{2}} - \left[-\frac{T}{2\pi} \cos \frac{2\pi t}{T} \right]_{\frac{T}{2}}^T \right) \\ &= -\frac{M}{\pi} \left(\underbrace{\cos \pi}_{-1} - \underbrace{\cos 0}_1 - \left(\underbrace{\cos 2\pi}_1 - \underbrace{\cos \pi}_{-1} \right) \right) = \frac{4M}{\pi} \end{aligned} \quad (28.34)$$

Thus, we can assume for input (28.31) an approximate output given by

$$\tilde{y}(t) = \frac{4M}{\pi} \sin \left(\frac{2\pi t}{T} \right), \quad 0 \leq t \leq 2\pi \quad (28.35)$$

this being the Fourier series truncated after the first term, as seen in Figure 28.8; and the describing function is

$$N(U, \omega) = \frac{\frac{4M}{\pi} \sin \left(\frac{2\pi t}{T} \right)}{U \sin \left(\frac{2\pi t}{T} \right)} = \frac{4M}{\pi U} \quad (28.36)$$

Compare this result with what you get from (28.29).

N(U) of static non-linearities does not depend on ω

Remark 28.2. Notice that, as is clear from Figure 28.8, the input $x(t)$ and the truncated Fourier series of the output $\tilde{y}(t)$ are in phase, always; this is an obvious consequence of the non-linearity being static. Thus the phase of the describing function is always 0° , and there is in fact no dependency at all from ω . This happens with all static non-linearities, not just on-off. \square

Describing function of a reset integrator

Example 28.5. Let us find the describing function of the dynamic non-linearity called reset integrator (or **Clegg integrator**), with which the output $y(t)$ of $C(t)$, which is an integral, is reset to zero whenever the input $u(t)$ is zero. By convention, this is represented as

$$y(s) = \overset{0}{\underset{\uparrow}{\int}} \frac{1}{s} u(s) \quad (28.37)$$

Let the input be

$$u(t) = \sin \omega t, \quad t > 0 \quad (28.38)$$

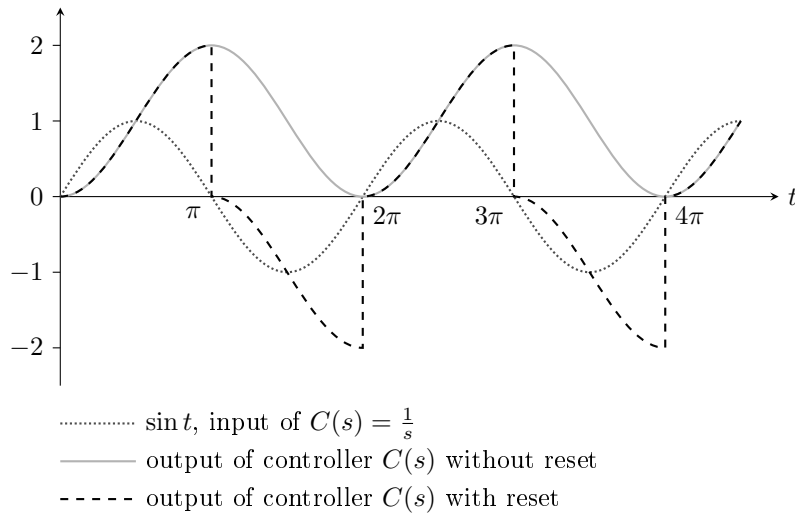
which has a frequency $\omega = \frac{2\pi}{T}$ and a period $T = \frac{2\pi}{\omega}$; the amplitude is 1, without loss of generality, since for this non-linearity the amplitude of the output is linear with respect to the amplitude of the input. Without reset, the output would be given by

$$y(t) = \frac{1}{\omega} - \frac{1}{\omega} \cos \omega t \quad (28.39)$$

(where the integration constant ensures $y(0) = 0$) and, with reset, by

$$y(t) = \begin{cases} \frac{1}{\omega} - \frac{1}{\omega} \cos \omega t, & kT \leq t < kT + \frac{T}{2}, \quad k \in \mathbb{N}_0 \\ -\frac{1}{\omega} - \frac{1}{\omega} \cos \omega t, & kT + \frac{T}{2} \leq t < (k+1)T, \quad k \in \mathbb{N}_0 \end{cases} \quad (28.40)$$

as seen in Figure 28.9.

Figure 28.9: Effects of reset control on the control action for controller $C(s) = \frac{1}{s}$

Using (28.19)–(28.20), the first coefficients of the Fourier series of (28.40) are

$$\begin{aligned}
 a &= \frac{\omega}{\pi} \int_0^{\frac{\pi}{\omega}} \left(\frac{1}{\omega} - \cos \omega t \right) \cos \omega t \, dt + \frac{\omega}{\pi} \int_{\frac{\pi}{\omega}}^{\frac{2\pi}{\omega}} \left(-\frac{1}{\omega} - \cos \omega t \right) \cos \omega t \, dt \\
 &= \frac{1}{\pi} \int_0^{\frac{\pi}{\omega}} \cos \omega t \, dt - \frac{1}{\pi} \int_{\frac{\pi}{\omega}}^{\frac{2\pi}{\omega}} \cos \omega t \, dt - \frac{1}{\pi} \int_0^{\frac{2\pi}{\omega}} \cos^2 \omega t \, dt \\
 &= \frac{1}{\pi} \underbrace{\left[\frac{1}{\omega} \sin \omega t \right]_0^{\frac{\pi}{\omega}}}_0 - \frac{1}{\pi} \underbrace{\left[\frac{1}{\omega} \sin \omega t \right]_{\frac{\pi}{\omega}}^{\frac{2\pi}{\omega}}}_0 - \frac{1}{\pi} \int_0^{\frac{2\pi}{\omega}} \frac{1 + \cos 2\omega t}{2} \, dt \\
 &= -\frac{1}{\pi} \left[\frac{t}{2} + \frac{1}{4\omega} \sin 2\omega t \right]_0^{\frac{2\pi}{\omega}} = -\frac{1}{\omega} \tag{28.41}
 \end{aligned}$$

$$\begin{aligned}
 b &= \frac{\omega}{\pi} \int_0^{\frac{\pi}{\omega}} \left(\frac{1}{\omega} - \cos \omega t \right) \sin \omega t \, dt + \frac{\omega}{\pi} \int_{\frac{\pi}{\omega}}^{\frac{2\pi}{\omega}} \left(-\frac{1}{\omega} - \cos \omega t \right) \sin \omega t \, dt \\
 &= \frac{1}{\pi} \int_0^{\frac{\pi}{\omega}} \sin \omega t \, dt - \frac{1}{\pi} \int_{\frac{\pi}{\omega}}^{\frac{2\pi}{\omega}} \sin \omega t \, dt - \frac{1}{\pi} \int_0^{\frac{2\pi}{\omega}} \sin \omega t \cos \omega t \, dt \\
 &= \frac{1}{\pi} \left[-\frac{1}{\omega} \cos \omega t \right]_0^{\frac{\pi}{\omega}} - \frac{1}{\pi} \left[-\frac{1}{\omega} \cos \omega t \right]_{\frac{\pi}{\omega}}^{\frac{2\pi}{\omega}} - \frac{1}{\pi} \int_0^{\frac{2\pi}{\omega}} \frac{1}{2} \sin 2\omega t \, dt \\
 &= -\frac{1}{\omega\pi} (-1 - 1) + \frac{1}{\omega\pi} (1 + 1) - \frac{1}{2\pi} \underbrace{\left[-\frac{1}{2\omega} \cos 2\omega t \right]_0^{\frac{2\pi}{\omega}}}_0 = \frac{4}{\omega\pi} \tag{28.42}
 \end{aligned}$$

Hence, applying (28.29) the describing function is found as

$$N(U, \omega) = \frac{4}{\omega\pi U} - \frac{1}{\omega U} j \tag{28.43}$$

and according to (28.27)–(28.28), the corresponding (approximations of) gain and phase are

$$g_{df}(U, \omega) = \sqrt{\frac{16}{\omega^2 \pi^2 U^2} + \frac{1}{\omega^2 U^2}} = \frac{\sqrt{16 + \pi^2}}{\omega \pi U} \tag{28.44}$$

$$g_{df}(1, 1) = \frac{\sqrt{16 + \pi^2}}{\pi} \approx 4 \text{ dB} \tag{28.45}$$

$$\varphi_{df}(U, \omega) = \arctan \frac{-\frac{1}{\omega U}}{\frac{4}{\omega \pi U}} = -\arctan \frac{\pi}{4} \approx -38^\circ, \forall U \tag{28.46}$$

The Bode diagram of $N(U, \omega)$ is compared in Figure 28.10 with that of $\frac{1}{s}$. It is clear that the phase margin increases, that the slope of the gain is the same, and that there is a gain offset. \square

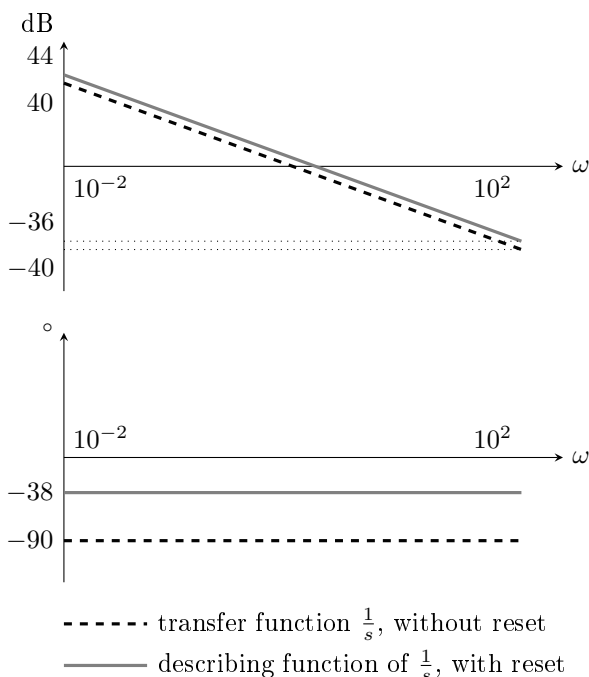


Figure 28.10: Bode diagram of controller $C(s) = \frac{1}{s}$, with and without reset control

$N(U, \omega)$ of dynamic nonlinearities depends on ω

Remark 28.3. Notice that this non-linearity has a describing function depending from ω , since it is dynamic. But, because the amplitude of the output is linearly dependent on the amplitude of the input (it is determined by an integral, which is linear if there is no reset), $N(U, \omega)$ does not depend on input amplitude U . \square

Just as Laplace transforms are in practice found from tables, so are describing functions. Table 28.1 gives describing functions of some common nonlinearities.

28.3 Predicting limit cycles

If a non-linearity causes a limit cycle in a control loop such as the one in Figure 28.2, the oscillations will take place also when the input is a step ending in 0, as seen in the bottom plot of Figure 28.7 from Example 28.3.

Consequently, since $r = 0$ and $y \neq 0$,

$$y = GN e = GN(-y) \Rightarrow y(1 + GN) = 0 \Rightarrow G = -\frac{1}{N} \quad (28.47)$$

Limit cycle if $G = -\frac{1}{N}$

We can thus expect a limit cycle whenever

$$G(j\omega) = -\frac{1}{N(U, \omega)} \quad (28.48)$$

Plot $-\frac{1}{N}$ on the Nyquist diagram

The best way to see this for a static non-linearity is to plot $-\frac{1}{N(U)}$ on the polar plot of $G(s)$ (or on the Nyquist diagram, considering only the polar plot), and verify if there are intersections:

Frequency and amplitude of the limit cycle

- the frequency ω for which there is an intersection will be the frequency of the limit cycle;
- the amplitude U for which there is an intersection is the amplitude of the limit cycle.

In the case of a dynamic non-linearity, different curves $-\frac{1}{N(U, \omega)}$ for several values of ω have to be drawn; a limit cycle will exist if there is an intersection between $G(j\omega)$ and the curve $-\frac{1}{N(U, \omega)}$ for the same value of frequency ω .

Two things must be had into account:

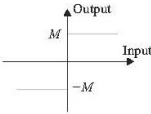
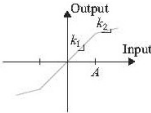
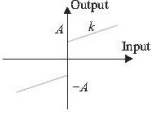
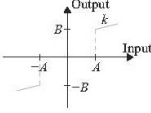
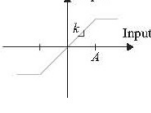
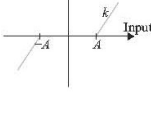
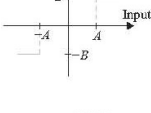
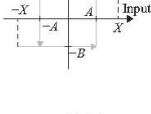
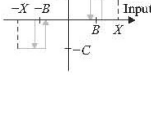
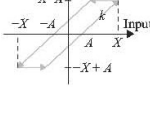
Nonlinearity	$N(X)$ with $f(\gamma) = \frac{2}{\pi} [\arcsin(\gamma) + \gamma \sqrt{1-\gamma^2}]$
	$N = \frac{4M}{\pi X}$
	$N = \begin{cases} k_2 + (k_1 - k_2)f(\frac{A}{X}), & X > A \\ k_1, & X \leq A \end{cases}$
	$N = k + \frac{4A}{\pi X}$
	$N = \begin{cases} k [1 - f(\frac{A}{X})] + \frac{4B}{\pi X} \sqrt{1 - (\frac{A}{X})^2}, & X > A \\ 0, & X \leq A \end{cases}$
	$N = \begin{cases} kf(\frac{A}{X}), & X > A \\ k, & X \leq A \end{cases}$
	$N = \begin{cases} k [1 - f(\frac{A}{X})], & X > A \\ 0, & X \leq A \end{cases}$
	$N = \begin{cases} \frac{4B}{\pi X} \sqrt{1 - (\frac{A}{X})^2}, & X > A \\ 0, & X \leq A \end{cases}$
	$N = \begin{cases} \frac{4B}{\pi X} \sqrt{1 - (\frac{A}{X})^2} - j \frac{4AB}{\pi X^2}, & X > A \\ 0, & X \leq A \end{cases}$
	$N = \begin{cases} \frac{2C}{\pi X} \left[\sqrt{1 - (\frac{B-A}{X})^2} + \sqrt{1 - (\frac{B+A}{X})^2} \right] - j \frac{4AC}{\pi X^2}, & X > A + B \\ 0, & X \leq A \end{cases}$
	$N = \frac{k}{2} \left[1 - f\left(\frac{2-X}{X}\right) \right] - j \frac{4kA(X-A)}{\pi X^2}, \quad X > A$

Table 28.1: Table of describing functions

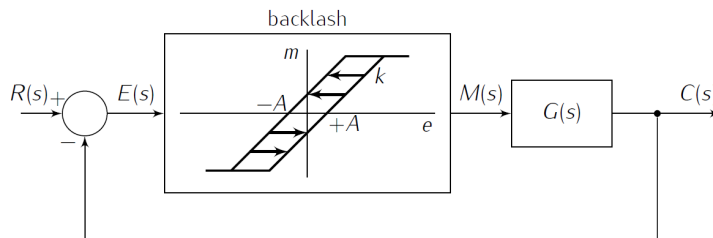


Figure 28.11: Block diagram of Example 28.6 and Exercise 5.

- The describing function is an approximation. Thus, the amplitude and frequency of the limit cycle found from (28.48) are approximations too.

Good approximation if $G(s)$ is low pass

- The approximation will be good if in the control loop of Figure 28.2 $G(s)$ is a low pass filter with a cut-off frequency such that the frequency of the limit cycle is in the pass band, and its integer multiples are in the rejection band. (Remember that the describing function is a truncated Fourier series.)

Example 28.6. The plant in Figure 28.11 includes a linear part

$$G(s) = \frac{1.4}{s(s+1)^2} \quad (28.49)$$

and a backlash non-linearity with $A = 1.1$ and $k = 1$. To verify if there is a limit cycle we plot curve $-\frac{1}{N}$ over the Nyquist diagram of $G(s)$ and look for intersections with the polar diagram. Figure 28.12 shows there are two. To find their values of frequency ω and amplitude X , we must solve

$$\frac{1.4}{j\omega(j\omega+1)^2} = -\frac{1}{\frac{1}{2} \left(1 - \frac{2}{\pi} \left[\arcsin \left(\frac{2-\frac{X}{1.1}}{\frac{X}{1.1}} \right) + \frac{2-\frac{X}{1.1}}{\frac{X}{1.1}} \cos \left(\arcsin \left(\frac{2-\frac{X}{1.1}}{\frac{X}{1.1}} \right) \right) \right] \right)} - 4j \frac{1.1(X-1.1)}{\pi X^2} \quad (28.50)$$

This is better found numerically with MATLAB; we find that

$$\begin{cases} X = 1.37 \leftrightarrow \omega = 0.31 \text{ rad/s} \\ X = 4.72 \leftrightarrow \omega = 0.76 \text{ rad/s} \end{cases} \quad \square \quad (28.51)$$

Finding limit cycles for static non-linearities

Remark 28.4. Sometimes solutions for (28.48) can be found analytically. This is easier if the non-linearity is static. In that case, $-\frac{1}{N}$ is real and negative, and ω is a phase crossover frequency.

(28.50) is one of those cases where a numerical solution has to be sought. \square

Some of the limit cycles found from (28.48) are never found in practice because they are **unstable**. The ones actually found are those which are **stable**. The Nyquist diagram is used to tell them apart:

Stable limit cycle

- Suppose that, when the amplitude of the limit cycle increases, curve $-\frac{1}{N}$ enters a zone of the Nyquist plane where the number of encirclements is such that the Nyquist stability criterion proves that the closed loop is unstable. In that case, the amplitude increases even more. Likewise, when the amplitude of the limit cycle decreases, curve $-\frac{1}{N}$ will enter a zone of the Nyquist plane where the number of encirclements is such that the Nyquist stability criterion proves that the closed loop is stable. Thus, the amplitude will decrease even more. Whichever case happens, the limit cycle will disappear.

Unstable limit cycle

- Suppose that, when the amplitude of the limit cycle increases, curve $-\frac{1}{N}$ enters a zone of the Nyquist plane where the number of encirclements is such that the Nyquist stability criterion proves that the closed loop is stable. In that case, the amplitude will stop increasing and will rather decrease. Likewise, when the amplitude of the limit cycle decreases, curve $-\frac{1}{N}$ will enter a zone of the Nyquist plane where the number of encirclements is such that the Nyquist stability criterion proves that the closed

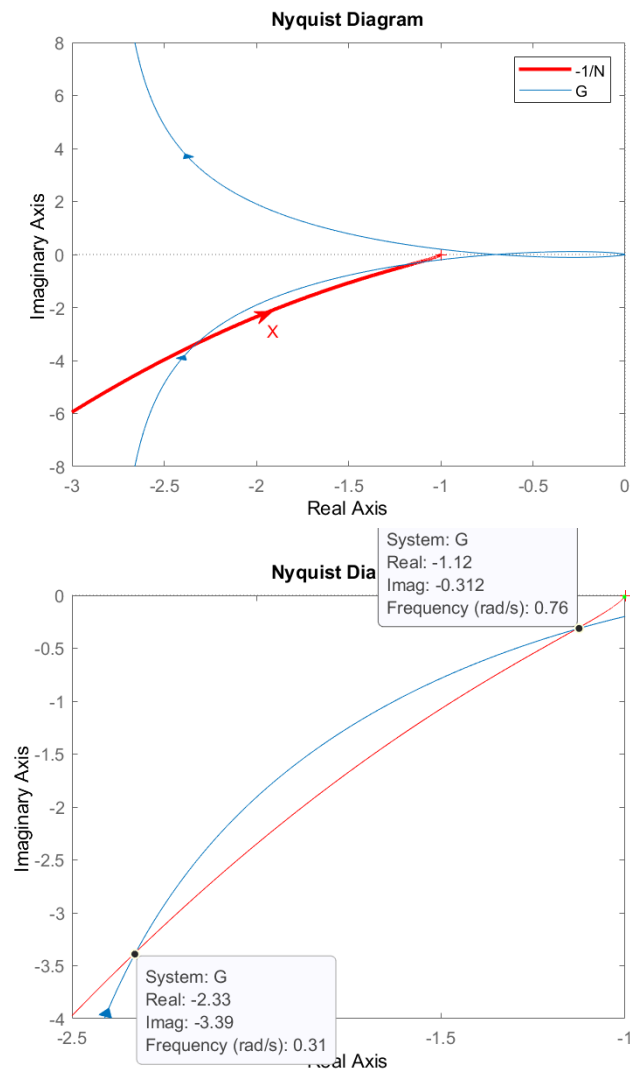


Figure 28.12: Nyquist diagram of (28.49) and $-\frac{1}{N}$ for the backlash in Figure 28.11.

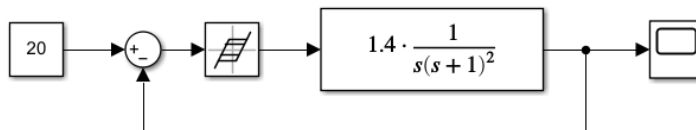


Figure 28.13: SIMULINK implementation of Figure 28.11 for Examples 28.6 and 28.7.

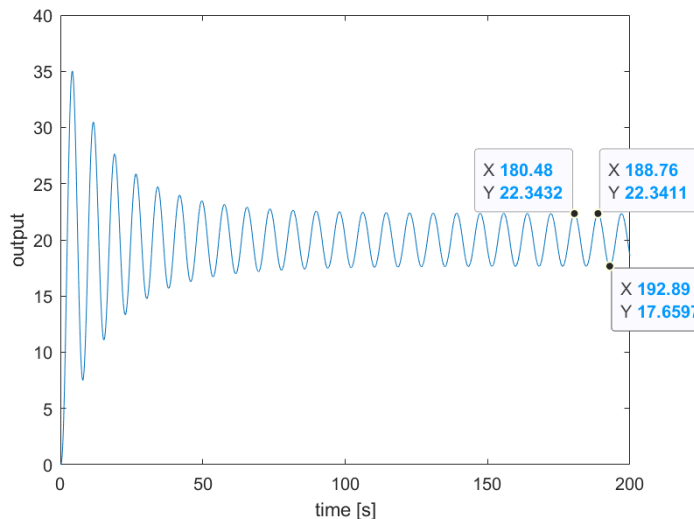


Figure 28.14: Output of the control system in Figure 28.11 for Example 28.7.

loop is unstable. Thus, the amplitude will stop decreasing and will rather increase. Whichever case happens, the amplitude of the limit cycle will remain roughly the same.

This is better seen through an example.

Example 28.7. Simulating the closed loop of Example 28.6 as shown in Figure 28.13, we obtain the result in Figure 28.14, with a limit cycle of frequency $\frac{2\pi}{188.76-180.48} = 0.76$ rad/s and an amplitude quite close to the expected one, $22.34 - 17.66 = 4.68$.

This is because, of the two limit cycles, this one is stable. Notice that $G(s)$ has no unstable poles. Consequently, by the Nyquist stability criterion, gains corresponding to no encirclements of the Nyquist plot lead to stable closed loop, and one or more encirclements correspond to an unstable closed loop. The Nyquist diagram in Figure 28.12 closes on the right, as is clear from

$$\lim_{s \rightarrow 0^+} G(s) = \lim_{s \rightarrow 0^+} \frac{1.4}{s(s+1)^2} = +\infty \quad (28.52)$$

So we see that, at the intersection in the upper right corner of the lower plot, corresponding to $X = 4.72$ and $\omega = 0.76$ rad/s:

- If the amplitude X increases, the diagram leaves the encirclement. This is a stable zone of the plane. Thus the amplitude will decrease back to $X = 4.72$.
- If the amplitude X decreases, the diagram enters the encirclement. This is an unstable zone of the plane. Thus the amplitude will increase back to $X = 4.72$.

The same Figure shows that, at the intersection in the lower left corner of the lower plot, corresponding to $X = 1.37$ and $\omega = 0.31$ rad/s:

- If the amplitude X increases, the diagram enters the encirclement. This is an unstable zone of the plane. Thus the amplitude will increase away from $X = 1.37$, approaching the limit cycle with $X = 4.72$.

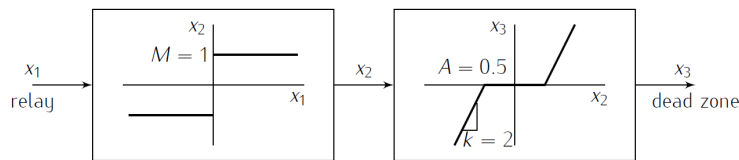


Figure 28.15: Block diagram of Exercise 1.

- If the amplitude X decreases, the diagram leaves the encirclement. This is a stable zone of the plane. Thus the amplitude will keep decreasing away from $X = 1.37$.

This limit cycle is thus unstable, and will not be observed in practice. \square

Glossary

He was not loved, Guy knew, either by his household or in the town. He was accepted and respected but he was not *simpatico*. Gräfin von Gluck, who spoke no word of Italian and lived in undisguised concubinage with her butler, was *simpatica*. Mrs. Garry was *simpatica*, who distributed Protestant tracts, interfered with the fishermen's methods of killing octopuses and filled her house with stray cats.

Guy's uncle, Peregrine, a bore of international repute whose dreaded presence could empty the room in any centre of civilization, — Uncle Peregrine was considered *molto simpatico*. The Wilmots were gross vulgarians; they used Santa Dulcina purely as a pleasure resort, subscribed to no local funds, gave rowdy parties and wore indecent clothes, talked of 'wops' and often left after the summer with their bills to the tradesmen unpaid; but they had four boisterous and ill-favoured daughters whom the Santa-Dulcinesi had watched grow up. Better than this, they had lost a son bathing here from the rocks. The Santa-Dulcinesi participated in these joys and sorrows. They observed with relish their hasty and unobtrusive departures at the end of the holidays. They were *simpatici*. Even Musgrave who had the Castelletto before the Wilmots and bequeathed it his name, Musgrave who, it was said, could not go to England or America because of warrants for his arrest, 'Musgrave the Monster', as the Crouchbacks used to call him — he was *simpatico*. Guy alone, whom they had known from infancy, who spoke their language and conformed to their religion, who was open-handed in all his dealings and scrupulously respectful of all their ways, whose grandfather built their school, whose mother had given a set of vestments embroidered by the Royal School of Needlework for the annual procession of St Dulcina's bones — Guy alone was a stranger among them.

Evelyn WAUGH (1903 — †1966), *Sword of Honour* (Men at arms, 1952), 1 I

describing function função de descrição

equivalent gain ganho equivalente

limit cycle ciclo limite

Exercises

1. Figure 28.15 shows two non-linear elements (a relay and a dead zone) in series. To which of the following single systems is this equivalent?
 - (a) A dead zone with $A = 0.5$.
 - (b) A relay with $M = 0.5$.
 - (c) A relay with $M = 1$.
 - (d) None of the above.

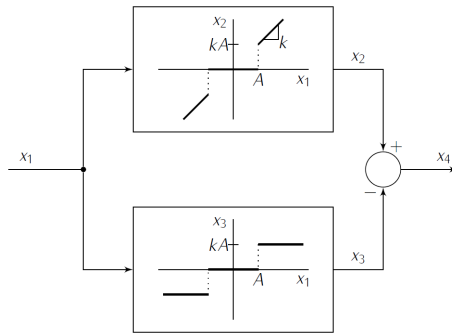


Figure 28.16: Block diagram of Exercise 2.

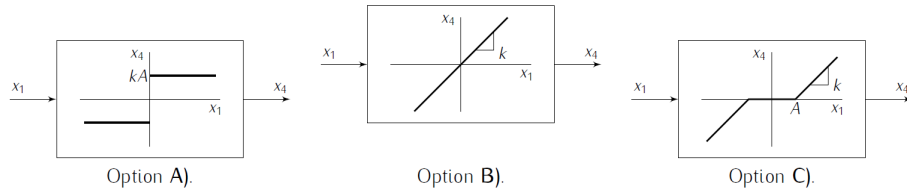


Figure 28.17: Options of Exercise 2.

2. Figure 28.16 shows two non-linear elements (a relay and a dead zone) in parallel. To which of the single systems in Figure 28.17 is this equivalent?

3. Repeat the calculations of Example 28.3, when the plant of the diagram in Figure 28.3 is

$$G(s) = \frac{s^2 + 7s + 10}{s^3 + 0.25s^2 + s} \tag{28.53}$$

4. The plant in Figure 28.18 includes a backlash non-linearity with $2A = 0.5$ and $B = 1, C = 1$. Let $G(s) = \frac{3}{s(s+1)^2}$. Use the describing function method to verify if this plant has a limit cycle. If it does, find its frequency and amplitude.

5. The plant in Figure 28.11 includes a backlash non-linearity with $A = 1$ and $k = 1$. Let $G(s) = \frac{1.5}{s(s+1)^2}$. Use the describing function method to verify if this plant has a limit cycle. If it does, find its frequency and amplitude.

6. Consider the plant in Figure 28.19.

- (a) Let $R = 0, A = 4, k = 1$ and $K_1 = 20$. Find the frequency and the amplitude of the limit cycle.
- (b) Is this limit cycle stable or unstable?
- (c) Increasing gain K_1 , how does the limit cycle change?

7. The plant in Figure 28.20 includes a saturation non-linearity with $M = 1$. Use the describing function method to verify if this plant has a limit cycle. If it does, find its frequency and amplitude.

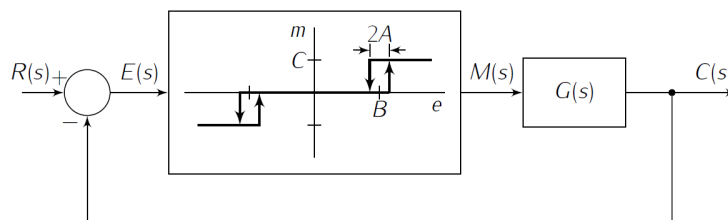


Figure 28.18: Block diagram of Exercise 4.

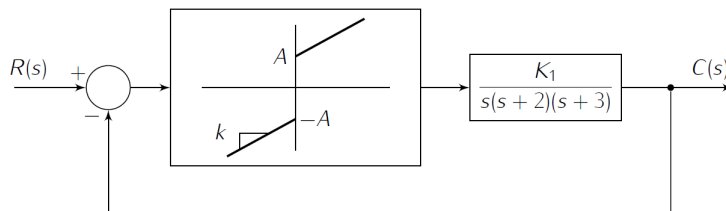


Figure 28.19: Block diagram of Exercise 6.

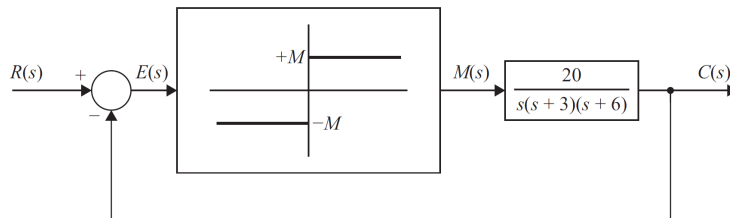


Figure 28.20: Block diagram of Exercise 7.

8. The plant in Figure 28.21 includes a dead zone non-linearity with $A = 1$ and $K = 1$. Use the describing function method to verify if this plant has a limit cycle. If it does, find its frequency and amplitude.
9. Repeat Exercise 7 when the transfer function is $G(s) = \frac{e^{-2s}}{s(s+2)}$ instead.
10. Repeat Exercise 7 when the transfer function is $G(s) = \frac{100}{s(s+2)(s+3)}$ instead.

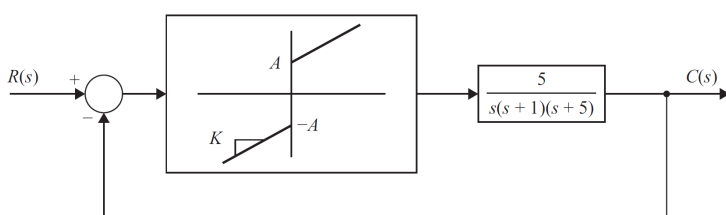


Figure 28.21: Block diagram of Exercise 8.

Chapter 29

Other aspects of controller design and implementation

Metido tenho a mão na consciencia,
E não fallo senão verdades puras,
Que m'ensinou a viua experiencia.

Luís Vaz de CAMÕES (1524? — †1580), *Rimas de Luis de Camões*, Soneto 87
(1598 edition)



Strictly proper systems. Placing extra poles. PID as PI+D. Windup. Reset. Modelling errors. Robustness. Gain modelling. Waterbed effect. The small gain theorem.

This chapter is still being written.

In the picture: National Pantheon, or Church of Saint Engratia, Lisbon (source: <http://www.panteaonacional.gov.pt/171-2/historia-2/>), some-when during its construction (1682–1966).

Glossary

“Listen to me for a moment,” said the steady voice of Moreau; “and then say what you will.”

“Well?” said I.

He coughed, thought, then shouted: “Latin, Prendick! bad Latin, schoolboy Latin; but try and understand. *Hi non sunt homines; sunt animalia qui nos habemus*—vivisected. A humanising process. I will explain. Come ashore.”

H. G. WELLS (1866 — †1946), *The Island of Doctor Moreau* (1896), XIII

word in English word in Portuguese
palavra em inglês palavra em português

Exercises

1. Question.

Part VI

System identification

‘Yes. Dr Hartman has a theory. In any investigation, my Bunter, it is most damnably dangerous to have a theory.’

‘I have heard you say so, my lord.’

‘Confound you — you know it as well as I do! What is wrong with the doctor’s theories, Bunter?’

‘You wish me to reply, my lord, that he only sees the facts which fit into the theory.’

‘Thought-reader!’ exclaimed Lord Peter bitterly.

‘And that he supplies them to the police, my lord.’

Dorothy L. SAYERS (1893 — †1957), *Lord Peter Views the Body* (1928), The vindictive story of the footsteps that ran

In this part of the lecture notes:

Chapter 30 addresses general questions related to system identification.

Chapter 31 is about the identification of transfer functions from time responses.

Chapter 32 concerns the identification of transfer functions from frequency responses.

Chapter 33 shows how to identify non-linearities.

Here is what you need to know beforehand to follow these chapters:

- The Laplace and Fourier transforms, from Chapter 2;
- Transfer functions, from Sections 4.1 and 4.2 of Chapter 4;
- System theory, from Part II;
- Filters, from Sections 12.2 and 12.3 of Chapter 12;
- Discrete transfer functions, from Chapter 25, Sections 26.1 to 26.5 of Chapter 26, and Sections 27.1, 27.2 and 27.6 of Chapter 27.

For Chapter 33, you also need to know the following:

- Soft non-linearities, from Section 8.3 of Chapter 8;
- Hard non-linearities, from Chapter 28;
- Delays, from Chapter 24.

Chapter 30

Overview and general issues

Je n'ay jamais jugé d'une mesme chose exactement de mesme, je ne puis juger d'un ouvrage en le faisant ; il faut que je fasse comme les peintres & que je m'en éloigne, mais non pas trop. De combien donc ? Devinez.

Blaise PASCAL (1623 — †1662), *Pensées diverses* (1670, posth.), I, 12/37

We already saw in Section 3.3 that system identification is the obtention of models from experimental data. More generally, it is the obtention of models for a plant from its behaviour. In some cases, a controller can be obtained identifying its model from its desired behaviour, determined analytically from the plant to control and the specifications to be followed. But the usual case is, indeed, the identification of a model for a plant from its behaviour experimentally determined, and that will be the situation presumed by default in what follows. The models we address are transfer functions; we will not consider the identification of NN and fuzzy models, mentioned in Figures 3.18 and 3.19, which are beyond our scope. This chapter sums up some general questions before entering into particular methods and algorithms.

Identification of transfer functions

30.1 Types of data, types of identification methods, and types of models

The experimental data always consists of responses in time. Chapter 31 addresses identification methods directly from time responses.

It is possible to obtain from time responses a frequency response: this is obvious if the responses in time that we have are responses to sinusoids, but from other types of time responses a frequency response can be obtained as well. In Chapter 32 we will see identification methods from frequency responses.

As we saw in Section 3.2 and in Chapter 13, nowadays experimental data is nearly always a digital signal, and thus sampled in time. We will address some graphical methods that seem to implicitly require an analogical signal. Remember that:

- if the sampling time was well chosen, according to the criteria in Section 25.4, the digital signal will reproduce the analogical signal with accuracy;
- fitting a curve (this includes a straight line, of course) or finding a slope can be done numerically with a set of points.

Even though signals are almost always digital, some identification methods provide a continuous transfer function, while others provide a discrete transfer function. If the desired transfer function is of the other type, a conversion using the Tustin approximation or a pole-zero match is then performed.

Continuous and discrete identified models

The identification methods addressed in this part presume SISO models. If the plant is linear, this poses no restriction at all:

How to deal with MIMO plants

- if there are several outputs, each of them is identified separately, i.e. the MIMO plant is treated as several MISO plants;

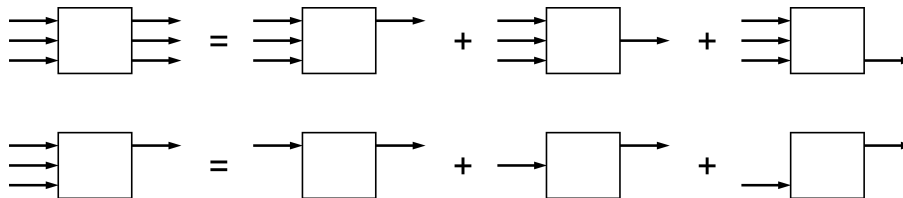


Figure 30.1: How to identify a MIMO plant as several MISO plants (top), and a linear MISO plant as several SISO plants (bottom).

- if there are several inputs, since the plant is linear, the output is a linear combination of SISO plants. Consequently:
 - all inputs are set to zero, save one of them, and a model for the relation between that input and the output is found;
 - the step above is repeated for each input;
 - the model of the MISO plant is the sum of the SISO models found.

All this is illustrated in Figure 30.1. But if there are non-linearities the above method fails. We will address the identification of non-linearities only for SISO models, in Chapter 33.

Identification methods can be performed using data as it becomes available.

Identification in real time

This is called **real time identification**, and if it works as expected the model obtained gets more accurate with time, as more data becomes available. (Obviously, if the model becomes worse, the identification is a failure.) Identification can also be carried out only after all data becomes available, using a batch of measured values all at once. This is called **offline identification**.

Offline identification

If there is a model based upon first principles

The case where there already is a model of the plant based upon first principles has been mentioned in Section 3.3. Remember that:

- in this case, identification from experimental data can be used to confirm, or correct, the values of the parameters of a model for which the structure is known;
- experimental data can also be used to confirm (or not) the said structure of the model. It may show that some neglected phenomenon (say, a non-linearity, or a friction force) is after all so significant that, for the model to have the desired accuracy, additional details must be included (a non-linearity, or additional poles or zeros, etc.).

If there is no model available before identification, then a model structure must be found before model parameters are identified; or, in other identification methods, both the model structure and its parameters are found simultaneously.

Identification in closed loop

Some plants have to be identified when they are already controlled in closed loop. This may happen for several reasons:

- it is somehow impossible to turn off or disconnect the controller;
- without control, outputs may reach unacceptable values (e.g. some temperature may become so high that the plant will be damaged; some pressure may rise to values that endanger operators' lives; forces will be applied in a structure that will yield);
- the plant is marginally stable or unstable, and without control outputs will reach saturation values so fast that almost no data at all can be recorded. Such plants usually have the problem of unacceptable values too.

In this case, a known controller C must be used, and what is identified is the closed loop transfer function F . If the closed loop is the one from Figure 9.13, plant G is found as

$$\begin{aligned}
 F &= \frac{CG}{1+CG} \Rightarrow F + FCG = CG \\
 &\Rightarrow F = CG(1-F) \\
 &\Rightarrow G = \frac{F}{C(1-F)}
 \end{aligned} \tag{30.1}$$

Different configurations of the closed loop are dealt with similarly.

30.2 Comparing model performance

Model accuracy has already been mentioned a few times. It is the difference between the experimental outputs and those of the model for the same inputs. Of course, a model can be accurate for some cases and inaccurate for others (hopefully less relevant). Some models are good during the transient regime but not in steady-state; for others, it is the other way round. Some models reproduce step responses but not those of other inputs, or the plant's frequency response; with other models the reverse can be true. It is surely desirable, but seldom possible, to have a model which is *always* accurate. Or perhaps increasing the accuracy of the model requires a model structure which is too complicated (with too many poles and zeros). Quite often, it is better to have a manageable model with a moderate number of zeros and poles than a very accurate model which is very complicated and difficult to study and simulate. What is to be considered *too complicated* depends on the case.

Trade-off between accuracy and simplicity

The following performance indexes are often used to measure accuracy. For each of them, two expressions are given: one for analogical signals, another for digital signals. The experimental variable is y , and the variable that approximates y is \hat{y} . In the digital case, there are $N + 1$ samples, numbered from 0 to N . The analogical case is given for time varying from 0 to t_f ; should the signal depend on frequency (say, from ω_l to ω_h), the difference in the calculations is obvious.

- **Mean absolute error (MAE):**

MAE

$$\text{MAE}(y(t) - \hat{y}(t)) = \frac{1}{t_f} \int_0^{t_f} |y(t) - \hat{y}(t)| dt \quad (30.2)$$

$$\text{MAE}(y_k - \hat{y}_k) = \frac{1}{N + 1} \sum_{k=0}^N |y_k - \hat{y}_k| \quad (30.3)$$

- **Mean square error (MSE):**

MSE

$$\text{MSE}(y(t) - \hat{y}(t)) = \frac{1}{t_f} \int_0^{t_f} (y(t) - \hat{y}(t))^2 dt \quad (30.4)$$

$$\text{MSE}(y_k - \hat{y}_k) = \frac{1}{N + 1} \sum_{k=0}^N (y_k - \hat{y}_k)^2 \quad (30.5)$$

- **Root mean square error (RMS):**

RMS

$$\text{RMS}(y(t) - \hat{y}(t)) = \sqrt{\frac{1}{t_f} \int_0^{t_f} (y(t) - \hat{y}(t))^2 dt} \quad (30.6)$$

$$\text{RMS}(y_k - \hat{y}_k) = \sqrt{\frac{1}{N + 1} \sum_{k=0}^N (y_k - \hat{y}_k)^2} \quad (30.7)$$

- **Coefficient of determination (R^2):**

R^2

$$R^2(y(t) - \hat{y}(t)) = 1 - \frac{\int_0^{t_f} (y(t) - \hat{y}(t))^2 dt}{\int_0^{t_f} \left(y(t) - \underbrace{\frac{1}{t_f} \int_0^{t_f} y(t) dt}_{\text{average value of } y} \right)^2 dt} \quad (30.8)$$

$$R^2(y_k - \hat{y}_k) = 1 - \frac{\sum_{k=0}^N (y_k - \hat{y}_k)^2}{\sum_{k=0}^N \left(y_k - \underbrace{\frac{1}{N + 1} \sum_{k=0}^N y_k}_{\text{average value of } y} \right)^2} \quad (30.9)$$

- **Variance accounted for (VAF):**

VAF

$$\text{VAF}(y - \hat{y}) = 1 - \frac{\sigma^2(y - \hat{y})}{\sigma^2(y)} \quad (30.10)$$

Variance

Here $\sigma^2(x)$ is the **variance** of x , be it continuous or discrete. Notice that $\sigma^2(y - \hat{y}) = \text{MSE}(y - \hat{y})$.

The best model according to one criterion may not be the best according to another, and the best performance index depends on the case. Very often it is a good idea to take a look at several of them. Notice that:

- MAE and RMS have the same units as y ;
- MSE has units of y^2 ;
- R^2 and VAF have no dimensions;
- MAE, MSE and RMS are 0 if there is no error;
- R^2 and VAF are 1 if there is no error;
- VAF is often given as a percentage.

There is no error if
 $MAE=MSE=RMS=0$ and
 $R^2=VAF=1=100\%$

The performance indexes above do not help to solve the trade-off between accuracy and simplicity mentioned above. For that purpose, the following ones are employed. Data is assumed to be discrete in time; models have K parameters.

AIC

- **Akaike Information Criterion (AIC):**

$$\begin{aligned} \text{AIC}(y_k - \hat{y}_k) &= N \log \text{MSE}(y_k - \hat{y}_k) + 2K + \overbrace{\frac{2K(K+1)}{N-K-1}}^{\text{correction for small data sets}} \\ &= N \log \text{MSE}(y_k - \hat{y}_k) + 2K \underbrace{\frac{N}{N-K-1}}_{\text{correction for small data sets}} \end{aligned} \quad (30.11)$$

The correction for small data sets can be neglected if $N \gg K$; say, if $N > 40K$. Indeed, the value of $\frac{N}{N-K-1}$ is

- only 1.053 if $K = 1$ and $N = 40K$ (and so neglecting it correspond to an error of about 5%),
- even lower, if $K > 1$,
- also lower, if $N > 40$.

BIC

- **Bayesian Information Criterion (BIC):**

$$\text{BIC}(y_k - \hat{y}_k) = N \log \text{MSE}(y_k - \hat{y}_k) + K \log N \quad (30.12)$$

This expression has no correction for small data sets, and so applies only if $N \gg K$. Compared with the AIC, the BIC favours models with less parameters.

- **Normalised AIC and BIC**, needed to compare models found from data with different lengths N :

$$\text{nAIC}(y_k - \hat{y}_k) = \frac{\text{AIC}(y_k - \hat{y}_k)}{N} \quad (30.13)$$

$$\text{nBIC}(y_k - \hat{y}_k) = \frac{\text{BIC}(y_k - \hat{y}_k)}{N} \quad (30.14)$$

The lower the value of the AIC or of the BIC, the better the model is considered; the values themselves have no significance, and only serve to compare different models. Given n models with outputs $i\hat{y}_k$, $i = 1, \dots, n$, it is possible to find from their values of the AIC or of the BIC a probability of each model being the best. (Defining precisely in which sense this probability is to be understood falls outside our scope.) This probability is given by the **Akaike weights** for

Akaike weights

Table 30.1: Values of the AIC, BIC, and corresponding weights for four models of plant (30.17).

	$G_1(s)$	$G_2(s)$	$G_3(s)$	$G_4(s)$
AIC	-0.98	-0.98	-4635.0	-4630.1
$w(\text{AIC})$	0.0%	0.0%	92.2%	7.8%
BIC	15.8	15.8	-4618.2	-4613.3
$w(\text{BIC})$	0.0%	0.0%	92.2%	7.8%

the AIC, or the **Schwarz weights** for the BIC:

$$w_i(\text{AIC}) = \frac{\exp\left(-\frac{\text{AIC}_i - \min_{m=1\dots n} \text{AIC}_m}{2}\right)}{\sum_{p=1}^n \exp\left(-\frac{\text{AIC}_p - \min_{m=1\dots n} \text{AIC}_m}{2}\right)}, \quad i = 1 \dots n \quad (30.15)$$

$$w_i(\text{BIC}) = \frac{\exp\left(-\frac{\text{BIC}_i - \min_{m=1\dots n} \text{BIC}_m}{2}\right)}{\sum_{p=1}^n \exp\left(-\frac{\text{BIC}_p - \min_{m=1\dots n} \text{BIC}_m}{2}\right)}, \quad i = 1 \dots n \quad (30.16)$$

Example 30.1. Consider the unit step response of plant

$$G(s) = \frac{10}{s^2 + 0.6s + 1} \quad (30.17)$$

sampled with $T_s = 0.01$ s from 0 s to 20 s, comprising thus $N = 2001$ points. Suppose that four models were somehow obtained from this response:

$$G_1(s) = \frac{11}{s^2 + 0.6s + 1} \quad (30.18)$$

$$G_2(s) = \frac{9}{s^2 + 0.6s + 1} \quad (30.19)$$

$$G_3(s) = \frac{10}{s^2 + 0.5s + 1} \quad (30.20)$$

$$G_4(s) = \frac{10}{s^2 + 0.726s + 1} \quad (30.21)$$

All these models have the same number of parameters: 2 poles and the gain, with no zeros, i.e. $K_1 = K_2 = K_3 = K_4 = 3$. Figure 30.2 compares the desired unit step response with those of the models, and gives the corresponding values of the MSE. The corresponding values of the AIC, the BIC, the Akaike weights, and the Schwarz weights are given in Table 30.1. Notice that:

- Since K is the same for all models, it is unsurprising that ranking the models using the MSE and the AIC gives the very same result.
- For the same reason, even though the values of the AIC and the BIC are different, the ranking and even the weights are the same.
- Since $G_1(s)$ and $G_2(s)$ have the same MSE and the same K , the AIC and the BIC are also equal for these two models.

Suppose that two additional models are also obtained:

$$G_5(s) = \frac{10(s + 11.2705)}{(s^2 + 0.6s + 1)(s + 11)} \quad (30.22)$$

$$G_6(s) = \frac{10(s + 11.01 \times 100 \times 50)}{(s^2 + 0.6s + 1)(s + 11)(s + 100)(s + 50)} \quad (30.23)$$

$G_5(s)$ has 3 poles, 1 zero, and the gain, i.e. $K_5 = 5$; while $G_6(s)$ has 5 poles, 1 zero, and the gain, i.e. $K_6 = 7$. Figure 30.3 compares the desired unit step response with those of the new models, and gives the corresponding values of the MSE. The MSE is lower for $G_6(s)$, but this model is clearly the more complex. Table 30.1 shows that:

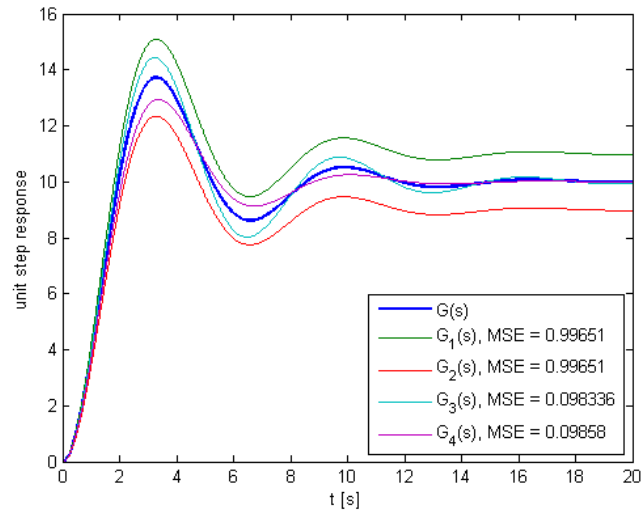


Figure 30.2: Unit step responses of plant (30.17) and of four models thereof.

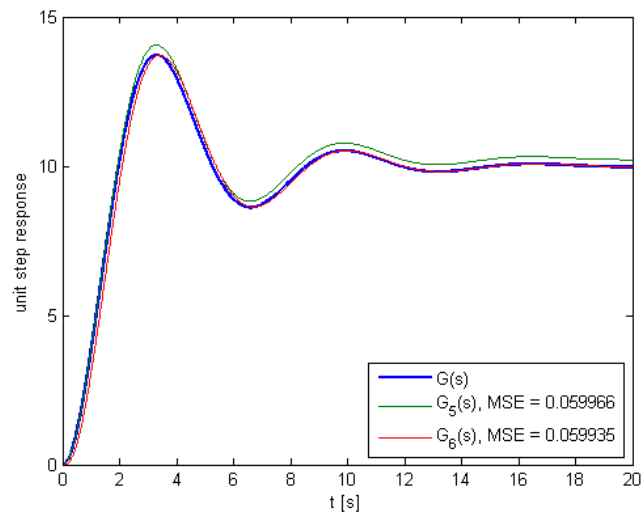


Figure 30.3: Unit step responses of plant (30.17) and of two models thereof.

- the lowest AIC is that of $G_5(s)$, which results in a very high probability of being the best;
- $G_5(s)$ is neither the model with less parameters nor the model with lower MSE;
- that the BIC favours $G_5(s)$ even more; in fact, a slightly higher MSE of $G_5(s)$ could make Akaike weights attribute a higher probability for $G_6(s)$, while the Schwarz weights would still favour $G_5(s)$;
- $G_3(s)$ and $G_4(s)$, which had significant weights when there were only four models, are now considered highly improbable.

□

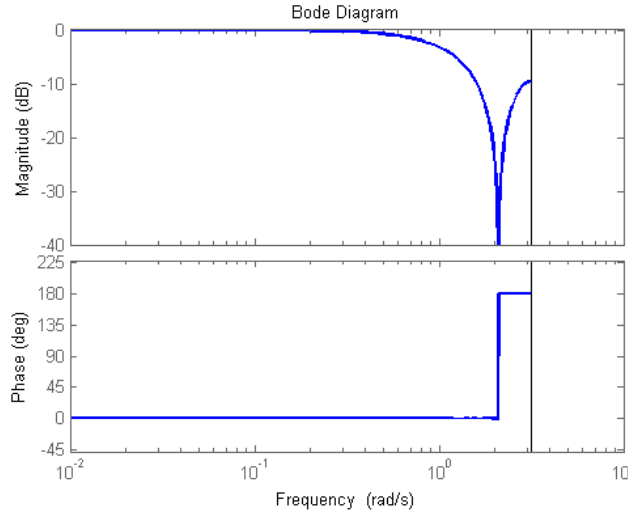
30.3 Noise

All experimental data is more or less corrupted by noise. Explicit consideration of noise in models and in the identification process leads to models of a type called stochastic models, addressed below in Part VIII. Here we address the way to try to eliminate it from data.

Suppose that a signal y is corrupted with noise y_n , so that the measured

Table 30.2: Values of the AIC, BIC, and corresponding weights for six models of plant (30.17).

	$G_1(s)$	$G_2(s)$	$G_3(s)$	$G_4(s)$	$G_5(s)$	$G_6(s)$
AIC	already given in Table 30.1				-5620.7	-5617.7
$w(\text{AIC})$	0.0%	0.0%	0.0%	0.0%	81.6%	18.4%
BIC	already given in Table 30.1				-5592.7	-5578.6
$w(\text{BIC})$	0.0%	0.0%	0.0%	0.0%	92.9%	0.1%

Figure 30.4: Bode diagram of (30.27), when $n = 1$ and $T_s = 1$ s.

value \tilde{y} is

$$\tilde{y} = y + y_n \quad (30.24)$$

Then the **signal to noise ratio (SNR)** is defined as

$$\text{SNR} = \frac{\text{RMS}(y)}{\text{RMS}(y_n)} \quad (30.25)$$

The SNR is often given in dB, i.e. as

$$20 \log_{10} \text{SNR} = 20 \log_{10} \frac{\text{RMS}(y)}{\text{RMS}(y_n)} \quad (30.26)$$

When $\text{SNR} < 1$ (i.e. below 0 dB) noise is larger than the signal itself. It is obviously desirable that $\text{SNR} \gg 1$ (i.e. $20 \log_{10} \text{SNR} \gg 0$ dB).

To eliminate noise, a filter is applied. Offline identification can use non-causal filters.

Example 30.2. A **centred moving average filter** of order $2n + 1$ is a non-causal digital low-pass filter given by *Centred moving average*

$$F(z) = \frac{z^{-n} + \dots + z^{-1} + 1 + z + \dots + z^n}{2n + 1} \quad (30.27)$$

It has a zero phase until the rejection band, thus introducing no phase distortion. Its gain is almost constant in a large passband. See Figure 30.4. \square

When a too small sampling time was employed, data should be **subsampling**, *Subsampling* i.e. only the measurements at regular intervals are kept. This corresponds to a low pass filtering of a signal. When there is so much experimental data that the computational effort of using it all would be excessive, there are two options:

- subsample the data, which eliminates high frequency behaviour (since subsampling has the effect of a low pass filter); *Truncating and subsampling eliminate low and high frequencies respectively*
- truncate the data, which eliminates low frequency behaviour (which is not captured in a shorter time range).

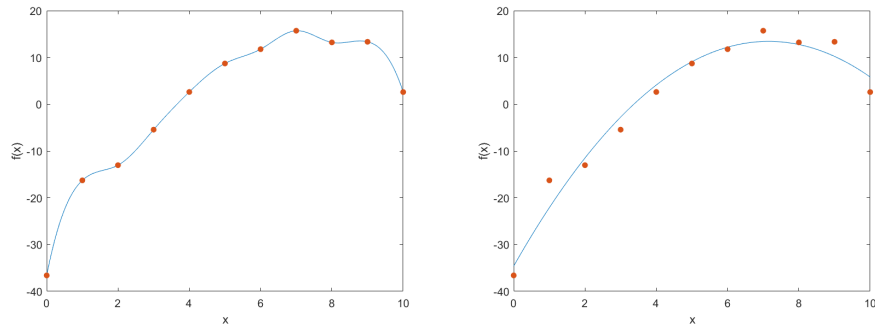


Figure 30.5: Interpolating (left) and fitting by least squares a second order polynomial (right) to 11 points.

The discussion above presumes that it is known what in the measured values is the signal and what is the noise. Sometimes this is not clear at the outset or by a mere inspection of the data. It is necessary to have some previous estimate of the bandwidth of the plant to know what is to be filtered. Remember that:

- noise with frequencies inside that bandwidth can only be eliminated damaging the response of the plant itself;
- filters attenuate but seldom, if ever, completely eliminate;
- any behaviour of the plant (or of the noise) that is filtered cannot be expected to be modelled accurately.

Noise is not everything that has to be eliminated from data: initial conditions must be set to zero; remember Example 8.1. Some identification methods require normalising all variables involved, but since the methods we are dealing with are for transfer function models, which are linear, this should not make any difference.

The results of identification methods must be critically analysed rather than accepted without reserve. It often happens that identification methods provide models with unnecessary poles and zeros, that is to say:

- poles and zeros that cancel out, or nearly cancel out and only do not because of slight numerical differences (these can be removed in MATLAB using function `minreal`);
- poles and zeros with frequencies outside the range of frequencies where experimental data can lead to acceptable results.

MATLAB's `command`
`minreal`
Range of frequencies where
good results can lie

Remember from what we saw in Chapter 25 that a signal measured from time 0 to time t_f with sampling frequency T_s cannot carry information outside frequency range

$$\frac{\pi}{t_f} \leq \omega \leq \frac{\pi}{T_s} \quad (30.28)$$

which corresponds to a bare minimum of two points per period; and that in practice, since we need at least some 10 points to sample a period of a sinusoid with an error in its amplitude below 5%, it is better to consider the narrower frequency range

$$10 \frac{2\pi}{t_f} \leq \omega \leq \frac{2\pi}{10T_s} \quad (30.29)$$

instead. Eliminating the poles and zeros of a model that fall outside this range of frequencies is also a sort of filtering, since less frequencies are being considered in the end.

30.4 Interpolation vs. curve fitting

There are two major ways of adjusting a model's parameters to data: **interpolation** and **curve fitting** (see Figure 30.5).

When data is interpolated:

- The model passes through all the data points. This makes more sense if there is little (or, better still, no) noise, since the model will account for all the noise in the data.
- The order of the model depends on how many points are interpolated. Passing through more data points requires more parameters, unless there is no noise.
- These two characteristics mean that increasing the number of data points usually leads to a model with many parameters that mostly tries to reproduce noise (in short, a poorer model).

When using curve fitting:

- The model seldom passes through all data points. This makes more sense if there is noise; parameter adjustment automatically involves some sort of noise filtering.
- The order of the model is fixed in advance. More points do not require a larger model.
- These two characteristics mean that increasing the number of data points usually leads to a better model if the model structure is good enough to explain the data. When a poorer model appears with increasing data, this often means that the model structure is insufficient.

Given this, it is not surprising that system identification relies more often on curve fitting than on interpolation. The most usual method of curve fitting is finding the least squares solution to an overdetermined system, which will only have an exact solution under ideal conditions — an exact model, no noise, no measurement errors, etc. — seldom, if ever, found in practice. The least-squares solution of a linear problem can be found using the pseudo-inverse of a matrix. To define it, some results are needed.

System identification uses curve fitting often, interpolation seldom

Definition 30.1. Let \mathbf{x} be a $n \times 1$ vector. Given a scalar function of vectorial variable $f(\mathbf{x})$, derivative $\frac{d}{d\mathbf{x}}f(\mathbf{x})$ is defined as

$$\frac{d}{d\mathbf{x}}f(\mathbf{x}) = \begin{bmatrix} \frac{\partial}{\partial \mathbf{x}_1} f(\mathbf{x}) \\ \frac{\partial}{\partial \mathbf{x}_2} f(\mathbf{x}) \\ \vdots \\ \frac{\partial}{\partial \mathbf{x}_n} f(\mathbf{x}) \end{bmatrix} \quad \square \quad (30.30)$$

Lemma 30.1.

$$\frac{d}{d\mathbf{x}}(\mathbf{b}^T \mathbf{x}) = \begin{bmatrix} \frac{\partial}{\partial \mathbf{x}_1} \sum_{k=1}^n \mathbf{b}_k \mathbf{x}_k \\ \frac{\partial}{\partial \mathbf{x}_2} \sum_{k=1}^n \mathbf{b}_k \mathbf{x}_k \\ \vdots \\ \frac{\partial}{\partial \mathbf{x}_n} \sum_{k=1}^n \mathbf{b}_k \mathbf{x}_k \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_n \end{bmatrix} = \mathbf{b} \quad \square \quad (30.31)$$

Lemma 30.2. Let \mathbf{C} be a square $n \times n$ matrix. Then

$$\begin{aligned}
\frac{d}{d\mathbf{x}} (\mathbf{x}^T \mathbf{C} \mathbf{x}) &= \frac{d}{d\mathbf{x}} \left(\begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix} \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} & \cdots & \mathbf{C}_{1n} \\ \mathbf{C}_{21} & \mathbf{C}_{22} & \cdots & \mathbf{C}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}_{n1} & \mathbf{C}_{n2} & \cdots & \mathbf{C}_{nn} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix} \right) \\
&= \frac{d}{d\mathbf{x}} \left(\begin{bmatrix} \sum_{k=1}^n \mathbf{x}_k \mathbf{C}_{k1} & \sum_{k=1}^n \mathbf{x}_k \mathbf{C}_{k2} & \cdots & \sum_{k=1}^n \mathbf{x}_k \mathbf{C}_{kn} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix} \right) \\
&= \begin{bmatrix} \frac{\partial}{\partial \mathbf{x}_1} \sum_{l=1}^n \sum_{k=1}^n \mathbf{x}_k \mathbf{C}_{kl} \mathbf{x}_l \\ \frac{\partial}{\partial \mathbf{x}_2} \sum_{l=1}^n \sum_{k=1}^n \mathbf{x}_k \mathbf{C}_{kl} \mathbf{x}_l \\ \vdots \\ \frac{\partial}{\partial \mathbf{x}_n} \sum_{l=1}^n \sum_{k=1}^n \mathbf{x}_k \mathbf{C}_{kl} \mathbf{x}_l \end{bmatrix} = \begin{bmatrix} \sum_{l=1}^n \mathbf{C}_{1l} \mathbf{x}_l + \sum_{k=1}^n \mathbf{x}_k \mathbf{C}_{k1} \\ \sum_{l=1}^n \mathbf{C}_{2l} \mathbf{x}_l + \sum_{k=1}^n \mathbf{x}_k \mathbf{C}_{k2} \\ \vdots \\ \sum_{l=1}^n \mathbf{C}_{nl} \mathbf{x}_l + \sum_{k=1}^n \mathbf{x}_k \mathbf{C}_{kn} \end{bmatrix} \\
&= \begin{bmatrix} 2 \sum_{m=1}^n \mathbf{C}_{m1} \mathbf{x}_m \\ 2 \sum_{m=1}^n \mathbf{C}_{m2} \mathbf{x}_m \\ \vdots \\ 2 \sum_{m=1}^n \mathbf{C}_{mn} \mathbf{x}_m \end{bmatrix} = 2 \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} & \cdots & \mathbf{C}_{1n} \\ \mathbf{C}_{21} & \mathbf{C}_{22} & \cdots & \mathbf{C}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}_{n1} & \mathbf{C}_{n2} & \cdots & \mathbf{C}_{nn} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix} \\
&= 2\mathbf{C}\mathbf{x} \quad \square
\end{aligned} \tag{30.32}$$

Lemma 30.3. Given any matrix \mathbf{B} , the product $\mathbf{B}^T \mathbf{B}$ is a symmetric matrix.

Proof. Let \mathbf{B} be a $n \times m$ matrix. Then

$$\begin{aligned}
\mathbf{B}^T \mathbf{B} &= \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{21} & \cdots & \mathbf{B}_{n1} \\ \mathbf{B}_{12} & \mathbf{B}_{22} & \cdots & \mathbf{B}_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B}_{1m} & \mathbf{B}_{2m} & \cdots & \mathbf{B}_{nm} \end{bmatrix} \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} & \cdots & \mathbf{B}_{1m} \\ \mathbf{B}_{21} & \mathbf{B}_{22} & \cdots & \mathbf{B}_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B}_{n1} & \mathbf{B}_{n2} & \cdots & \mathbf{B}_{nm} \end{bmatrix} \\
&= \begin{bmatrix} \sum_{k=1}^n \mathbf{B}_{k1} \mathbf{B}_{k1} & \sum_{k=1}^n \mathbf{B}_{k1} \mathbf{B}_{k2} & \cdots & \sum_{k=1}^n \mathbf{B}_{k1} \mathbf{B}_{km} \\ \sum_{k=1}^n \mathbf{B}_{k2} \mathbf{B}_{k1} & \sum_{k=1}^n \mathbf{B}_{k2} \mathbf{B}_{k2} & \cdots & \sum_{k=1}^n \mathbf{B}_{k2} \mathbf{B}_{km} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=1}^n \mathbf{B}_{km} \mathbf{B}_{k1} & \sum_{k=1}^n \mathbf{B}_{km} \mathbf{B}_{k2} & \cdots & \sum_{k=1}^n \mathbf{B}_{km} \mathbf{B}_{km} \end{bmatrix}
\end{aligned} \tag{30.33}$$

Thus, in the $m \times m$ resulting matrix, both element l, c and element c, l are given by $\sum_{k=1}^n \mathbf{B}_{kl} \mathbf{B}_{kc}$. \square

We are now in condition to prove the following.

Theorem 30.1. Let \mathbf{A} be a $m \times n$ matrix, and \mathbf{b} a $m \times 1$ vector. The solution in the least-squares sense of problem

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{30.34}$$

is the $n \times 1$ vector \mathbf{x} given by

$$\mathbf{x} = \mathbf{A}^+ \mathbf{b} \tag{30.35}$$

do-inverse

where the $n \times m$ matrix \mathbf{A}^+ is the **pseudo-inverse** (or Moore-Penrose inverse) of matrix \mathbf{A} , given by

$$\mathbf{A}^+ = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \quad (30.36)$$

Proof. The error is the $m \times 1$ vector

$$\boldsymbol{\varepsilon} = \mathbf{b} - \mathbf{A}\mathbf{x} \quad (30.37)$$

and the least-squares solution minimises the MSE, that is

$$\begin{aligned} \text{MSE} &= \frac{1}{n} \sum_{k=1}^n \varepsilon_k^2 = \frac{1}{n} \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} \\ &= \frac{1}{n} (\mathbf{b} - \mathbf{A}\mathbf{x})^T (\mathbf{b} - \mathbf{A}\mathbf{x}) \\ &= \frac{1}{n} (\mathbf{b}^T - \mathbf{x}^T \mathbf{A}^T) (\mathbf{b} - \mathbf{A}\mathbf{x}) \\ &= \frac{1}{n} (\mathbf{b}^T \mathbf{b} - \mathbf{b}^T \mathbf{A}\mathbf{x} - \mathbf{x}^T \mathbf{A}^T \mathbf{b} + \mathbf{x}^T \mathbf{A}^T \mathbf{A}\mathbf{x}) \end{aligned} \quad (30.38)$$

Because MSE is a scalar, we know immediately that both $\mathbf{b}^T \mathbf{A}\mathbf{x}$ and $\mathbf{x}^T \mathbf{A}^T \mathbf{b} = (\mathbf{b}^T \mathbf{A}\mathbf{x})^T$ are equal. Thus

$$\text{MSE} = \frac{1}{n} (\mathbf{b}^T \mathbf{b} - 2\mathbf{b}^T \mathbf{A}\mathbf{x} + \mathbf{x}^T \mathbf{A}^T \mathbf{A}\mathbf{x}) \quad (30.39)$$

We apply (30.31), (30.32) and (30.33) to get

$$\begin{aligned} \frac{d}{d\mathbf{x}} \text{MSE} &= 0 \\ \Leftrightarrow 0 - 2(\mathbf{b}^T \mathbf{A})^T + 2\mathbf{A}^T \mathbf{A}\mathbf{x} &= 0 \\ \Leftrightarrow \mathbf{A}^T \mathbf{A}\mathbf{x} &= \mathbf{A}^T \mathbf{b} \\ \Leftrightarrow \mathbf{x} &= \underbrace{(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}}_{\mathbf{A}^+} \quad \square \end{aligned} \quad (30.40)$$

Glossary

“Just a few novels, with ‘Paul Alexis’ inside, and some with nothing at all, and one or two paper-backed books written in Chinese.”

“Chinese?”

“Well, it looked like it. Russian, maybe. Not in proper letters, anyhow.”

Dorothy L. SAYERS (1893 — †1957), *Have his carcase* (1932), 10

centred moving average média móvel centrada

coefficient of determination coeficiente de determinação

mean absolute error erro absoluto médio

mean square error erro quadrático médio

offline identification identificação em diferido

real time identification identificação em tempo real

root mean square error raiz do erro quadrático médio, valor eficaz do erro

pseudo-inverse matrix matriz pseudo-inversa

signal to noise ratio razão sinal-ruído, relação sinal-ruído

stochastic model modelo estocástico

variance accounted for variância contabilizada

Exercises

1. Plot the Bode diagram of (30.27) for different values of n . For each case, find the passband and the largest gain in the rejection band.

Chapter 31

Identification from time responses

—Oh ! avec les chiffres on prouve tout ce qu'on veut !

—Et avec les faits, mon garçon, en est-il de même ?

Jules VERNE (1828 — †1905), *Voyage au centre de la Terre* (1864), VI

This chapter concerns identification methods from responses in time.

Some identification methods are based upon time responses for a particular input. The most frequently used inputs for this purpose are:

- steps, not necessarily of unit amplitude (remember the passing comments in Section 10.1);
- squares waves, in lieu of a step (remember the passing comments in Section 10.4);
- triangle waves, in lieu of a ramp (remember the passing comments in Section 10.4).

Responses to sinusoidal inputs are usually obtained to find the frequency response. They can, however, be used as responses to any other signal. Responses to impulses are seldom attempted and never really attained: as we saw in Section 10.1, an impulse is a mathematical abstraction, and in practice the best approximation is a pulse. Things being so, it is better to explicitly consider a square wave instead.

No experimental impulse responses

Other identification methods, as we will see, work with any input whatsoever.

31.1 Identification of the order of the model

The characteristics of the time response can be used to estimate the number of poles and zeros that a model needs to have. Remember in particular the results from Section 11.6.

The following observations apply to step responses; similar considerations for other types of responses are found in a similar manner.

- If there is an overshoot, the plant has at least two poles.
- An overshoot is frequently the result of a pair of complex conjugate poles.
- An overshoot can also be caused by a zero with a frequency which is low when compared with the frequency of the poles.
- Oscillations (which are not the same as an overshoot: an overshoot can occur without being followed by oscillations) are either the result of a pair of complex conjugate poles, or a limit cycle caused by a non-linearity.
- An undershoot implies a non-minimum phase zero.

- The difference between the numbers of zeros and poles can be found from the way the step response behaves for $t = 0$, according to Theorem 11.3. This becomes difficult if noise is so significant that this behaviour in a short period of time is hard to assess.
- The signal can be differentiated to better see how the behaviour for $t = 0$ is. But remember that differentiating amplifies high-frequency noise.
- If the output settles at a steady state value y_∞ , there are no poles at the origin, and the gain of the plant K is the ratio between that steady state value and the amplitude of the step u_∞ :

$$K = \frac{y_\infty}{u_\infty} \quad (31.1)$$

- If the plant is marginally stable, or is unstable because of two or more poles at the origin (integrations), the response can be differentiated one or more times, until a steady state is found. The number of differentiations is the number of poles at the origin. But it is never too often repeated that differentiating amplifies high-frequency noise.

31.2 Identification from analytical characteristics of the response

Once a desired model structure is known, its parameters, or at least some of them, may be found from characteristics of the response. Remember in particular the results from Sections 11.1 to 11.3 and 11.6.

The following observations apply to step responses; similar considerations for other types of responses are found in a similar manner.

- The gain of the plant is found from (31.1).
- The pole of a (stable) first order model, or a (stable) dominant real pole of a model with more poles and zeros, can be found from the settling time, or from any other point given by (11.12)–(11.18) or in the general case by (11.11). The slope at $t = 0$, given by (11.19), can be used to find the gain. Since the response at $t = 0$ is linear with time, this slope can be found fitting a straight line with least squares during a short interval. This corresponds to a low pass filtering of noise.
- The two real poles of an overdamped second order model without zeros can be found as follows:
 - the dominant pole is found as if it were the only one;
 - the other pole is found from the inflection point using (11.54).
- The coefficients of an underdamped second order model without zeros can be found as follows:
 - the damping factor is found from the maximum overshoot using (11.72);
 - the natural frequency is then found from the settling time using (11.73) and the particular values given immediately after;
 - the frequency of the oscillations $\omega_n \sqrt{1 - \xi^2}$ can be used to confirm these two values or to correct one of them once the other is known.

Example 31.1. An electrical motor can be modelled as

$$\frac{\theta(s)}{U(s)} = \frac{10}{s(1 + 0.15s)} \quad (31.2)$$

and its unit step response is shown in Figure 31.1. A linear regression easily shows that the response is going to infinity linearly (not exponentially, as when there is an unstable pole; not quadratically, as when there are two poles at the origin; etc.) and thus that there is one pole at the origin. As an alternative,

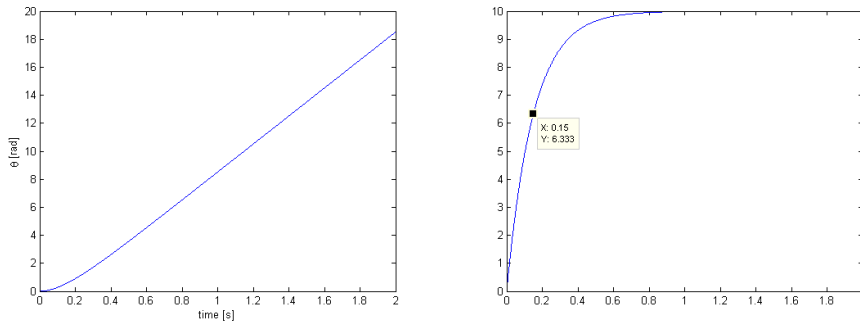


Figure 31.1: Left: unit step response of (31.2), from Example 31.1. Right: its first derivative.

the derivative of $\theta(t)$ can be numerically calculated (with a first order finite difference) and plotted, and, as seen also in Figure 31.1, it is easy to establish for $\frac{\hat{\theta}(s)}{U(s)}$ a first order model with steady state gain 10 and time constant 0.15:

$$\frac{\hat{\theta}(s)}{U(s)} = \frac{10}{1 + 0.15s} \quad (31.3)$$

From this, (31.2) comes immediately. \square

31.3 Identification by minimisation of error

Identifying a model as described in the previous section consists in finding parameters from the values of only some of the measured outputs. Most of the measured values are wasted; this is the price to pay for fairly simple calculations. Using all the measured outputs may lead to more accurate results.

Example 31.2. The response of model $G(s) = \frac{b}{s+a}$ to a step of amplitude K is $y(t) = \frac{bK}{a}(1 - e^{-at})$, as we saw in (11.11).

From the values of $y(t)$ for large t , we can get steady state gain $\frac{b}{a}$. Then we calculate

$$\log\left(y(t) - \frac{bK}{a}\right) = -at \quad (31.4)$$

and a linear regression on t gives a . See Exercise 4 of Chapter 11. \square

While in the Example above the calculations are still fairly simple, this is not the case for models with more than one pole.

Example 31.3. An overdamped second order model with poles which are not too close may still be identified in a likewise manner. Consider the unit step response of

$$G(s) = \frac{1}{(s+1)(s+10)} \quad (31.5)$$

which is given by (see Table 2.1)

$$y(t) = \frac{1}{10} + \frac{1}{10(-9)}(10e^{-t} - e^{-10t}) \quad (31.6)$$

and has a steady state given by $\lim_{t \rightarrow +\infty} y(t) = \frac{1}{10}$, as seen in Figure 31.2.

Suppose that our model uses the correct structure

$$\hat{G}(s) = \frac{\frac{1}{10}p_1p_2}{(s+p_1)(s+p_2)} \quad (31.7)$$

and already includes the correct steady-state gain $\frac{1}{10}$ easily read in Figure 31.2. Plotting

$$\log\left(-\left(y(t) - \frac{1}{10}\right)\right) = \log(90(10e^{-t} - e^{-10t})) = \log 90 + \log(10e^{-t} - e^{-10t}) \quad (31.8)$$

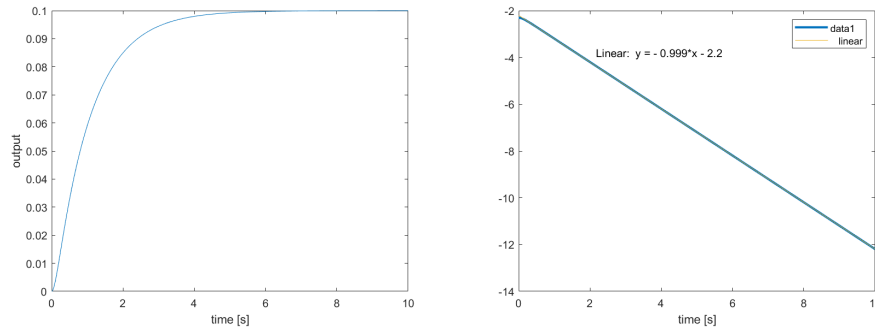


Figure 31.2: Left: unit step response of (31.5), from Example 31.3. Right: plot of (31.8).

gives us a practically straight line, since the faster pole (which is -10 , as we saw in Section 11.6) has a vanishing effect that can only be seen near $t = 0$. The slope shown in Figure 31.2 gives us the pole $p_1 = -1$.

As to the other pole, it could be found from the inflection point using (11.54), but it is quite hard to find it using a method similar to the one used for p_1 , since we now know that

$$\begin{aligned}
 y(t) &= \frac{p_2}{10} \frac{1}{p_2} \left(1 + \frac{1}{1-p_2} (p_2 e^{-t} - e^{-p_2 t}) \right) \Rightarrow \\
 (10y(t) - 1)(1 - p_2) &= p_2 e^{-t} - e^{-p_2 t} \Rightarrow \\
 -p_2 t &= \log(p_2 e^{-t} - (10y(t) - 1)(1 - p_2)) \quad (31.9)
 \end{aligned}$$

This expression has p_2 on both sides, and so we would have to find some value p_2 with which the right hand side of 31.9 is a straight line for small values of t , with slope $-p_2$. Searching for this value can be done by trial and error until $p_2 = 10$ is found, but there are better methods than that. \square

Remark 31.1. The method of Example 31.3 can be used for other plants with a dominant real negative pole. \square

In the general case, finding the parameters of a continuous model that provide the best fit between its response and the experimental data is a non-linear problem. Consequently, optimisation methods such as the Nelder-Mead simplex search method (implemented in MATLAB with function `fminsearch`), or metaheuristic methods (of which genetic algorithms and particle swarm optimisation are well known examples) can be used. We will not study such methods, but once you learn them in a course on that subject you can apply them easily in identification:

- The objective is to minimise one of the performance indexes (30.2)–(30.10), or a linear combination thereof.
- An initial estimate of the parameters is provided if available; otherwise, random values are used to begin with.
- Some algorithms use only one initial estimate; others use several different sets of values.
- These algorithms are iterative, and in each iteration the time response of the plant for the parameters currently being evaluated is found and compared with the experimental data.
- In this way, identification can be performed with responses to any input, not just steps.

Metaheuristics for identification

Metaheuristics do not guarantee the optimal result, and not even good results

Identifying parameters of models in z^{-1}

This strategy can work for complicated problems, but can also get stuck because of numerical problems, and these algorithms do not ensure that the parameters they find actually correspond to the minimum value of the performance index.

The problem is far easier if discrete models

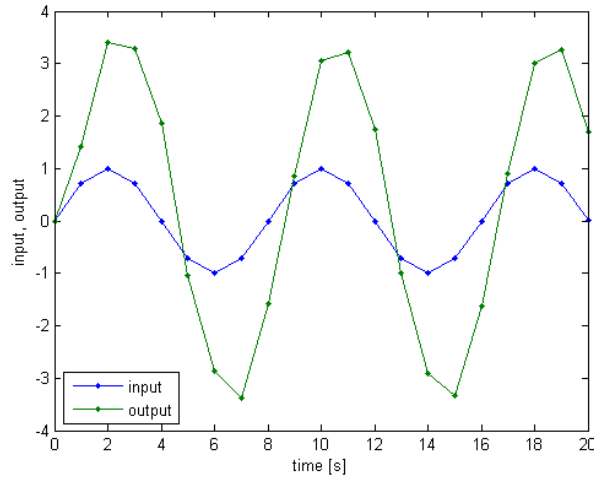


Figure 31.3: Output of (31.14) for a sinusoidal output and normally distributed error.

$$\frac{y_k}{u_k} = \frac{b_0 + b_1 z^{-1} + \dots + b_m z^{-m}}{1 + a_1 z^{-1} + \dots + a_n z^{-n}} \Rightarrow y_k = b_0 u_k + b_1 u_{k-1} + \dots + b_m u_{k-m} - a_1 y_{k-1} - \dots - a_n y_{k-n} \quad (31.10)$$

are used instead, since we can arrange data and parameters to be identified as

$$\begin{matrix} \mathbf{y} \\ \left[\begin{array}{c} y_k \\ y_{k-1} \\ y_{k-2} \\ \vdots \end{array} \right] \end{matrix} = \begin{matrix} \mathbf{A} \\ \left[\begin{array}{ccccccc} u_k & u_{k-1} & \cdots & u_{k-m} & -y_{k-1} & \cdots & -y_{k-n} \\ u_{k-1} & u_{k-2} & \cdots & u_{k-m-1} & -y_{k-2} & \cdots & -y_{k-n-1} \\ u_{k-2} & u_{k-3} & \cdots & u_{k-m-2} & -y_{k-3} & \cdots & -y_{k-n-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{array} \right] \end{matrix} \begin{matrix} \boldsymbol{\theta} \\ \left[\begin{array}{c} b_0 \\ b_1 \\ \vdots \\ b_m \\ a_1 \\ \vdots \\ a_n \end{array} \right] \end{matrix} \quad (31.11)$$

and then solve for $\boldsymbol{\theta}$. We need \mathbf{y} and \mathbf{A} to have at least $n + m + 1$ rows, and thus $n + m + \max\{n, m\}$ samples of u_k and y_k are needed. Likely there will be more, and $\mathbf{A}\boldsymbol{\theta} = \mathbf{y}$ will be overdetermined. A solution in the least-squares sense is then found, as seen in Section 30.4:

$$\boldsymbol{\theta} = \mathbf{A}^+ \mathbf{y} \quad (31.12)$$

Example 31.4. Consider plant

$$G(z^{-1}) = \frac{2 + 3z^{-1}}{1 + 0.5z^{-1} - 0.4z^{-2}} \quad (31.13)$$

with sample time $T_s = 1$ s, which, allowing for additive noise at the output, corresponds to the difference equation

$$\begin{aligned} y_k &= -0.5y_{k-1} + 0.4y_{k-2} + 2u_k + 3u_{k-1} + e_k \\ \Leftrightarrow y_k &= a_1(-y_{k-1}) + a_2(-y_{k-2}) + b_0 u_k + b_1 u_{k-1} + e_k, \\ a_1 &= 0.5, a_2 = -0.4, b_0 = 2, b_1 = 3 \end{aligned} \quad (31.14)$$

Its output is tabulated in Table 31.1 and shown in Figure 31.3 for a sinusoidal input, when the error is normally distributed with variance 0.001.

Suppose we know that the model has two poles and one zero, i.e. we know that it is given by (31.14), and want to know the values of parameters a_1 , a_2 , b_0 and b_1 . We know that there will be an error, and since we cannot measure it or estimate it the best we can do is to replace it by its expected value, which

Table 31.1: Input and output data of Figure 31.3.

$t[s]$	0	1	2	3	4	5	6
input	0.0000	0.7071	1.0000	0.7071	0.0000	-0.7071	-1.0000
output	0.0007	1.4237	3.4108	3.2732	1.8618	-1.0272	-2.8700
$t[s]$	7	8	9	10	11	12	13
input	-0.7071	0.0000	0.7071	1.0000	0.7071	0.0000	-0.7071
output	-3.3999	-1.5776	0.8401	3.0635	3.2140	1.7455	-1.0050
$t[s]$	14	15	16	17	18	19	20
input	-1.0000	-0.7071	0.0000	0.7071	1.0000	0.7071	0.0000
output	-2.9193	-3.3469	-1.6351	0.8849	3.0279	3.2467	1.7052

we assume to be $E[e_k] = 0$. We can write

$$y_{20} = a_1(-y_{19}) + a_2(-y_{18}) + b_0u_{20} + b_1u_{19} \quad (31.15)$$

$$y_{19} = a_1(-y_{18}) + a_2(-y_{17}) + b_0u_{19} + b_1u_{18} \quad (31.16)$$

$$y_{18} = a_1(-y_{17}) + a_2(-y_{16}) + b_0u_{18} + b_1u_{17} \quad (31.17)$$

⋮

$$y_2 = a_1(-y_1) + a_2(-y_0) + b_0u_2 + b_1u_1 \quad (31.18)$$

$$y_1 = a_1(-y_0) + a_2(-y_{-1}) + b_0u_1 + b_1u_0 \quad (31.19)$$

$$y_0 = a_1(-y_{-1}) + a_2(-y_{-2}) + b_0u_0 + b_1u_{-1} \quad (31.20)$$

We can now replace values, assuming that before $t = 0$ s both input and output were 0 (if we could not assume this, we would just discard the last two equations right away):

$$1.7052 = -3.2467a_1 - 3.0279a_2 + 0.0000b_0 + 0.7071b_1 \quad (31.21)$$

$$3.2467 = -3.0279a_1 - 0.8849a_2 + 0.7071b_0 + 1.0000b_1 \quad (31.22)$$

$$3.0279 = -0.8849a_1 + 1.6351a_2 + 1.0000b_0 + 0.7071b_1 \quad (31.23)$$

⋮

$$3.4108 = -1.4237a_1 - 0.0007a_2 + 1.0000b_0 + 0.7071b_1 \quad (31.24)$$

$$1.4237 = -0.0007a_1 + 0.0000a_2 + 0.7071b_0 + 0.0000b_1 \quad (31.25)$$

$$0.0007 = 0.0000a_1 + 0.0000a_2 + 0.0000b_0 + 0.0000b_1 \quad (31.26)$$

The last equation being clearly impossible — which is a result of the output being nothing but noise without depending from the parameters in any way —, we discard it; the other 20 equations can be easily put in the form of (31.11). Solving for θ ,

$$\theta = \begin{bmatrix} a_1 \\ a_2 \\ b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} 0.4969 \\ -0.4024 \\ 2.0114 \\ 2.9812 \end{bmatrix} \quad \square \quad (31.27)$$

31.4 Deconvolution

We know from (10.25) that the output of a transfer function is given by the convolution of its impulse response with the input. If signals are discretised in time, we can approximate the integral in (10.25)

$$y(t) = \int_0^t g(t - \tau)u(\tau) d\tau \quad (31.28)$$

using a rectangular forward approximation (illustrated in Figure 31.4):

$$y_k = T_s \sum_{i=0}^{k-1} u_i g_{k-1-i} \quad (31.29)$$

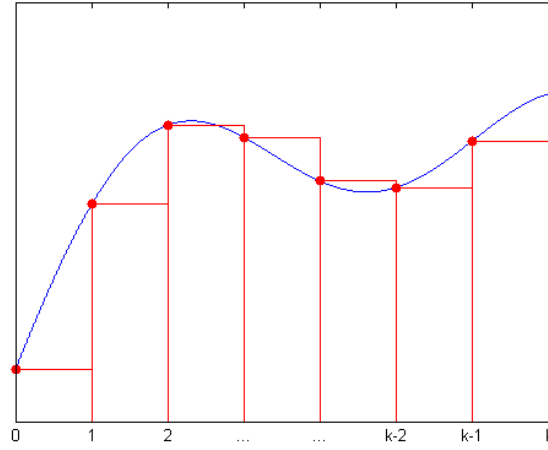


Figure 31.4: Rectangular forward approximation of an integral, consisting of the areas of the rectangles built from the highlighted points.

Data to be identified, from $t = 0$ to $t_f = N T_s$, can be arranged in matrixes:

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_N \end{bmatrix}}_{\substack{N \text{ rows} \\ \mathbf{y}}} = T_s \underbrace{\begin{bmatrix} u_0 & 0 & 0 & \cdots & 0 \\ u_1 & u_0 & 0 & \cdots & 0 \\ u_2 & u_1 & u_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ u_{N-1} & u_{N-2} & u_{N-3} & \cdots & u_0 \end{bmatrix}}_{\substack{N \times N \text{ matrix} \\ \mathbf{U}}} \underbrace{\begin{bmatrix} g_0 \\ g_1 \\ g_2 \\ \vdots \\ g_{N-1} \end{bmatrix}}_{\substack{N \text{ rows} \\ \mathbf{g}}} \quad (31.30)$$

The impulse response \mathbf{g} is found as $\mathbf{g} = \mathbf{U}^{-1}\mathbf{y}$. Since \mathbf{U} is square, there is no need to use a pseudo-inverse, and, because it is triangular, the solution can be found solving the equations line by line:

$$g_0 = \frac{1}{u_0} \frac{y_1}{T_s} \quad (31.31)$$

$$g_1 = \frac{1}{u_0} \left(\frac{y_2}{T_s} - u_1 g_0 \right) \quad (31.32)$$

$$g_2 = \frac{1}{u_0} \left(\frac{y_3}{T_s} - u_2 g_0 - u_1 g_1 \right) \quad (31.33)$$

\vdots

$$g_k = \frac{1}{u_0} \left(\frac{y_{k+1}}{T_s} - u_k g_0 - u_{k-1} g_1 - \cdots - u_1 g_{k-1} \right) \quad (31.34)$$

\vdots

$$g_{N-1} = \frac{1}{u_0} \left(\frac{y_N}{T_s} - u_{N-1} g_0 - u_{N-2} g_1 - \cdots - u_1 g_{N-2} \right) \quad (31.35)$$

On the other hand, we can force (31.30) to be overdetermined by calculating the impulse response g_k in less time instants than the N time instants for which this is possible: *Filtering noise in deconvolution*

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_N \end{bmatrix}}_{N \times 1} = T_s \underbrace{\begin{bmatrix} u_0 & 0 & 0 & \cdots & 0 \\ u_1 & u_0 & 0 & \cdots & 0 \\ u_2 & u_1 & u_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ u_{N-1} & u_{N-2} & u_{N-3} & \cdots & u_0 \end{bmatrix}}_{N \times K} \underbrace{\begin{bmatrix} g_0 \\ g_1 \\ g_2 \\ \vdots \\ g_{K-1} \end{bmatrix}}_{K \times 1}, \quad K < N \quad (31.36)$$

In this case the impulse response vector \mathbf{g} must be found in a least-squares sense:

$$\mathbf{g} = \frac{1}{T_s} \mathbf{U}^+ \mathbf{y} \quad (31.37)$$

This alternative is useful to filter high-frequency noise.

Since a convolution finds the response to an arbitrary input from the input and the impulse response, and now the impulse response is determined from an arbitrary input and its response, this operation (31.31)–(31.35) or (31.37) is called **deconvolution**.

Once the impulse response of the model is found, it can be used in several different ways.

- It can be used to find the model's response to an arbitrary input, by convolution.
- It can in particular be used to find the model's response to sinusoids, so as to obtain the model's frequency response.
- The Laplace transform of the impulse response can be found numerically. We know from Theorem 10.1 that the result is the transfer function of the model. Since we will only have numerical values, and not an analytical expression, this may not be very useful. But instead of the Laplace transform (with $s \in \mathbb{C}$) we can just find the Fourier transform (2.87), thereby obtaining the model's frequency response without the cumbersome and tedious intermediate step of sinusoidal inputs.
- The impulse response can be used to obtain the parameters of a model, either using (31.11) or using a simplified formulation as in the following Example, which is possible because the impulse is different from zero in one time instant only.

Example 31.5. Consider model

$$G(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}} \Rightarrow$$

$$y_k = b_0 u_k + b_1 u_{k-1} + b_2 u_{k-2} - a_1 y_{k-1} - a_2 y_{k-2} \quad (31.38)$$

Replacing the impulse

$$u_0 = \frac{1}{T_s} \quad (31.39)$$

$$u_k = 0, \quad \forall k \neq 0 \quad (31.40)$$

in (31.38), the output y_k will be the impulse response g_k :

$$g_0 = b_0 \frac{1}{T_s} \quad (31.41)$$

$$g_1 = b_1 \frac{1}{T_s} - a_1 g_0 \quad (31.42)$$

$$g_2 = b_2 \frac{1}{T_s} - a_1 g_1 - a_2 g_0 \quad (31.43)$$

$$g_3 = -a_1 g_2 - a_2 g_1 \quad (31.44)$$

$$g_4 = -a_1 g_3 - a_2 g_2 \quad (31.45)$$

⋮

There are 5 unknowns (2 on the denominator and 3 on the numerator), which can be found with at least 5 values of the impulse response, corresponding to the 5 equations above.

Using more values of g_k , and thus more equations, the solution of the overdetermined system will once more be found in a least-squares sense (and be more robust to noise). \square

Remark 31.2. (31.41)–(31.45) can be arranged in matrix form, collecting the terms on numerator coefficients b_k on one side and all the other terms on the other:

$$\frac{1}{T_s} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ 0 \\ 0 \\ \vdots \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \cdots \\ a_1 & 1 & 0 & 0 & 0 & \cdots \\ a_2 & a_1 & 1 & 0 & 0 & \cdots \\ 0 & a_2 & a_1 & 1 & 0 & \cdots \\ 0 & 0 & a_2 & a_1 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} g_0 \\ g_1 \\ g_2 \\ g_3 \\ g_4 \\ \vdots \end{bmatrix} \quad (31.46)$$

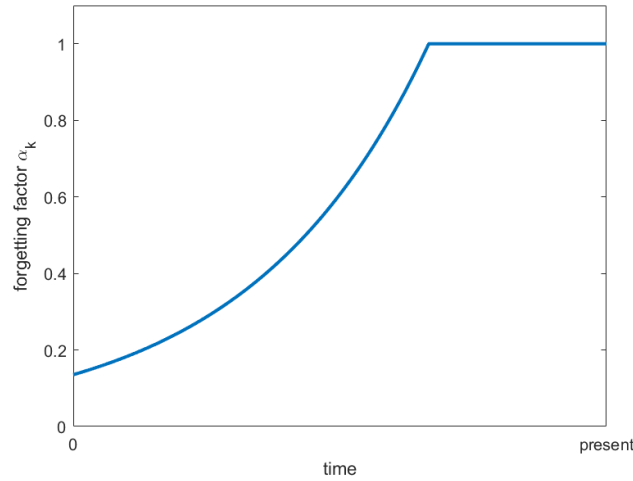


Figure 31.5: Example of forgetting factor α_i for online identification. In this case, the weight decreases exponentially; a power law or a linear decrease can be used instead.

This is not as useful as (31.11) or (31.30), since the coefficients are in the matrix, but gives a better idea of the general case. \square

31.5 Real time identification

When identification is carried out in real time, applying (31.12) in each sampling instant k means that a (possibly overdetermined) linear system of equations, with an ever-increasing size, will have to be solved, to find a completely different set of parameters θ_k . It is possible to improve this method in two ways:

- the estimate of the parameters θ_k at time instant k is obtained from the previous estimate as

$$\theta_k = \theta_{k-1} + \Delta\theta_k \quad (31.47)$$

and so only the variation $\Delta\theta_k$ has to be determined;

- a **forgetting factor** α_i , $i = 0, 1, \dots, k$ is used, so that older measurements carry less weight than recent ones. In this way, it is possible to identify a *quasi-stationary* process, i.e. a process which is not time invariant but which undergoes changes in parameters much slower than its dynamics. Of course, $\alpha_i = 1, \forall i$ can be used to get the same effect of employing no forgetting factor at all. Otherwise, something like what is shown in Figure 31.5 is used.

Forgetting factor

Quasi-stationary process

These changes can be used in offline identification as well: in this case, the algorithm is run while being fed previously recorded data successively, as if it were being measured online. The algorithm can be represented in a block diagram as shown in Figure 31.6.

It is expedient to separate the several lines of matrix \mathbf{A} from (31.11) as follows:

$$\mathbf{a}_k = \overbrace{[u_k \quad u_{k-1} \quad \cdots \quad u_{k-m} \quad -y_{k-1} \quad \cdots \quad -y_{k-n}]}^{1 \times (m+n+1)} \quad (31.48)$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_k \\ \mathbf{a}_{k-1} \\ \mathbf{a}_{k-2} \\ \vdots \\ \mathbf{a}_0 \end{bmatrix} \quad (31.49)$$

$(k+1) \times (m+n+1)$

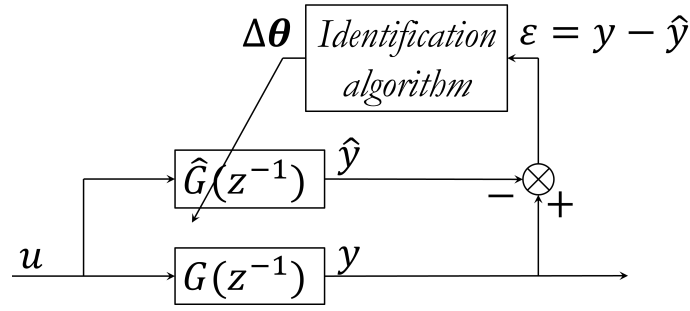


Figure 31.6: Block diagram for online identification by error minimisation. The arrow crossing the block with model $\hat{G}(s)$ of process $G(s)$ means that the transfer function is modified.

This allows rewriting the least-squares solution (31.12) as

$$\begin{aligned}
 \boldsymbol{\theta} &= \mathbf{A}^+ \mathbf{y} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y} \\
 &= \left(\begin{bmatrix} \mathbf{a}_k^T & \mathbf{a}_{k-1}^T & \cdots & \mathbf{a}_0^T \end{bmatrix} \begin{bmatrix} \mathbf{a}_k \\ \mathbf{a}_{k-1} \\ \vdots \\ \mathbf{a}_0 \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{a}_k^T & \mathbf{a}_{k-1}^T & \cdots & \mathbf{a}_0^T \end{bmatrix} \begin{bmatrix} y_k \\ y_{k-1} \\ \vdots \\ y_0 \end{bmatrix} \\
 &= \underbrace{\left(\sum_{i=0}^k \mathbf{a}_i^T \mathbf{a}_i \right)^{-1}}_{\mathbf{R}_k^{-1}} \sum_{i=0}^k \mathbf{a}_i^T y_i \tag{31.50}
 \end{aligned}$$

The forgetting factor can then be added both to matrix \mathbf{R}_k and output y_k :

$$\boldsymbol{\theta} = \mathbf{R}_k^{-1} \sum_{i=0}^k \alpha_i y_i \mathbf{a}_i^T \tag{31.51}$$

$$\mathbf{R}_k = \underbrace{\sum_{i=0}^k \alpha_i \mathbf{a}_i^T \mathbf{a}_i}_{(n+m+1) \times (n+m+1)} \tag{31.52}$$

Lemma 31.1. The least-squares solution (31.12) of the identification problem, with a forgetting factor α_k , can be recursively found as

$$\boldsymbol{\theta}_k = \boldsymbol{\theta}_{k-1} + \alpha_k \mathbf{R}_k^{-1} \mathbf{a}_k^T (y_k - \mathbf{a}_k \boldsymbol{\theta}_{k-1}) \tag{31.53}$$

$$\mathbf{R}_k = \mathbf{R}_{k-1} + \alpha_k \mathbf{a}_k^T \mathbf{a}_k \tag{31.54}$$

Proof. From (31.51), we can get two expressions for the estimate $\boldsymbol{\theta}$ at time instant $k-1$ and time instant k :

$$\mathbf{R}_{k-1} \boldsymbol{\theta}_{k-1} = \sum_{i=0}^{k-1} \alpha_i y_i \mathbf{a}_i^T \tag{31.55}$$

$$\begin{aligned}
 \mathbf{R}_k \boldsymbol{\theta}_k &= \sum_{i=0}^k \alpha_i y_i \mathbf{a}_i^T \\
 &= \sum_{i=0}^{k-1} \alpha_i y_i \mathbf{a}_i^T + \alpha_k y_k \mathbf{a}_k^T \\
 &= \mathbf{R}_{k-1} \boldsymbol{\theta}_{k-1} + \alpha_k y_k \mathbf{a}_k^T \tag{31.56}
 \end{aligned}$$

From (31.52), we can get an expression for matrix \mathbf{R} at time instant $k-1$:

$$\begin{aligned}
 \mathbf{R}_k &= \underbrace{\sum_{i=0}^{k-1} \alpha_i \mathbf{a}_i^T \mathbf{a}_i}_{\mathbf{R}_{k-1}} + \alpha_k \mathbf{a}_k^T \mathbf{a}_k \\
 \Rightarrow \mathbf{R}_{k-1} &= \mathbf{R}_k - \alpha_k \mathbf{a}_k^T \mathbf{a}_k \tag{31.57}
 \end{aligned}$$

Replacing this in (31.56),

$$\begin{aligned}\boldsymbol{\theta}_k &= \mathbf{R}_k^{-1} (\mathbf{R}_{k-1} \boldsymbol{\theta}_{k-1} + \alpha_k y_k \mathbf{a}_k^T) \\ &= \mathbf{R}_k^{-1} (\mathbf{R}_k - \alpha_k \mathbf{a}_k^T \mathbf{a}_k) \boldsymbol{\theta}_{k-1} + \mathbf{R}_k^{-1} \alpha_k y_k \mathbf{a}_k^T \\ &= \mathbf{R}_k^{-1} \mathbf{R}_k \boldsymbol{\theta}_{k-1} - \alpha_k \mathbf{R}_k^{-1} \mathbf{a}_k^T \mathbf{a}_k \boldsymbol{\theta}_{k-1} + \alpha_k y_k \mathbf{R}_k^{-1} \mathbf{a}_k^T\end{aligned}\quad (31.58)$$

whence (31.53) follows immediately. \square

With (31.53)–(31.54), there are no longer matrixes with an ever-increasing size, but the problem of having to solve a linear system of equations in each iteration remains (there is a matrix inverse in (31.53)). This can be solved with the following lemma:

Lemma 31.2.

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} (\mathbf{C}^{-1} + \mathbf{DA}^{-1} \mathbf{B})^{-1} \mathbf{DA}^{-1} \quad (31.59)$$

provided that matrixes have compatible dimensions, and that matrixes \mathbf{A} and \mathbf{C} are square and invertible.

Proof.

$$\begin{aligned}& (\mathbf{A} + \mathbf{BCD}) \left(\mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} (\mathbf{C}^{-1} + \mathbf{DA}^{-1} \mathbf{B})^{-1} \mathbf{DA}^{-1} \right) \\ &= \mathbf{AA}^{-1} + \mathbf{BCDA}^{-1} - \mathbf{AA}^{-1} \mathbf{B} (\mathbf{C}^{-1} + \mathbf{DA}^{-1} \mathbf{B})^{-1} \mathbf{DA}^{-1} - \mathbf{BCDA}^{-1} \mathbf{B} (\mathbf{C}^{-1} + \mathbf{DA}^{-1} \mathbf{B})^{-1} \mathbf{DA}^{-1} \\ &= \mathbf{I} + \mathbf{BCDA}^{-1} - (\mathbf{B} + \mathbf{BCDA}^{-1} \mathbf{B}) (\mathbf{C}^{-1} + \mathbf{DA}^{-1} \mathbf{B})^{-1} \mathbf{DA}^{-1} \\ &= \mathbf{I} + \underbrace{\mathbf{BCDA}^{-1} - \mathbf{BC} \overbrace{(\mathbf{C}^{-1} + \mathbf{DA}^{-1} \mathbf{B})^{-1} (\mathbf{C}^{-1} + \mathbf{DA}^{-1} \mathbf{B})}^{\mathbf{I}} \mathbf{DA}^{-1}}_{\mathbf{0}} = \mathbf{I}\end{aligned}\quad (31.60)$$

In the above, the inverses of \mathbf{A} and \mathbf{C} appear, and if they exist the calculations hold, irrespective of the values of \mathbf{B} and \mathbf{D} . \square

Theorem 31.1. The least-squares solution (31.12) of the identification problem, with a forgetting factor α_k , can be recursively found as

These are the equations implemented for real time identification

$$\boldsymbol{\theta}_k = \boldsymbol{\theta}_{k-1} + \frac{\mathbf{P}_{k-1} \mathbf{a}_k^T}{\frac{1}{\alpha_k} + \mathbf{a}_k \mathbf{P}_{k-1} \mathbf{a}_k^T} (y_k - \mathbf{a}_k \boldsymbol{\theta}_{k-1}) \quad (31.61)$$

$$\mathbf{P}_k = \mathbf{P}_{k-1} - \frac{\mathbf{P}_{k-1} \mathbf{a}_k^T \mathbf{a}_k \mathbf{P}_{k-1}}{\frac{1}{\alpha_k} + \mathbf{a}_k \mathbf{P}_{k-1} \mathbf{a}_k^T} \quad (31.62)$$

Proof. Using (31.59) in (31.54),

$$\begin{aligned}\mathbf{R}_k &= \overbrace{\mathbf{R}_{k-1}}^{\mathbf{A}} + \overbrace{\mathbf{a}_k^T}^{\mathbf{B}} \overbrace{\alpha_k}^{\mathbf{C}} \overbrace{\mathbf{a}_k}^{\mathbf{D}} \\ \Rightarrow \mathbf{R}_k^{-1} &= \mathbf{R}_{k-1}^{-1} - \mathbf{R}_{k-1}^{-1} \mathbf{a}_k^T \left(\frac{1}{\alpha_k} + \mathbf{a}_k \mathbf{R}_{k-1}^{-1} \mathbf{a}_k^T \right)^{-1} \mathbf{a}_k \mathbf{R}_{k-1}^{-1}\end{aligned}\quad (31.63)$$

Notice that:

- matrix $\mathbf{C} = \alpha_k$ is in fact a scalar;
- matrixes \mathbf{R} and \mathbf{R}^{-1} are square, with dimensions $(n+m+1) \times (n+m+1)$;
- vector \mathbf{a}_k has dimensions $1 \times (n+m+1)$, and thus vector \mathbf{a}_k^T has dimensions $(n+m+1) \times 1$;
- thus, product $\mathbf{a}_k \mathbf{R}_{k-1}^{-1} \mathbf{a}_k^T$ is a scalar, and so is $\left(\frac{1}{\alpha_k} + \mathbf{a}_k \mathbf{R}_{k-1}^{-1} \mathbf{a}_k^T \right)^{-1}$;
- product $\mathbf{a}_k^T \mathbf{a}_k$ has dimensions $(n+m+1) \times (n+m+1)$, and so does product $\mathbf{R}_{k-1}^{-1} \mathbf{a}_k^T \mathbf{a}_k \mathbf{R}_{k-1}^{-1}$.

We can thus write (31.63) as

$$\mathbf{R}_k^{-1} = \mathbf{R}_{k-1}^{-1} - \frac{\mathbf{R}_{k-1}^{-1} \mathbf{a}_k^T \mathbf{a}_k \mathbf{R}_{k-1}^{-1}}{\frac{1}{\alpha_k} + \mathbf{a}_k \mathbf{R}_{k-1}^{-1} \mathbf{a}_k^T} \quad (31.64)$$

Replacing this result in (31.53),

$$\begin{aligned} \boldsymbol{\theta}_k &= \boldsymbol{\theta}_{k-1} + \alpha_k \left(\mathbf{R}_{k-1}^{-1} - \frac{\mathbf{R}_{k-1}^{-1} \mathbf{a}_k^T \mathbf{a}_k \mathbf{R}_{k-1}^{-1}}{\frac{1}{\alpha_k} + \mathbf{a}_k \mathbf{R}_{k-1}^{-1} \mathbf{a}_k^T} \right) \mathbf{a}_k^T (y_k - \mathbf{a}_k \boldsymbol{\theta}_{k-1}) \\ &= \boldsymbol{\theta}_{k-1} + \left(\alpha_k \mathbf{R}_{k-1}^{-1} \mathbf{a}_k^T - \frac{\alpha_k \mathbf{R}_{k-1}^{-1} \mathbf{a}_k^T \mathbf{a}_k \mathbf{R}_{k-1}^{-1} \mathbf{a}_k^T}{\frac{1}{\alpha_k} + \mathbf{a}_k \mathbf{R}_{k-1}^{-1} \mathbf{a}_k^T} \right) (y_k - \mathbf{a}_k \boldsymbol{\theta}_{k-1}) \\ &= \boldsymbol{\theta}_{k-1} + \frac{\alpha_k \mathbf{R}_{k-1}^{-1} \mathbf{a}_k^T \left(\frac{1}{\alpha_k} + \mathbf{a}_k \mathbf{R}_{k-1}^{-1} \mathbf{a}_k^T \right) - \alpha_k \mathbf{R}_{k-1}^{-1} \mathbf{a}_k^T \mathbf{a}_k \mathbf{R}_{k-1}^{-1} \mathbf{a}_k^T}{\frac{1}{\alpha_k} + \mathbf{a}_k \mathbf{R}_{k-1}^{-1} \mathbf{a}_k^T} (y_k - \mathbf{a}_k \boldsymbol{\theta}_{k-1}) \\ &= \boldsymbol{\theta}_{k-1} + \frac{\mathbf{R}_{k-1}^{-1} \mathbf{a}_k^T}{\frac{1}{\alpha_k} + \mathbf{a}_k \mathbf{R}_{k-1}^{-1} \mathbf{a}_k^T} (y_k - \mathbf{a}_k \boldsymbol{\theta}_{k-1}) \end{aligned} \quad (31.65)$$

(31.61)–(31.62) are in fact (31.65) and (31.64), making $\mathbf{R}^{-1} = \mathbf{P}$ to stress that there is no matrix inversion involved (there are only matrix multiplications and divisions by scalars). \square

In practice:

Initialising the algorithm

- if there is an estimate of $\boldsymbol{\theta}$, this is used as initial value;
- otherwise, the initial value $\boldsymbol{\theta} = \mathbf{0}$ is used;
- we should have

$$\mathbf{R}_0 = \begin{bmatrix} u_0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{bmatrix} u_0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix} \quad (31.66)$$

and thus $\mathbf{P}_0 = \mathbf{R}_0^{-1}$ should have $\frac{1}{u_0}$ in line 1, column 1, and be zero otherwise. But this often brings numerical problems. Consequently, it is usual to make $\mathbf{P}_0 = c\mathbf{I}$, where $c > 0$ is a scalar. Larger values of c make $\boldsymbol{\theta}$ update (and thus hopefully converge) faster, but sometimes may cause the identification algorithm to become unstable.

- (31.61)–(31.62) are then applied in each sampling instant.

31.6 Digital model validation

To validate a model's time responses, we give it an input and compare its output with the experimental response. Suppose that the model is discrete in time:

$$\begin{aligned} \frac{y_k}{u_k} &= \frac{b_0 + b_1 z^{-1} + \dots + b_m z^{-m}}{1 + a_1 z^{-1} + \dots + a_n z^{-n}} \Leftrightarrow \\ y_k &= b_0 u_k + b_1 u_{k-1} + \dots + b_m u_{k-m} - a_1 y_{k-1} - \dots - a_n y_{k-n} \end{aligned} \quad (31.67)$$

Let the experimental data at instant k be y_k , and the estimate provided by the model be \hat{y}_k . We can validate the model in several different ways.

- Giving the model an input and comparing it to the experimental response, as mentioned above, is

$$\hat{y}_k = b_0 u_k + b_1 u_{k-1} + \dots + b_m u_{k-m} - a_1 \hat{y}_{k-1} - \dots - a_n \hat{y}_{k-n} \quad (31.68)$$

Table 31.2: Unit step response of Exercise 2.

time	output	time	output	time	output
0.0	0.0000	0.7	0.1225	1.4	0.4900
0.1	0.0025	0.8	0.1600	1.5	0.5625
0.2	0.0100	0.9	0.2025	1.6	0.6400
0.3	0.0225	1.0	0.2500	1.7	0.7225
0.4	0.0400	1.1	0.3025	1.8	0.8100
0.5	0.0625	1.2	0.3600	1.9	0.9025
0.6	0.0900	1.3	0.4225	2.0	1.0000

step ahead prediction

- On the other hand, since the experimental output is known, we can make

$$\hat{y}_k = b_0 u_k + b_1 u_{k-1} + \dots + b_m u_{k-m} - a_1 y_{k-1} - \dots - a_n y_{k-n} \quad (31.69)$$

Clearly (31.69) is not as demanding on the model as (31.68). This is called **one step ahead prediction**, since the model is only required to provide an estimate of the output one time step after the data it is provided with.

- One step ahead prediction can be generalised into N step ahead prediction: *N step ahead prediction*

$$\hat{y}_k = b_0 u_k + b_1 u_{k-1} + \dots + b_m u_{k-m} - a_1 \hat{y}_{k-1} - \dots - a_{N-1} \hat{y}_{k-N+1} - a_N y_{k-N} - \dots - a_n y_{k-n} \quad (31.70)$$

The more estimated outputs \hat{y} are used instead of experimental outputs y , the more is being demanded from the model.

Glossary

Though it be appointed in the afore written preface, that al thinges shalbe read and sōg in the churchē, in the Englishe tongue, to thende $\frac{1}{2}$ the congregacion maie be therby edified: yet it is not meant, but when men saye Matins and Euensong priuately, they maye saie the same in any language that they themselues do understande.

Thomas CRANMER (attrib.; 1489 — †1556), *The booke of the common prayer and administracion of the Sacramentes, and other rites and ceremonies of the Churchē: after the use of the Churchē of England* (1549), The preface

deconvolution desconvolução

forgetting factor fator de esquecimento

genetic algorithms algoritmos genéticos

metaheuristic meta-heurística

Nelder-Mead simplex search method método do simplex, método de Nelder-Mead

N step ahead prediction predição a N passos

one step ahead prediction predição a um passo

particle swarm optimisation otimização por enxame de partículas

quasi-stationary process processo quase-estacionário

Exercises

1. Find numerically the unit step response of

$$G(s) = \frac{1}{(s + 0.1)(s^2 + 0.3s + 1)} \quad (31.71)$$

for the first 60 s, using a sampling time you find appropriate. Use the method of Example 31.3 to find the dominant pole from the response.

2. Find a model for a plant with the unit step response $y(t)$ in Table 31.2.
3. Consider the data from Exercise 4 of Chapter 11.

- (a) Find discrete models for this plant, using (31.11). Try different numbers of zeros and poles, and see how some performance index changes.
 - (b) Find the plant's impulse response, using (31.37) with different values of K . Explain the different results obtained.
4. Repeat Exercise 3 using the data from Exercise 2 instead.
5. Find an expression similar to (31.46) for the case in which plant (31.38) has 3 zeros and poles, 4 zeros and poles, and so on. What happens if there are more poles than zeros?

Chapter 32

Identification from frequency responses

He thought he saw an Albatross
That fluttered round the lamp:
He looked again, and found it was
A Penny-Postage-Stamp.
“You’d best be getting home,” he said:
“The nights are very damp!”

Lewis CARROLL (1832 — †1898), *Sylvie and Bruno* (1889), 12

This chapter concerns identification methods from frequency responses.

32.1 Finding frequency response data

A frequency response to identify can be found in different ways.

- Sinusoids of different frequencies are used as inputs. After the transient regime dies off, gain and phase are determined comparing the input and output sinusoids.
- The Fourier transform of an impulse response, obtained by deconvolution, is numerically calculated, as mentioned in Section 31.4.
- A controller design method provides a particular frequency response for a controller, desirable for the plant to be controller. (See e.g. Exercise 2.)
- A frequency response is established using one of the methods studied below in Part VIII.
- Instead of using sinusoids as inputs, a **chirp** is used.

A chirp is a sinusoid with a time-varying frequency. An **up-chirp** begins with a low frequency ω_0 , which increases with time until frequency ω_f , reached at final time t_f . A **down-chirp** begins with a high frequency ω_0 , which decreases with time until frequency ω_f , reached at final time t_f . More formally:

Definition 32.1. A chirp is a signal $u(t)$, defined from $t = 0$ to $t = t_f$, given by

$$u(t) = A \sin(\omega(t)t) \quad \square \quad (32.1)$$

Example 32.1. In the case of an up-chirp with a frequency increasing linearly with time, (32.1) becomes

$$\omega(t) = \omega_0 + \frac{\omega_f - \omega_0}{t_f} t, \quad \omega_0 < \omega_f \quad \square \quad (32.2)$$

Example 32.2. In the case of an down-chirp with a frequency decreasing exponentially with time, (32.1) becomes

$$\omega(t) = \omega_0 \left(\frac{\omega_f}{\omega_0} \right)^{\frac{t}{t_f}}, \quad \omega_0 > \omega_f \quad \square \quad (32.3)$$

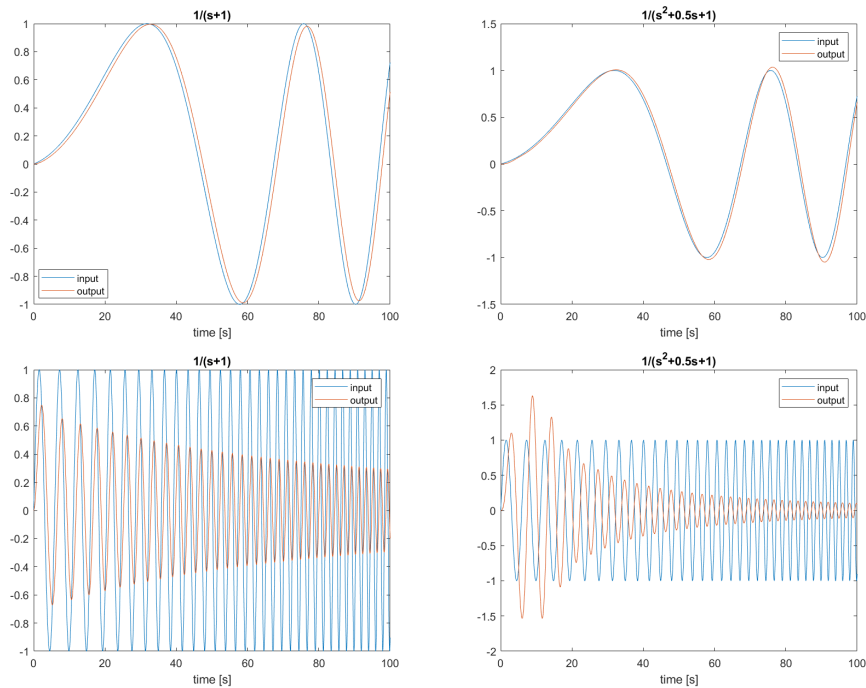


Figure 32.1: Left: responses to chirps of a first order plant. Right: responses to chirps of a second order plant. Top: responses of first and second order plants to a chirp in the $[0.01, 1]$ rad/s frequency range. Bottom: responses of first and second order plants to a chirp in the $[1, 100]$ rad/s frequency range.

Using a chirp instead of several sinusoids with different frequencies to obtain the frequency response of a plant has the major advantage of being faster. The major drawback is that, of course, a steady-state is never reached. Consequently, measurements of both gain and amplitude cannot but be approximations. However, if the rate of frequency increase is low enough, this approximation may be acceptable.

Replacing sinusoids with a chirp

Example 32.3. Figure 32.1 shows the response of

$$G_1(s) = \frac{1}{s+1} \quad (32.4)$$

$$G_2(s) = \frac{1}{s^2+s+1} \quad (32.5)$$

to two exponential up-chirps: one in frequency range $[0.01, 1]$ rad/s, another in frequency range $[1, 100]$ rad/s. Notice the resonance peak of the underdamped ($\xi = 0.25$) second order plant. Also notice that steady state responses are never reached. For instance:

- at $\omega = 100$ rad/s, $|G(100j)| = 0.01$, but the amplitude of the output is still at about 0.3, when the amplitude of the input is 1;
- the maximum gain of $G_2(s)$ is $0.5\sqrt{1-0.25^2} = 0.48$, according to (11.80), i.e. 48% above the low frequency range, but the maximum amplitude of the chirp response reaches 1.63.

Clearly this chirp would have had to last longer to explore the frequency response in this way. \square

Remark 32.1. It is difficult to tell, before identification, how fast or slow the transient regime of a plant may be, and consequently how much time a chirp ought to take to sweep any particular range of frequencies. Still, responses such as those in Figure 32.1 can even so be used in identification, with the advantage of having explored the system's response at many frequencies, but using instead the techniques we will study in Part VIII. \square

32.2 Identification from the Bode diagram

We studied how a Bode diagram can be built from poles, zeros and gain in Section 11.4. A reverse reasoning can lead us from the Bode diagram to the poles, zeros and gain of the model. In particular:

- A constant gain and a phase of 0° or $\pm 180^\circ$ at low frequency mean that there are neither poles nor zeros at the origin.
- In that case, if the phase is 0° , the gain is positive; if it is $\pm 180^\circ$, the gain is negative.
- A negative slope of $-20n$ dB/decade and a phase of $-90^\circ \times n$ at low frequency mean that there are n poles at the origin.
- A positive slope of $20n$ dB/decade and a phase of $90^\circ \times n$ at low frequency mean that there are n zeros at the origin.
- Remember that the phase is periodic and so a phase of θ is the same as a phase of $\theta \pm 360^\circ k$, $k = 1, 2, \dots$
- A negative slope of $-20n$ dB/decade and a phase of $-90^\circ \times n$ at high frequency mean that there are n poles more than zeros.
- A positive slope of $20n$ dB/decade and a phase of $90^\circ \times n$ at high frequency mean that there are n zeros more than poles.
- A decrease of the slope of the gain of $-20n$ dB/decade and a decrease of the phase of $-90^\circ \times n$ at some frequency mean that there are n poles at that frequency.
- An increase of the slope of the gain of $20n$ dB/decade and an increase of the phase of $90^\circ \times n$ at some frequency mean that there are n zeros at that frequency.
- Pairs of complex conjugate poles with damping coefficient $\xi < \frac{\sqrt{2}}{2}$ can be detected from the resonance peak, and (11.80) can be used to find ξ ; then (11.79) can be used to find ω_n .
- Gain slopes can be estimated by eye, or analytically determined with a linear regression. In the later case, care must be taken to choose wisely the frequencies to include.

Review Section 11.4 for further insights. See Exercises 10 to 12 of Chapter 11.

Notice that this method can be applied if all that is known about the frequency response is the gain. This may happen because the phase was not registered, or is more affected by noise than the gain. However, it will be impossible to determine if poles are stable or unstable, and if zeros are minimum phase or non-minimum phase. Only with the phase can that be known. Unless there is some reason to suppose otherwise, stable poles and minimum phase zeros are presumed.

If only the gain is known

Something similar happens if all that is known about the frequency response is the phase (a more rare occurrence than knowing only the gain).

If only the phase is known

32.3 Levy's method

Levy's method fits a transfer function to a frequency response using least squares.

Theorem 32.1. Let a plant's frequency response $G(j\omega_p)$ be known at f frequencies, i.e. $p = 1, 2, \dots, f$. The model

$$\hat{G}(s) = \frac{N(s)}{D(s)} = \frac{\sum_{k=0}^m b_k s^k}{1 + \sum_{k=1}^n a_k s^k} \quad (32.6)$$

with parameters to be determined

$$\mathbf{b} = [b_0 \quad \cdots \quad b_m]^T \quad (32.7)$$

$$\mathbf{a} = [a_1 \quad \cdots \quad a_n]^T \quad (32.8)$$

What *Levy's method* minimises which quadratic error

$$\begin{aligned} \varepsilon &= \sum_{p=1}^f \left| \left(G(j\omega_p) - \hat{G}(j\omega_p) \right) D(j\omega_p) \right|^2 \\ &= \sum_{p=1}^f \left| G(j\omega_p) D(j\omega_p) - N(j\omega_p) \right|^2 \end{aligned} \quad (32.9)$$

is found solving

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{e} \\ \mathbf{g} \end{bmatrix} \quad (32.10)$$

$$\mathbf{A} = \sum_{p=1}^f \mathbf{A}_p \quad (32.11)$$

$$\mathbf{B} = \sum_{p=1}^f \mathbf{B}_p \quad (32.12)$$

$$\mathbf{C} = \sum_{p=1}^f \mathbf{C}_p \quad (32.13)$$

$$\mathbf{D} = \sum_{p=1}^f \mathbf{D}_p \quad (32.14)$$

$$\mathbf{e} = \sum_{p=1}^f \mathbf{e}_p \quad (32.15)$$

$$\mathbf{g} = \sum_{p=1}^f \mathbf{g}_p \quad (32.16)$$

where the elements in line l and column c of matrixes \mathbf{A}_p , \mathbf{B}_p , \mathbf{C}_p and \mathbf{D}_p are given by

$$\begin{aligned} \mathbf{A}_{p;l,c} &= -\Re [(j\omega_p)^l] \Re [(j\omega_p)^c] - \Im [(j\omega_p)^l] \Im [(j\omega_p)^c], \\ & \quad l = 0 \dots m \wedge c = 0 \dots m \end{aligned} \quad (32.17)$$

$$\begin{aligned} \mathbf{B}_{p;l,c} &= \Re [(j\omega_p)^l] \Re [(j\omega_p)^c] \Re [G(j\omega_p)] + \Im [(j\omega_p)^l] \Re [(j\omega_p)^c] \Im [G(j\omega_p)] \\ & \quad - \Re [(j\omega_p)^l] \Im [(j\omega_p)^c] \Re [G(j\omega_p)] + \Im [(j\omega_p)^l] \Im [(j\omega_p)^c] \Re [G(j\omega_p)], \\ & \quad l = 0 \dots m \wedge c = 1 \dots n \end{aligned} \quad (32.18)$$

$$\begin{aligned} \mathbf{C}_{p;l,c} &= -\Re [(j\omega_p)^l] \Re [(j\omega_p)^c] \Re [G(j\omega_p)] + \Im [(j\omega_p)^l] \Re [(j\omega_p)^c] \Im [G(j\omega_p)] \\ & \quad - \Re [(j\omega_p)^l] \Im [(j\omega_p)^c] \Re [G(j\omega_p)] - \Im [(j\omega_p)^l] \Im [(j\omega_p)^c] \Re [G(j\omega_p)], \\ & \quad l = 1 \dots n \wedge c = 0 \dots m \end{aligned} \quad (32.19)$$

$$\begin{aligned} \mathbf{D}_{p;l,c} &= \left(\Re [G(j\omega_p)]^2 + \Im [G(j\omega_p)]^2 \right) \left(\Re [(j\omega_p)^l] \Re [(j\omega_p)^c] + \Im [(j\omega_p)^l] \Im [(j\omega_p)^c] \right), \\ & \quad l = 1 \dots n \wedge c = 1 \dots n \end{aligned} \quad (32.20)$$

and the elements of vectors \mathbf{e}_p and \mathbf{g}_p are given by

$$\begin{aligned} \mathbf{e}_{p;l,1} &= -\Re [(j\omega_p)^l] \Re [G(j\omega_p)] - \Im [(j\omega_p)^l] \Im [G(j\omega_p)], \\ & \quad l = 0 \dots m \end{aligned} \quad (32.21)$$

$$\begin{aligned} \mathbf{g}_{p;l,1} &= -\Re [(j\omega_p)^l] \left(\Re [G(j\omega_p)]^2 + \Im [G(j\omega_p)]^2 \right), \\ & \quad l = 1 \dots n \end{aligned} \quad (32.22)$$

Proof. Error (32.9) is given by

$$\begin{aligned}
\varepsilon &= |GD - N|^2 \\
&= \left| \left(\Re[G] + j\Im[G] \right) \left(\Re[D] + j\Im[D] \right) - \left(\Re[N] + j\Im[N] \right) \right|^2 \\
&= \left| \left(\Re[G]\Re[D] - \Im[G]\Im[D] - \Re[N] \right) + j \left(\Re[G]\Im[D] + \Im[G]\Re[D] - \Im[N] \right) \right|^2 \\
&= \left(\Re[G]\Re[D] - \Im[G]\Im[D] - \Re[N] \right)^2 + \left(\Re[G]\Im[D] + \Im[G]\Re[D] - \Im[N] \right)^2
\end{aligned} \tag{32.23}$$

To minimise ε , we want its derivatives in order to coefficients a_k and b_k to be zero.

Let us suppose that there is only one frequency ω_p . First we find

$$\frac{d\Re[G]}{da_i} = \frac{d\Re[G]}{db_i} = \frac{d\Im[G]}{da_i} = \frac{d\Im[G]}{db_i} = 0 \tag{32.24}$$

$$\frac{d\Re[N]}{da_i} = 0 \tag{32.25}$$

$$\frac{d\Re[N]}{db_i} = \frac{d}{db_i} \sum_{k=0}^m b_k \Re[(j\omega_p)^k] = \Re[(j\omega_p)^i] \tag{32.26}$$

$$\frac{d\Im[N]}{da_i} = 0 \tag{32.27}$$

$$\frac{d\Im[N]}{db_i} = \frac{d}{db_i} \sum_{k=0}^m b_k \Im[(j\omega_p)^k] = \Im[(j\omega_p)^i] \tag{32.28}$$

$$\frac{d\Re[D]}{da_i} = \frac{d}{da_i} \left\{ 1 + \sum_{k=1}^n a_k \Re[(j\omega_p)^k] \right\} = \Re[(j\omega_p)^i] \tag{32.29}$$

$$\frac{d\Re[D]}{db_i} = 0 \tag{32.30}$$

$$\frac{d\Im[D]}{da_i} = \frac{d}{da_i} \left\{ 1 + \sum_{k=1}^n a_k \Im[(j\omega_p)^k] \right\} = \Im[(j\omega_p)^i] \tag{32.31}$$

$$\frac{d\Im[D]}{db_i} = 0 \tag{32.32}$$

(Notice that j^k can only assume four values: ± 1 and $\pm j$. In the first case, there will be a real part, and the imaginary part is zero; in the later case, there will be an imaginary part, and the real part is zero.) Consequently,

$$\begin{aligned}
\frac{\partial \varepsilon}{\partial a_i} &= 2 \left(\Re[G]\Re[D] - \Im[G]\Im[D] - \Re[N] \right) \left(\Re[G]\Re[(j\omega_p)^i] - \Im[G]\Im[(j\omega_p)^i] \right) \\
&\quad + 2 \left(\Re[G]\Im[D] + \Im[G]\Re[D] - \Im[N] \right) \left(\Re[G]\Im[(j\omega_p)^i] + \Im[G]\Re[(j\omega_p)^i] \right)
\end{aligned} \tag{32.33}$$

$$\begin{aligned}
\frac{\partial \varepsilon}{\partial b_i} &= 2 \left(\Re[G]\Re[D] - \Im[G]\Im[D] - \Re[N] \right) \left(-\Re[(j\omega_p)^i] \right) \\
&\quad + 2 \left(\Re[G]\Im[D] + \Im[G]\Re[D] - \Im[N] \right) \left(-\Im[(j\omega_p)^i] \right)
\end{aligned} \tag{32.34}$$

Equating to zero,

$$\begin{aligned}
\frac{\partial \varepsilon}{\partial a_i} = 0 &\Leftrightarrow \Re[D] \left(\Im[G]^2 + \Re[G]^2 \right) \Re[(j\omega_p)^i] + \\
&\quad \Im[D] \left(\Im[G]^2 + \Re[G]^2 \right) \Im[(j\omega_p)^i] + \\
&\quad \Re[N] \left(\Im[G]\Im[(j\omega)^i] - \Re[G]\Re[(j\omega_p)^i] \right) + \\
&\quad \Im[N] \left(-\Im[G]\Re[(j\omega_p)^i] - \Re[G]\Im[(j\omega_p)^i] \right) = 0
\end{aligned} \tag{32.35}$$

$$\begin{aligned}
\frac{\partial \varepsilon}{\partial b_i} = 0 &\Leftrightarrow \left(\Re[G]\Re[D] - \Im[G]\Im[D] - \Re[N] \right) \Re[(j\omega_p)^i] + \\
&\quad \left(\Re[G]\Im[D] + \Im[G]\Re[D] - \Im[N] \right) \Im[(j\omega_p)^i] = 0
\end{aligned} \tag{32.36}$$

The $m + 1$ equations given by (32.35) and the n equations given by (32.36) form linear system (32.10), with matrixes and vectors defined by (32.17)–(32.22).

Considering f frequencies instead of only one, we arrive at (32.11)–(32.16). \square

Why the error is defined as it was

Remark 32.2. Notice that error was defined as in (32.9) so that the minimisation problem should be linear on coefficients \mathbf{b} and \mathbf{a} . That is why the denominator $D(j\omega_p)$ is used as a sort of weighting function. Otherwise, the problem would not have as solution a system of linear equations such as (32.10). \square

High frequencies are given more weight

Remark 32.3. Using $D(j\omega_p)$ is used as a weighting function in error ε gives more weight to high frequencies, since $D(j\omega_p)$ is a polynomial and thus its absolute value must increase with frequency ω . If Levy's method is applied, and the performance of the resulting model is poor at low frequencies, this effect can be counteracted, weighting the contributions of different frequencies in (32.11)–(32.16) with weights that increase the influence of low frequencies. \square

Use weights to improve low frequencies results

Alternative formulation of Levy's method

Remark 32.4. Instead of summing matrixes \mathbf{A}_p , \mathbf{B}_p , \mathbf{C}_p and \mathbf{D}_p , and vectors \mathbf{e}_p and \mathbf{g}_p , they can be stacked. In that case, instead of (32.10)–(32.16), the coefficients in \mathbf{b} and \mathbf{a} will be found solving

$$\underbrace{\begin{bmatrix} \mathbf{A}_1 & \mathbf{B}_1 \\ \mathbf{C}_1 & \mathbf{D}_1 \\ \mathbf{A}_2 & \mathbf{B}_2 \\ \mathbf{C}_2 & \mathbf{D}_2 \\ \vdots & \vdots \\ \mathbf{A}_f & \mathbf{B}_f \\ \mathbf{C}_f & \mathbf{D}_f \end{bmatrix}}_{\mathbf{H}} \begin{bmatrix} \mathbf{b} \\ \mathbf{a} \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{e}_1 \\ \mathbf{g}_1 \\ \mathbf{e}_2 \\ \mathbf{g}_2 \\ \vdots \\ \mathbf{e}_f \\ \mathbf{g}_f \end{bmatrix}}_{\mathbf{v}} \quad (32.37)$$

This overdetermined system will have a solution in least squares sense:

$$\begin{bmatrix} \mathbf{b} \\ \mathbf{a} \end{bmatrix}^T = \mathbf{H}^+ \mathbf{v} \quad \square \quad (32.38)$$

What to do if there are poles at the origin

Remark 32.5. Model (32.6) is of type 0. If the frequency data corresponds to a model with poles at the origin, the effect of such poles must be removed from the data. Once the model is identified, the poles at the origin are then added.

Zeros at the origin cause no problem. \square

Example 32.4. The frequency response $G(j\omega)$ of

$$G(s) = \frac{5s + 10}{s(s^2 + 0.2s + 1)} \quad (32.39)$$

is sampled in 31 logarithmically spaced frequencies in $[0.1, 100]$ rad/s. The numbers of zeros and poles $m = 1$ and $n = 3$ can be easily from the Bode diagram. Levy's method identifies

$$\hat{G}(s) = \frac{0.02341s - 6.569}{0.002116s^3 - 1.317s^2 - 0.1569s - 1} \quad (32.40)$$

because a model of type 0 is presumed. This is shown in Figure 32.2.

Since there is obviously a pole at the origin, we should find the frequency response of this pole and subtract it from $G(j\omega)$:

$$G_{\text{correction}}(s) = \frac{1}{s} \quad (32.41)$$

$$20 \log_{10} |G^*(j\omega)| = 20 \log_{10} |G(j\omega)| - 20 \log_{10} |G_{\text{correction}}(j\omega)| \quad (32.42)$$

$$\angle G^*(j\omega) = \angle G(j\omega) - \angle G_{\text{correction}}(j\omega) \quad (32.43)$$

From frequency response $G^*(j\omega)$, Levy's method with $m = 1$ and $n = 2$ identifies

$$\hat{G}(s) = \frac{5s + 10}{s^2 + 0.2s + 1} \quad (32.44)$$

from which (32.39) is immediately recovered, as shown in Figure 32.3. \square

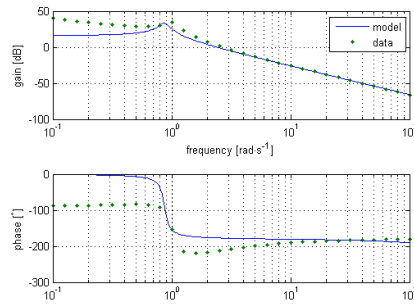


Figure 32.2: Frequency response of $G(s)$, given by (32.39), and frequency response of identified model $\hat{G}(s)$, given by (32.40).

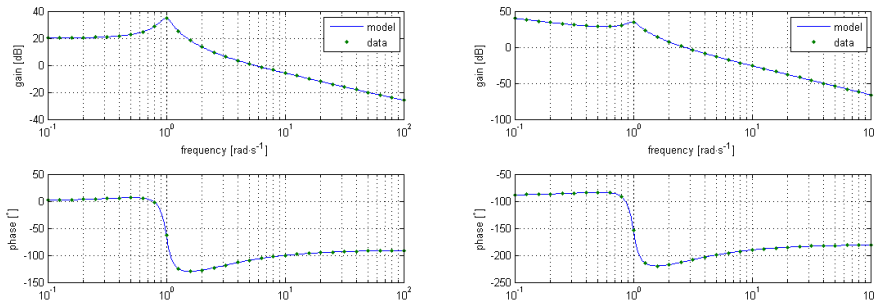


Figure 32.3: Left: frequency response $G^*(j\omega)$, given by (32.42)–(32.43), and frequency response of identified model $\hat{G}(s)$, given by (32.40). Right: frequency response of $G(s)$, given by (32.39), and frequency response of the identified model, which is the same.

Remark 32.6. If, in the example above, the model should have k poles at the origin, then

$$G_{\text{correction}}(s) = \frac{1}{s^k} \quad \square \tag{32.45}$$

Remark 32.7. Levy's method requires knowing in advance the number of zeros and poles needed in the model. The inspection of the Bode diagram should be used beforehand for this purpose. \square

32.4 Matsuda's method

Matsuda's method identifies a model from the gain of the frequency behaviour of a plant. Stable poles and minimum phase zeros are assumed. This method relies on continued fractions.

Definition 32.2. A **continued fraction** is an entity, defined from two sequences a_k and b_k , with the form *Continued fractions*

$$a_0 + \frac{b_1}{a_1 + \frac{b_2}{a_2 + \frac{b_3}{a_3 + \frac{b_4}{a_4 + \dots}}}} \quad \square \tag{32.46}$$

Instead of writing continued fractions as in (32.46), they are usually notated, for the benefit of clarity, using either

- the Abramowitz notation: *Abramowitz notation*

$$a_0 + \frac{b_1}{a_1 +} \frac{b_2}{a_2 +} \frac{b_3}{a_3 +} \frac{b_4}{a_4 +} \dots \tag{32.47}$$

- the Pringsheim notation: *Pringsheim notation*

$$\left[a_0; \frac{b_1}{a_1}, \frac{b_2}{a_2}, \frac{b_3}{a_3}, \frac{b_4}{a_4}, \dots \right] = \left[a_0, \frac{b_k}{a_k} \right]_{k=1}^{+\infty} \tag{32.48}$$

Remark 32.8. Just as a series $\sum_{k=1}^{+\infty} a_k$ is built from a sequence a_k , so a continued fraction is built from two sequences a_k and b_k , and just as series may or may not converge, continued fractions may converge or not. \square

Continued fraction expansion of real numbers

Every $x_0 \in \mathbb{R} \setminus \mathbb{Z}$ can be written as

$$x_0 = [x_0] + \frac{1}{\frac{1}{x_0 - [x_0]}} = [x_0] + \underbrace{\frac{1}{\frac{1}{x_0 - [x_0]}}}_{x_1} \tag{32.49}$$

and applying this repeatedly we find that

$$x_0 = [x_0] + \frac{1}{[x_1] + \frac{1}{[x_2] + \frac{1}{[x_3] + \dots}}} = \left[[x_0]; \frac{1}{[x_k]} \right]_{k=1}^{+\infty} \tag{32.50}$$

$$x_k = \frac{1}{x_{k-1} - [x_{k-1}]}, \quad k \in \mathbb{N} \tag{32.51}$$

If $x_0 \in \mathbb{Q}$, the continued fraction expansion above will terminate (that is to say, sooner or later one of the x_k will be integer, and thus all the following coefficients x_n , $n > k$ will be zero). If x_0 is irrational, all the x_k will be different from zero. Whatever the case, the continued fraction can be truncated after some terms, providing a rational approximation of x_0 which is the best possible for the order of magnitude of the denominator; that is to say, no other fraction closer to x_0 can be found with a denominator of the same order of magnitude.

Continued fractions as approximations

Evaluating a continued fraction

Terminating continued fractions can be evaluated beginning with its last (innermost) fraction, but it is more expedient to use the following result whether the continued fraction terminates or not.

Theorem 32.2. Let a_k and b_k be two sequences defining a continued fraction given by (32.46)–(32.48), and let

$$P_{-1} = 1 \tag{32.52}$$

$$P_0 = a_0 \tag{32.53}$$

$$P_k = a_k P_{k-1} + b_k P_{k-2}, \quad k \in \mathbb{N} \tag{32.54}$$

$$Q_{-1} = 0 \tag{32.55}$$

$$Q_0 = 1 \tag{32.56}$$

$$Q_k = a_k Q_{k-1} + b_k Q_{k-2}, \quad k \in \mathbb{N} \tag{32.57}$$

Then $R_k = \frac{P_k}{Q_k}$ is the value of the continued fraction, truncated after k terms.

Proof. This is proved by mathematical induction. For $k = 1$,

$$R_1 = \frac{P_1}{Q_1} = \frac{a_1 a_0 + b_1}{a_1 + 0} = a_0 + \frac{b_1}{a_1} \tag{32.58}$$

To prove the inductive step, notice that adding a further pair of coefficients means that the last denominator a_k will be added a new fraction $\frac{b_{k+1}}{a_{k+1}}$. So, if in

$$R_k = \frac{P_k}{Q_k} = \frac{a_k P_{k-1} + b_k P_{k-2}}{a_k Q_{k-1} + b_k Q_{k-2}} \text{ we replace } a_k \text{ with } a_k + \frac{b_{k+1}}{a_{k+1}},$$

$$\begin{aligned} R_{k+1} &= \frac{\left(a_k + \frac{b_{k+1}}{a_{k+1}}\right) P_{k-1} + b_k P_{k-2}}{\left(a_k + \frac{b_{k+1}}{a_{k+1}}\right) Q_{k-1} + b_k Q_{k-2}} \\ &= \frac{a_{k+1} \overbrace{\left(a_k P_{k-1} + b_k P_{k-2}\right)}^{P_k} + b_{k+1} P_{k-1}}{a_k \underbrace{\left(a_k Q_{k-1} + b_k Q_{k-2}\right)}_{Q_k} + b_{k+1} Q_{k-1}} = \frac{P_{k+1}}{Q_{k+1}} \square \end{aligned} \tag{32.59}$$

Remark 32.9. By computing the P_k and Q_k in this way, additional terms can be added to the truncated continued fraction without having to begin the computation anew. \square

Just as series can be used to define functions by letting the terms of the summed sequence depend on t , so can continued fractions.

Theorem 32.3. If function $f_0(t)$ is known at $t_0, t_1, t_2, \dots, t_p$, it can be interpolated by continued fraction

$$f_0(t) \approx f_0(t_0) + \frac{t-t_0}{f_1(t_1)+} \frac{t-t_1}{f_2(t_2)+} \frac{t-t_2}{f_3(t_3)+} \dots = \left[f_0(t_0); \frac{t-t_k}{f_{k+1}(t_{k+1})+} \right]_{k=0}^p \quad (32.60)$$

$$f_{k+1}(t) = \frac{t-t_k}{f_k(t) - f_k(t_k)}, \quad k \in \mathbb{N}_0 \quad (32.61)$$

Proof. Let $p \geq k+1$. The continued fraction (32.60), evaluated at t_{k+1} , becomes

$$\begin{aligned} & f_0(t_0) + \frac{t_{k+1}-t_0}{f_1(t_1)+} \frac{t_{k+1}-t_1}{f_2(t_2)+} \dots \frac{t_{k+1}-t_{k-2}}{f_{k-1}(t_{k-1})+} \frac{t_{k+1}-t_{k-1}}{f_k(t_k)+} \frac{t_{k+1}-t_k}{f_{k+1}(t_{k+1})+0} \\ &= f_0(t_0) + \frac{t_{k+1}-t_0}{f_1(t_1)+} \frac{t_{k+1}-t_1}{f_2(t_2)+} \dots \frac{t_{k+1}-t_{k-2}}{f_{k-1}(t_{k-1})+} \frac{t_{k+1}-t_{k-1}}{f_k(t_k)+} \frac{t_{k+1}-t_k}{\frac{t_{k+1}-t_k}{f_k(t_{k+1})-f_k(t_k)}} \\ &= f_0(t_0) + \frac{t_{k+1}-t_0}{f_1(t_1)+} \frac{t_{k+1}-t_1}{f_2(t_2)+} \dots \frac{t_{k+1}-t_{k-2}}{f_{k-1}(t_{k-1})+} \frac{t_{k+1}-t_{k-1}}{f_k(t_{k+1})} \\ &= f_0(t_0) + \frac{t_{k+1}-t_0}{f_1(t_1)+} \frac{t_{k+1}-t_1}{f_2(t_2)+} \dots \frac{t_{k+1}-t_{k-2}}{f_{k-1}(t_{k-1})+} \frac{t_{k+1}-t_{k-1}}{\frac{t_{k+1}-t_k}{f_{k-1}(t_{k+1})-f_{k-1}(t_{k-1})}} \\ &= f_0(t_0) + \frac{t_{k+1}-t_0}{f_1(t_1)+} \frac{t_{k+1}-t_1}{f_2(t_2)+} \dots \frac{t_{k+1}-t_{k-2}}{f_{k-1}(t_{k+1})} \end{aligned} \quad (32.62)$$

And so on, until

$$\begin{aligned} & f_0(t_0) + \frac{t_{k+1}-t_0}{f_1(t_1)+} \frac{t_{k+1}-t_1}{f_2(t_{k+1})} \\ &= f_0(t_0) + \frac{t_{k+1}-t_0}{f_1(t_1)+} \frac{t_{k+1}-t_1}{\frac{t_{k+1}-t_1}{f_1(t_{k+1})-f_1(t_1)}} \\ &= f_0(t_0) + \frac{t_{k+1}-t_0}{f_1(t_{k+1})} \\ &= f_0(t_0) + \frac{t_{k+1}-t_0}{\frac{t_{k+1}-t_0}{f_0(t_{k+1})-f_0(t_0)}} \\ &= f_0(t_{k+1}) \square \end{aligned} \quad (32.63)$$

Remark 32.10. Notice that points t_k do not need to be ordered. The same points, ordered in different manners, lead to different interpolating continued fractions. \square

(32.60) interpolates a real function of a real variable; Matsuda's identification method is a generalisation for complex variables. Given a frequency behaviour $G(j\omega)$, known at frequencies $\omega_0, \omega_1, \dots, \omega_N$, which do not need to be ordered, *Matsuda's method*

$$G(s) \approx d_0(\omega_0) + \frac{s-\omega_0}{d_1(\omega_1)+} \frac{s-\omega_1}{d_2(\omega_2)+} \frac{s-\omega_2}{d_3(\omega_3)+} \dots = \left[d_0(\omega_0); \frac{s-\omega_{k-1}}{d_k(\omega_k)+} \right]_{k=1}^N \quad (32.64)$$

$$d_0(\omega) = |G(j\omega)| \quad (32.65)$$

$$d_k(\omega) = \frac{\omega - \omega_{k-1}}{d_{k-1}(\omega) - d_{k-1}(\omega_{k-1})}, \quad k = 1, 2, \dots, N \quad (32.66)$$

Remark 32.11. Models obtained with (32.64) have $\lceil \frac{N}{2} \rceil$ zeros and $\lfloor \frac{N}{2} \rfloor$ poles, which means that they will be causal only if N is even (and thus the number of frequencies, which is $N+1$, is odd). \square

32.5 Oustaloup's method

Oustaloup's method identifies a model from the phase of the frequency behaviour of a plant. Stable poles and minimum phase zeros are assumed. Proving the method is exceedingly tedious; we will only state the result.

Suppose we want to find a transfer function

$$\hat{G}(s) = G_0 \frac{\prod_{k=1}^m 1 + \frac{s}{b_k}}{\prod_{k=1}^n 1 + \frac{s}{a_k}}, \quad G_0 > 0 \quad (32.67)$$

with the phase behaviour ϕ_i at frequencies ω_i , $i = 1 \dots f$; that is to say, we want that

$$\arg \hat{G}(j\omega_i) = \arg \frac{\prod_{k=1}^m 1 + \frac{j\omega_i}{b_k}}{\prod_{k=1}^n 1 + \frac{j\omega_i}{a_k}} = \phi_i, \quad i = 1 \dots f \quad (32.68)$$

Let \mathbf{C} be a $f \times (m+n+1)$ matrix given by

$$\mathbf{C} = \begin{bmatrix} \tan \phi_1 & 1 & -\tan \phi_1 & -1 & \tan \phi_1 & 1 & -\tan \phi_1 & \dots \\ \tan \phi_2 & \frac{\omega_2}{\omega_1} & -\tan \phi_2 \left(\frac{\omega_2}{\omega_1}\right)^2 & -\left(\frac{\omega_2}{\omega_1}\right)^3 & \tan \phi_2 \left(\frac{\omega_2}{\omega_1}\right)^4 & \left(\frac{\omega_2}{\omega_1}\right)^5 & -\tan \phi_2 \left(\frac{\omega_2}{\omega_1}\right)^6 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \dots \\ \tan \phi_f & \frac{\omega_f}{\omega_1} & -\tan \phi_f \left(\frac{\omega_f}{\omega_1}\right)^2 & -\left(\frac{\omega_f}{\omega_1}\right)^3 & \tan \phi_f \left(\frac{\omega_f}{\omega_1}\right)^4 & \left(\frac{\omega_f}{\omega_1}\right)^5 & -\tan \phi_f \left(\frac{\omega_f}{\omega_1}\right)^6 & \dots \end{bmatrix} \quad (32.69)$$

and let \mathbf{s} be a $(m+n+1) \times 1$ vector such that $\mathbf{C}\mathbf{s} = \mathbf{0}$. Because in practice there will be a solution only if there is no noise and the model structure is correct, we make

$$\mathbf{s} = \arg \min_{\mathbf{s}} \|\mathbf{C}\mathbf{s}\|^2 \quad (32.70)$$

instead. We then find the $m+n$ roots y_k , $k = 1 \dots m+n$ of a polynomial built with the elements of \mathbf{s} as coefficients:

$$\sum_{k=0}^{m+n} (-1)^k s_k y^{m+n-k} = 0 \quad (32.71)$$

Then for each $k = 1, \dots, m+n$ we make

$$a_k = \frac{\omega_1}{y_k}, \quad \text{if } y_k > 0 \quad (32.72)$$

$$b_k = -\frac{\omega_1}{y_k}, \quad \text{if } y_k < 0 \quad (32.73)$$

and thus obtain the $m+n$ poles and zeros of model (32.67). Notice that, while the total number of poles and zeros is in fact $m+n$, it is possible that the number of poles and the number of zeros are not n and m .

Glossary

Every concept that can ever be needed, will be expressed by exactly one word, with its meaning rigidly defined and all its subsidiary meanings rubbed out and forgotten. Already, in the Eleventh Edition, we're not far from that point. But the process will still be continuing long after you and I are dead. Every year fewer and fewer words, and the range of consciousness always a little smaller.

George ORWELL (1903 — †1950), *Nineteen Eighty-Four* (1949), I 5

chirp *chirp*

continued fraction fração contínua

down-chirp *chirp* descendente

up-chirp *chirp* ascendente

Exercises

1. Find a model for the frequency response in Table 32.1 using every method in this chapter.
2. A controller is wanted for plant

$$G(s) = \frac{s+1}{s^3 + 2.2s^2 + 100.4s + 20} \quad (32.74)$$

You want a constant phase margin of 45° in $[1, 10]$ rad/s.

Table 32.1: Frequency response of Exercise 1.

f [Hz]	gain [dB]	phase [°]	f [Hz]	gain [dB]	phase [°]
0.1	7.1	-6.35	1.0	3.7	-48.1
0.2	7.0	-12.6	1.2	2.7	-53.3
0.3	6.7	-18.5	1.5	1.4	-59.1
0.4	6.4	-24.1	2.0	-0.6	-65.9
0.5	6.0	-29.2	2.5	-2.3	-70.3
0.6	5.6	-33.8	3.0	-3.7	-73.4
0.7	5.1	-38.0	4.0	-6.0	-77.4
0.8	4.6	-41.8	5.0	-7.9	-79.8
0.9	4.1	-45.1	7.0	-10.8	-82.7
			10.0	-13.8	-84.9

- Find the frequency response of $G(s)$ in a suitable frequency range.
- Find the desirable phase of the frequency response of the controller. Sample it at suitable frequency values.
- Use Oustaloup's identification method to design a controller. The phase margin will never be really constant in the desired range, but may be within a narrow band centred on the desired value. Use a suitable performance index to verify the influence on results of the frequency values chosen to sample the desired controller behaviour.

3. Consider plant

$$G(s) = \frac{5s}{s^3 + s^2 + 0.2s + 1} \quad (32.75)$$

- Find its frequency response in some set of frequency values you find appropriate.
 - Use Levy's method to identify a transfer function from that frequency response.
 - Repeat this for different sets of frequency values, and for different model structures (with more, and less, poles and zeros than those of $G(s)$).
 - Use Matsuda's method and Oustaloup's method to identify models from the gain and the phase of a frequency response that had good results with Levy's method.
4. Figure 32.4 shows the Bode diagram of a LTI system. Find its transfer function.

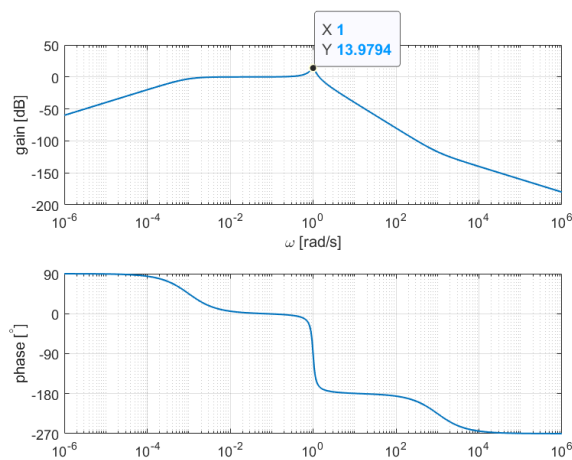


Figure 32.4: Bode diagram of Exercise 4.

5. Figure 32.5 shows the Bode diagram of a LTI system. Find its transfer function.

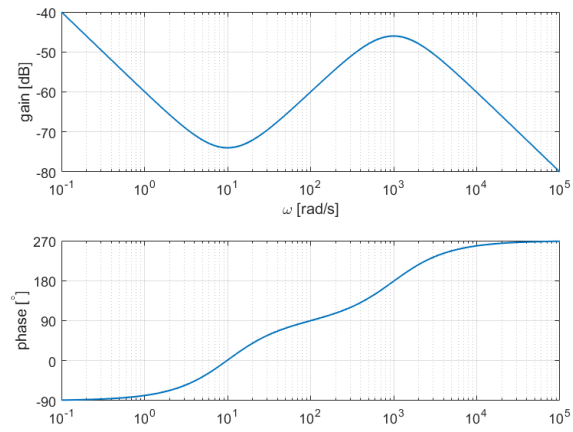


Figure 32.5: Bode diagram of Exercise 5.

6. Figure 32.6 shows steady state responses of a LTI system to sinusoidal inputs with amplitude 1.
- Find its transfer function, assuming that it has no non-minimum phase zeros, from the asymptotes of the Bode diagram.
 - Do the same, using Matsuda's method.
 - Could you use Levy's method? If so, how?

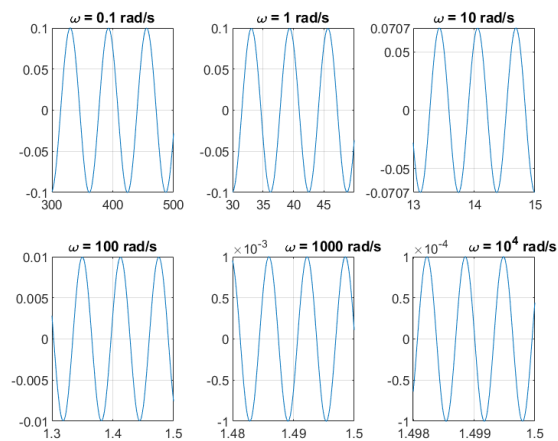


Figure 32.6: Time responses of Exercise 6.

Chapter 33

Identification of non-linearities

‘Somewhere to run,’ suggested Paul.

‘Why, God bless my soul, they’ve got the whole park! How did you manage yesterday for the heats?’

‘We judged the distance by eye.’

Evelyn WAUGH (19035 — †1966), *Decline and fall* (1928), 1 VIII

This chapter concerns the identification of non-linearities, introduced in Section 8.3 and already studied above in Chapter 28 as to their influence in control systems.

33.1 Identifying the presence of a non-linearity

There are three main ways of identifying the presence of a non-linearity.

The first is to verify if the amplitude of the output is proportional to the amplitude of the input. This can be seen, for instance, providing steps or sinusoids of different amplitudes as inputs. If the corresponding outputs are not proportional to the amplitude of the input, then there is a non-linearity.

The second is to check if the steady-state response of a system to a sinusoidal input is sinusoidal. Remember that Theorem 10.4 was established assuming that the system is linear. So, if a stable system has a sinusoidal input, and the output never becomes sinusoidal, there clearly is a non-linearity involved, even if it is not easy to see which; review Example 10.12. Remember that, because of this, non-linear systems have no frequency response; describing functions (introduced in Section 28.2) are only *approximations* of frequency responses.

The third is to study the behaviour of the system in closed loop, looking for

- responses that are unexpectedly stable or unstable, indicating an equivalent gain in the closed loop caused by a non-linearity (see Section 28.1);
- limit cycles (see Section 28.2). We will address this possibility further below in Section 33.3.

How to detect a non-linearity

Non-linear systems do not have outputs proportional to inputs

Non-linear systems do not have sinusoidal outputs for sinusoidal inputs

33.2 Identification from a time response

Recall the distinction introduced in Section 8.3 between **soft non-linearities** and **hard non-linearities**. As mentioned there, soft non-linearities can be approximated by a linearised model around some point of operation, typically the steady-state. If necessary, several different linear models, corresponding to different operation points, can be found.

Hard non-linearities can have no effect in time responses if the input is such that the output never reaches values where the non-linearity has its effect. This is obvious for a saturation, for example: if the outputs are always small enough, saturation values are never reached and can seem non-existent.

Hard non-linearities can be distinguished one from another looking carefully at how time responses behave.

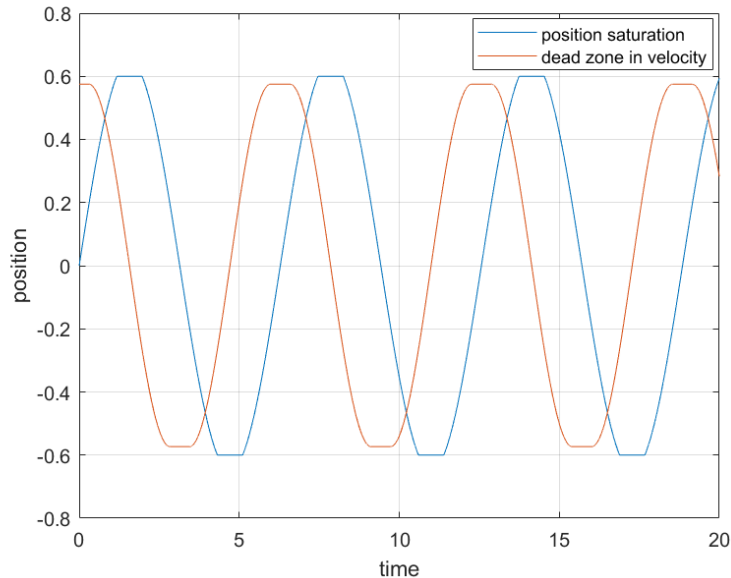


Figure 33.1: Effects of non-linearities. The variable is assumed to be a position. In one case, the position saturates, i.e. there are end stops. In the other, the velocity has a dead zone.

Example 33.1. Figure 33.1 shows that the effects of a saturation and of a dead zone can be similar.

Saturation vs. dead zone in time responses

Figure 33.2 shows experimental data of the output of a system (controlled in closed loop) for a sinusoidal input. While one might think that this plot shows the existence of output saturation at values corresponding to sensor measurements about ± 0.6 V, it is in fact a dead zone that causes the non-linear behaviour. Indeed, the value of the apparent saturation is not always the same over different periods of the sinusoid; even during the same period there are small fluctuations that are hardly a consequence of sensor noise. Even though this makes clear that the visible non-linearity is not a saturation, we could confirm this using sinusoids of different amplitudes as input: a saturation would always correspond to outputs limited at the same value, \square

33.3 Identification from a limit cycle

We saw in Sections 28.2 and 28.3 that a closed loop with a non-linearity will have a limit cycle if the curve of the frequency response of the linear part and the curve of $-\frac{1}{N}$ intersect, as expressed by condition (28.48):

$$G(j\omega) = -\frac{1}{N(U, \omega)} \quad (33.1)$$

The amplitude of the limit cycle will be U and its frequency ω . Additionally, the limit cycle should be stable to be observed.

If a limit cycle is observed, the describing function N of the non-linearity can be found from the measured values of U and ω . These non-linear calculations, however, are often affected by numerical problems in practice.

Calculations of the limit cycle, backwards

Example 33.2. Consider the situation of Examples 28.6 and 28.7. The plant has already been identified in open loop as

$$G(s) = \frac{1.4}{s(s+1)^2} \quad (33.2)$$

When controlling the plant in closed loop, a limit cycle is found with frequency 0.76 rad/s and amplitude 4.68. Since there are no marginally stable poles, this means that there is a non-linearity. If it can be ascertained that the non-linearity

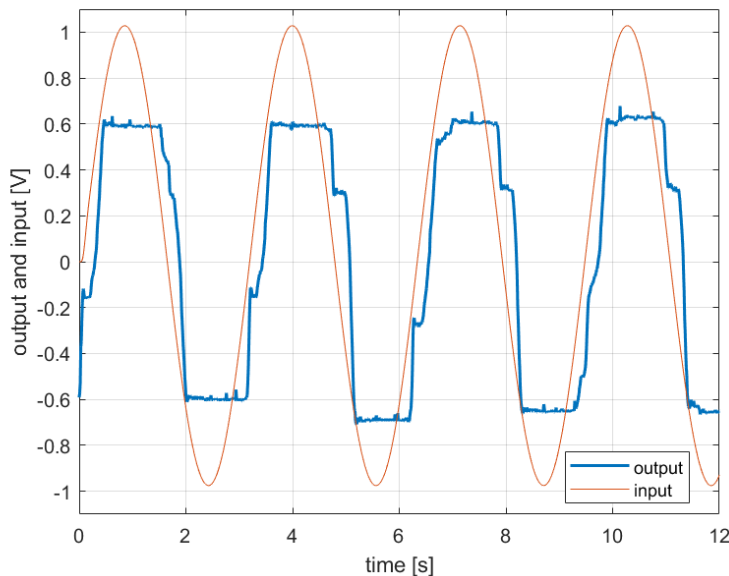


Figure 33.2: Plant with a dead zone. (Data collection: Mariana Preto Nunes and André Nakamura.)

is due to backlash, then

$$\frac{1.4}{j\omega(j\omega + 1)^2} = -\frac{1}{\frac{k}{2} \left(1 - \frac{2}{\pi} \left[\arcsin\left(\frac{2 - \frac{4.68}{A}}{\frac{4.68}{A}}\right) + \frac{2 - \frac{4.68}{A}}{\frac{4.68}{A}} \cos\left(\arcsin\left(\frac{2 - \frac{4.68}{A}}{\frac{4.68}{A}}\right)\right) \right] \right) - 4kA j \frac{(4.68 - A)}{\pi 4.68^2}} \quad (33.3)$$

Compare this with (28.50). There are two unknowns, both of them real numbers. The real and the imaginary parts of (33.3) provide a system of equations to find them. \square

33.4 Identification of a pure delay

In Chapter 24, we saw that a pure delay is a non-linear element when using continuous transfer functions, and in Chapter 25 we saw that it is a linear element when using discrete transfer functions. Whatever the case, its identification is done in the same way:

- In the case of a time response, the system should first be brought to steady state, and then a change in the input applied. A step is not necessary, as long as it is possible to measure the time to a change in the output. This should be done for several different types of input changes (steps with different amplitudes, ramps...), to make sure that the delay observed is always the same. If it is not, then the apparent delay is in reality due to some other non-linearity.
- In the case of a frequency response, since

$$\begin{aligned} \angle e^{-\theta j\omega} &= \angle [\cos(-\theta\omega) + j \sin(-\theta\omega)] \\ &= \angle [\cos(\theta\omega) - j \sin(\theta\omega)] \\ &= \arctan \frac{-\sin \theta\omega}{\cos \theta\omega} = -\theta\omega \end{aligned} \quad (33.4)$$

a delay can be expected if the phase decreases at high frequencies while the gain has a constant slope. It is identified fitting a straight line by minimum squares to the phase values at high frequency. Do not forget that, as we saw in Chapter 24, phase (33.4) appears in a Bode diagram as an exponential, because of the semi-logarithmic scale of frequencies.

The phase of the delay in the Bode diagram is exponential

In Chapter 39 we will see additional techniques that may be used to identify a pure delay.

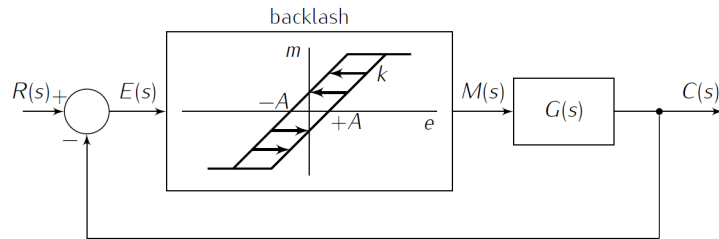


Figure 33.3: Block diagram of Exercise 1.

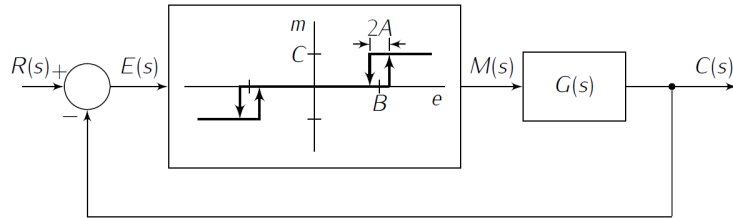


Figure 33.4: Block diagram of Exercise 2.

Exercises

1. The control system in Figure 33.3 includes plant $G(s) = \frac{1.5}{s(s+1)^2}$ and an actuator with backlash. A limit cycle with period 7.757 s and amplitude 5.45 is observed. Find A and k .
2. The control system in Figure 33.4 includes plant $G(s) = \frac{3}{s(s+1)^2}$ and the represented non-linearity, with $C = 1$. A limit cycle with 0.882 rad/s and amplitude 2.155 is observed. Find A and B .
3. The control system in Figure 33.5 includes the represented non-linearity, with $k = 1$. A limit cycle with 0.39 Hz and amplitude 10.2 is observed. Find A and K_1 .

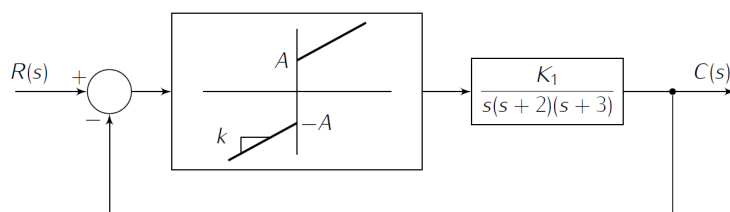


Figure 33.5: Block diagram of Exercise 3.

Part VII

Fractional order systems

‘I just take the train from platform nine and three-quarters at eleven o’clock,’ he read.

His aunt and uncle stared.

‘Platform what?’

‘Nine and three-quarters.’

‘Don’t talk rubbish,’ said Uncle Vernon, ‘there is no platform nine and three-quarters.’

J. K. ROWLING (1965 — . . .), *Harry Potter and the Philosopher’s Stone* (1997), 6

In this part of the lecture notes:

Chapter 34 introduces fractional order systems from their frequency responses and the corresponding identification methods.

Chapter 35 shows how fractional derivatives are the origin of fractional order systems.

Chapter 36 concerns the time responses of fractional order systems and the corresponding identification methods.

Here is what you need to know beforehand to follow these chapters:

- The Laplace and Fourier transforms, from Chapter 2;
- Transfer functions, from Sections 4.1 and 4.2 of Chapter 4;
- System theory, from Part II;
- Filters, from Sections 12.2 and 12.3 of Chapter 12;
- System identification, from Chapters 30 to 32.

Chapter 34

Fractional order systems and their frequency responses

Bien que la notion de dérivation non entière date du début du XIXème siècle à travers les travaux de CAUCHY, elle reste néanmoins une énigme dans le domaine de la physique appliquée et notamment en automatique.

Alain OUSTALOUP (1950 — ...), *La commande CRONE: commande robuste d'ordre non entier* (1991), Avant-propos

In this chapter we take a look at frequency responses that need to be modelled with fractional powers of s .

34.1 Frequency responses that require fractional order systems

We know from Chapter 11 that the frequency response of

$$G(s) = s^k, \quad k \in \mathbb{Z} \quad (34.1)$$

is

$$G(j\omega) = (j\omega)^k \quad (34.2)$$

$$20 \log_{10} |G(j\omega)| = 20k \log_{10} \omega \quad (34.3)$$

$$\arg G(j\omega) = k 90^\circ \quad (34.4)$$

At both high and low frequencies, given a transfer function

$$G(s) = \frac{b_m s^m + b_{m-1} s^{m-1} + \dots + b_1 s + b_0}{s^n + a_{n-1} s^{n-1} + \dots + a_1 s + a_0} \quad (34.5)$$

there will be one term that is increasingly larger than all others, and that is why, in both cases:

- the gain is linear with a slope which is an integer multiple of 20 dB/decade;
- the phase is constant and equal to an integer multiple of 90°.

As we saw in Sections 11.4 and 32.2, at low frequencies these values give us the number of poles or zeros at the origin, and at high frequencies $n - m$ (the difference between the number of poles and zeros).

Frequency responses with other gain slopes or constant values for phases require transfer functions with non-integer powers of the Laplace transform variable s , called **fractional transfer functions**. Such is the case of the response in Figure 34.1, which has, at high frequencies, a gain slope of -10 dB/decade, and a phase of 45° . It is easy to see that such a frequency response can be obtained with transfer function

$$G_1(s) = \frac{1}{s^{\frac{1}{2}} + 1} \quad (34.6)$$

$$G_1(j\omega) = \frac{1}{(j\omega)^{\frac{1}{2}} + 1} \quad (34.7)$$

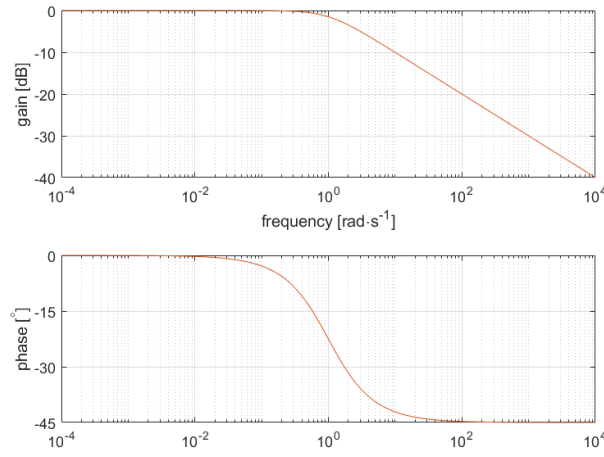


Figure 34.1: Frequency response of $\left(\frac{1}{s+1}\right)^{\frac{1}{2}}$.

since, for high frequencies, $\omega \gg 1$ and

$$G_1(j\omega) \approx (j\omega)^{-\frac{1}{2}} \quad (34.8)$$

$$20 \log_{10} |G_1(j\omega)| \approx 20 \log_{10} \omega^{-\frac{1}{2}} = -10 \log_{10} \omega \quad (34.9)$$

$$\angle G_1(j\omega) \approx \angle j^{-\frac{1}{2}} = -45^\circ \quad (34.10)$$

and also by transfer function

$$G_2(s) = \left(\frac{1}{s+1}\right)^{\frac{1}{2}} \quad (34.11)$$

$$G_2(j\omega) = \left(\frac{1}{j\omega+1}\right)^{\frac{1}{2}} \quad (34.12)$$

since, for high frequencies,

$$G_2(j\omega) \approx (j\omega)^{-\frac{1}{2}} \quad (34.13)$$

as well.

Notice, however, that the frequency responses of $G_1(s)$ and $G_2(s)$ are *not* the same. (Figure 34.1 actually corresponds to $G_2(j\omega)$, given by (34.12).) Transfer functions $G_1(s)$ and $G_2(s)$ belong to two different types of fractional transfer functions.

Definition 34.1. A fractional order transfer function is a transfer function with non-integer powers of the Laplace transform variable s and can be:

Implicit fractional transfer function

- **implicit**, given by:

$$\begin{aligned} G(s) &= \left(\frac{b_0 + b_1 s + \dots + b_m s^m}{a_0 + a_1 s + \dots + a_n s^n}\right)^\alpha \\ &= \left(\frac{\sum_{k=0}^m b_k s^k}{\sum_{k=0}^n a_k s^k}\right)^\alpha, \quad \alpha \in \mathbb{R}^+ \setminus \mathbb{N} \end{aligned} \quad (34.14)$$

Explicit fractional transfer function

- **explicit**, in which case it is the ratio of two linear combinations of powers of s , and may be

Commensurable fractional transfer function

- **commensurable**, if all powers of s are integer multiples of some $\alpha \in \mathbb{R}^+ \setminus \mathbb{N}$, and thus the transfer function is a ratio of two polynomials

in s^α :

$$\begin{aligned}
 G(s) &= \frac{b_0 + b_1 s^\alpha + \dots + b_m s^{m\alpha}}{a_0 + a_1 s^\alpha + \dots + a_n s^{n\alpha}} \\
 &= \frac{\sum_{k=0}^m b_k s^{k\alpha}}{\sum_{k=0}^n a_k s^{k\alpha}} \tag{34.15}
 \end{aligned}$$

– **non-commensurable**, if it cannot be put in the form of (34.15): *Non-commensurable fractional transfer function*

$$\begin{aligned}
 G(s) &= \frac{b_0 + b_1 s^{\beta_1} + \dots + b_m s^{\beta_m}}{a_0 + a_1 s^{\alpha_1} + \dots + a_n s^{\alpha_n}} \\
 &= \frac{b_0 + \sum_{k=1}^m b_k s^{\beta_k}}{a_0 + \sum_{k=1}^n a_k s^{\alpha_k}}, \quad \alpha_k, \beta_k > 0, \exists_k \alpha_k \notin \mathbb{N} \vee \beta_k \notin \mathbb{N} \tag{34.16}
 \end{aligned}$$

By contrast with these, transfer functions given by (9.3) may be called integer order transfer functions.

Remark 34.1. (34.16) reduces to (34.15) if it turns out that all the β_k in the numerator and all the α_k in the denominator are integer multiples of some **commensurability order** α . Both (34.14) and (34.15) reduce to integer order transfer functions if condition $\alpha \notin \mathbb{N}$ is relaxed. \square

Remark 34.2. *Non-integer* transfer functions would be a better name, since “fractional” orders may in fact be irrational, but the name *fractional* stuck. \square *“Fractional” orders can be irrational*

Remark 34.3. Implicit fractional transfer functions have this name because fractional powers of s are only implicitly found in (34.14). They are explicitly seen in (34.15)–(34.16). \square

We know that, as (2.45) shows, the Laplace transforms of derivatives originate integer powers of s . We will defer to the next chapter the study of fractional order derivatives that originate fractional powers of s , and proceed with the study of frequency responses of fractional transfer functions.

Where do fractional powers of s come from?

34.2 Frequency responses of fractional order systems

We must study separately each of the types of fractional order transfer functions in Definition 34.1.

Theorem 34.1. The gain and the phase of an implicit fractional transfer function $G(s)$ given by

$$G(s) = [G_i(s)]^\alpha \tag{34.17}$$

Frequency response of an implicit fractional transfer function

are the gain and the phase of the integer transfer function $G_i(s)$ scaled by α , i.e.

$$20 \log_{10} |G(j\omega)| = 20\alpha \log_{10} |G_i(j\omega)| \tag{34.18}$$

$$\angle G(j\omega) = \alpha \angle G_i(j\omega) \tag{34.19}$$

Proof. This is a straightforward result of the properties of the logarithm. \square

Example 34.1. The frequency response of (34.12) shown in Figure 34.1 is the frequency response of $\frac{1}{s+1}$ with the values of gains and phases halved. \square

A commensurable fractional transfer function $G(s)$ can be written as the product of simpler terms with only one or two powers of s . From their frequency responses it is possible to find that of $G(s)$, as done in Section 11.4 for integer transfer functions. In addition to (11.84)–(11.155), the other frequency responses that now serve as building blocks are the following:

Frequency response of a commensurable fractional transfer function

- s^α , a fractional power of s with $\alpha > 0$:

Frequency response of

$$G(j\omega) = (j\omega)^\alpha \quad (34.20)$$

$$20 \log_{10} |G(j\omega)| = 20\alpha \log_{10} \omega \text{ dB} \quad (34.21)$$

$$\angle G(j\omega) = 90^\circ \alpha \quad (34.22)$$

Frequency response of $\frac{1}{s^\alpha}$

- $\frac{1}{s^\alpha}$, a fractional power of $\frac{1}{s}$ with $\alpha > 0$:

$$G(j\omega) = \frac{1}{(j\omega)^\alpha} \quad (34.23)$$

$$20 \log_{10} |G(j\omega)| = -20\alpha \log_{10} \omega \text{ dB} \quad (34.24)$$

$$\angle G(j\omega) = -90^\circ \alpha \quad (34.25)$$

Frequency response of $\frac{1}{(\frac{s}{a})^\alpha + 1}$

- $\frac{1}{(\frac{s}{a})^\alpha + 1}$, a first-order polynomial on s^α in the denominator with $\alpha, a \in \mathbb{R}^+$:

$$\begin{aligned} G(j\omega) &= \frac{1}{(\frac{j\omega}{a})^\alpha + 1} \\ &= \frac{1}{(\cos \frac{\alpha\pi}{2} + j \sin \frac{\alpha\pi}{2}) (\frac{\omega}{a})^\alpha + 1} \end{aligned} \quad (34.26)$$

$$\begin{aligned} |G(j\omega)| &= \frac{1}{\sqrt{(1 + (\frac{\omega}{a})^\alpha \cos \frac{\alpha\pi}{2})^2 + (\frac{\omega}{a})^{2\alpha} \sin^2 \frac{\alpha\pi}{2}}} \\ &= \frac{1}{\sqrt{1 + (\frac{\omega}{a})^{2\alpha} + 2 (\frac{\omega}{a})^\alpha \cos \frac{\alpha\pi}{2}}} \end{aligned} \quad (34.27)$$

$$20 \log_{10} |G(j\omega)| \approx 0 \text{ dB}, \quad \omega \ll a \quad (34.28)$$

$$\angle G(j\omega) \approx 0^\circ, \quad \omega \ll a \quad (34.29)$$

$$\begin{aligned} 20 \log_{10} |G(ja)| &= 20 \log_{10} \left| \frac{1}{\cos \frac{\alpha\pi}{2} + j \sin \frac{\alpha\pi}{2} + 1} \right| \\ &= 20 \log_{10} \frac{1}{|1 + 2 \cos \frac{\alpha\pi}{2} + \cos^2 \frac{\alpha\pi}{2} + \sin^2 \frac{\alpha\pi}{2}|} \\ &= 20 \log_{10} \frac{1}{\sqrt{2 + 2 \cos \frac{\alpha\pi}{2}}} \\ &= -10 \log_{10} \left(2 + 2 \cos \frac{\alpha\pi}{2} \right) \end{aligned} \quad (34.30)$$

$$\begin{aligned} \angle G(ja) &= \arctan \frac{-\sin \frac{\alpha\pi}{2}}{1 + \cos \frac{\alpha\pi}{2}} = \arctan \frac{-2 \sin \frac{\alpha\pi}{4} \cos \frac{\alpha\pi}{4}}{2 \cos \frac{\alpha\pi}{4} \cos \frac{\alpha\pi}{4}} \\ &= -\alpha \frac{\pi}{4} \pm 2k\pi, \quad k \in \mathbb{Z} \end{aligned} \quad (34.31)$$

$$20 \log_{10} |G(j\omega)| \approx 20\alpha \log_{10} a - 20\alpha \log_{10} \omega \text{ dB}, \quad \omega \gg a \quad (34.32)$$

$$\angle G(j\omega) \approx -90^\circ \alpha, \quad \omega \gg a \quad (34.33)$$

Notice that the gain may decrease monotonously or it may first go up, have a peak value, and then decrease, depending on the value of α . Indeed, equating the derivative of (34.27) to zero,

$$\begin{aligned} \frac{d}{d\omega} \frac{1}{\sqrt{\zeta(\omega)}} &= 0 \\ \Leftrightarrow \frac{\frac{d}{d\omega} \zeta(\omega)}{-2\zeta(\omega)\sqrt{\zeta(\omega)}} &= 0 \\ \Leftrightarrow \frac{2\alpha\omega^{2\alpha-1}}{a^{2\alpha}} + \frac{2\alpha\omega^{\alpha-1}}{a^\alpha} \cos \frac{\alpha\pi}{2} &= 0 \\ \Leftrightarrow \left(\frac{\omega}{a}\right)^\alpha &= -\cos \frac{\alpha\pi}{2} \end{aligned} \quad (34.34)$$

Notice that the left hand side of the equation is always positive, while the right hand side will be positive if $\cos \frac{\alpha\pi}{2} < 0$. So (34.34) will have a

solution when

$$\begin{aligned} \frac{\pi}{2} + 2k\pi < \frac{\alpha\pi}{2} < \frac{3\pi}{2} + 2k\pi, k \in \mathbb{Z}_0^+ &\Leftrightarrow \\ \Leftrightarrow 4k + 1 < \alpha < 4k + 3, k \in \mathbb{Z}_0^+ &\quad (34.35) \end{aligned}$$

In other words, the gain will have a peak when $1 < \alpha < 3 \vee 5 < \alpha < 7 \vee 9 < \alpha < 11 \vee \dots$ (We know that this point where the derivative is equal to zero is a maximum, and not a minimum or a saddle point, because the gain is constant for low frequencies and decreases for high frequencies.) The frequency at which there is a peak is found from (34.34):

When the gain of $\frac{1}{(\frac{s}{a})^{\alpha+1}}$ has a peak

$$\omega_{\text{peak}} = a \left(-\cos \frac{\alpha\pi}{2} \right)^{\frac{1}{\alpha}} \quad (34.36)$$

The peak of the gain is found replacing this in (34.27):

$$\begin{aligned} \max_{\omega} |G(j\omega)| &= \frac{1}{\sqrt{1 + \cos^2 \frac{\alpha\pi}{2} + 2 \left(-\cos \frac{\alpha\pi}{2} \right) \cos \frac{\alpha\pi}{2}}} \\ &= \frac{1}{\sqrt{1 - \cos^2 \frac{\alpha\pi}{2}}} = \frac{1}{|\sin \frac{\alpha\pi}{2}|} \end{aligned} \quad (34.37)$$

The phase may decrease monotonously or increase monotonously, depending on the value of α . Indeed, the derivative of the phase

$$\begin{aligned} \angle G(j\omega) &= \angle \frac{1}{\left(\cos \frac{\alpha\pi}{2} + j \sin \frac{\alpha\pi}{2} \right) \left(\frac{\omega}{a} \right)^{\alpha} + 1} \\ &= -\arctan \frac{\overbrace{\left(\frac{\omega}{a} \right)^{\alpha} \sin \frac{\alpha\pi}{2}}^{S(\omega)}}{\underbrace{1 + \left(\frac{\omega}{a} \right)^{\alpha} \cos \frac{\alpha\pi}{2}}_{C(\omega)}} \end{aligned} \quad (34.38)$$

is given by

$$\frac{d}{d\omega} \angle G(j\omega) = - \frac{\frac{\alpha\omega^{\alpha-1}}{a^{\alpha}} \sin \frac{\alpha\pi}{2} \left(1 + \left(\frac{\omega}{a} \right)^{\alpha} \cos \frac{\alpha\pi}{2} \right) - \left(\frac{\omega}{a} \right)^{\alpha} \sin \frac{\alpha\pi}{2} \frac{\alpha\omega^{\alpha-1}}{a^{\alpha}} \cos \frac{\alpha\pi}{2}}{\left(1 + C(\omega) \right)^2} \frac{1}{1 + \left(\frac{S(\omega)}{1 + C(\omega)} \right)^2} \quad (34.39)$$

Notice that the denominator is always positive, and the numerator is a fraction with a denominator which cannot be negative. So, supposing that the derivative exists, its sign σ is

$$\begin{aligned} \sigma \left(\frac{d}{d\omega} \angle G(j\omega) \right) &\quad (34.40) \\ &= \sigma \left(- \underbrace{\frac{\alpha\omega^{\alpha-1}}{a^{\alpha}} \sin \frac{\alpha\pi}{2}}_{\in \mathbb{R}^+} - \underbrace{\frac{\alpha\omega^{2\alpha-1}}{a^{2\alpha}} \sin \frac{\alpha\pi}{2} \cos \frac{\alpha\pi}{2} + \frac{\alpha\omega^{2\alpha-1}}{a^{2\alpha}} \sin \frac{\alpha\pi}{2} \cos \frac{\alpha\pi}{2}}_0 \right) \end{aligned}$$

This means that the sign of the derivative is that of $-\sin \frac{\alpha\pi}{2}$. Consequently:

When the phase of $\frac{1}{(\frac{s}{a})^{\alpha+1}}$ goes up or down

– the phase will decrease from 0° to $\lfloor \frac{\alpha}{4} \rfloor 360^\circ - \alpha 90^\circ$ if

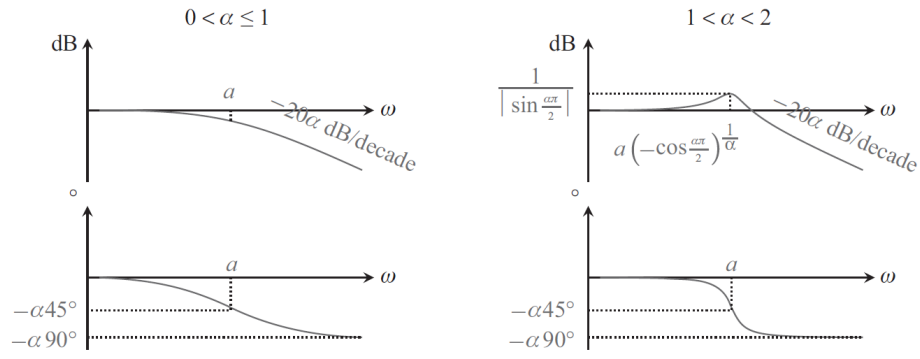
$$\begin{aligned} 2k\pi < \frac{\alpha\pi}{2} < (2k+1)\pi, k \in \mathbb{Z}_0^+ &\Leftrightarrow \\ 4k < \alpha < 4k+2, k \in \mathbb{Z}_0^+ &\quad (34.41) \end{aligned}$$

that is to say, if $0 < \alpha < 2 \vee 4 < \alpha < 6 \vee 8 < \alpha < 10 \vee \dots$;

– the phase will increase from 0° to $(1 + \lfloor \frac{\alpha}{4} \rfloor) 360^\circ - \alpha 90^\circ$ if

$$\begin{aligned} (2k-1)\pi < \frac{\alpha\pi}{2} < 2k\pi, k \in \mathbb{Z}^+ &\Leftrightarrow \\ 4k-2 < \alpha < 4k, k \in \mathbb{Z}^+ &\quad (34.42) \end{aligned}$$

that is to say, if $2 < \alpha < 4 \vee 6 < \alpha < 8 \vee 10 < \alpha < 12 \vee \dots$;


 Figure 34.2: Bode diagram of $\frac{1}{(\frac{s}{a})^\alpha + 1}$, $\alpha > 0$.

– the phase will remain constant and equal to 0° if

$$\begin{aligned} \frac{\alpha\pi}{2} &= 2k\pi, k \in \mathbb{Z}^+ \Leftrightarrow \\ \alpha &= 4k, k \in \mathbb{Z}^+ \end{aligned} \quad (34.43)$$

that is to say, if $\alpha \in \{4, 8, 12, 16 \dots\}$, since in that case the derivative is equal to 0;

– the phase will have a discontinuity and jump from 0° to $\pm 180^\circ$ if

$$\begin{aligned} \frac{\alpha\pi}{2} &= (2k+1)\pi, k \in \mathbb{Z}_0^+ \Leftrightarrow \\ \alpha &= 2 + 4k, k \in \mathbb{Z}_0^+ \end{aligned} \quad (34.44)$$

that is to say, if $\alpha \in \{2, 6, 10, 14 \dots\}$, since in that case, when $\omega = a$ rad/s, the denominator $(1 + C(\omega))^2$ in (34.39) is equal to $(1 + C(a))^2 = 1 + \cos \frac{\alpha\pi}{2} = 0$, and so the derivative is infinite at that point.

The two most common cases are $0 < \alpha \leq 1$ and $1 < \alpha < 2$ (as we will see in Chapter 36, no other cases can be stable). They correspond to the Bode diagrams in Figure 34.2.

Frequency response of $(\frac{s}{a})^\alpha + 1$

• $(\frac{s}{a})^\alpha + 1$, a first-order polynomial on s^α in the numerator with $\alpha, a \in \mathbb{R}^+$:

$$\begin{aligned} G(j\omega) &= \left(\frac{j\omega}{a}\right)^\alpha + 1 \\ &= \left(\cos \frac{\alpha\pi}{2} + j \sin \frac{\alpha\pi}{2}\right) \left(\frac{\omega}{a}\right)^\alpha + 1 \end{aligned} \quad (34.45)$$

$$20 \log_{10} |G(j\omega)| \approx 0 \text{ dB}, \omega \ll a \quad (34.46)$$

$$\angle G(j\omega) \approx 0^\circ, \omega \ll a \quad (34.47)$$

$$\begin{aligned} 20 \log_{10} |G(ja)| &= 20 \log_{10} \left| \cos \frac{\alpha\pi}{2} + j \sin \frac{\alpha\pi}{2} + 1 \right| \\ &= 20 \log_{10} \sqrt{2 + 2 \cos \frac{\alpha\pi}{2}} \\ &= 10 \log_{10} \left(2 + 2 \cos \frac{\alpha\pi}{2}\right) \end{aligned} \quad (34.48)$$

$$\begin{aligned} \angle G(ja) &= \arctan \frac{1 + \cos \frac{\alpha\pi}{2}}{-\sin \frac{\alpha\pi}{2}} = \arctan \frac{2 \cos \frac{\alpha\pi}{4} \cos \frac{\alpha\pi}{4}}{-2 \sin \frac{\alpha\pi}{4} \cos \frac{\alpha\pi}{4}} \\ &= \alpha \frac{\pi}{4} \pm 2k\pi, k \in \mathbb{Z} \end{aligned} \quad (34.49)$$

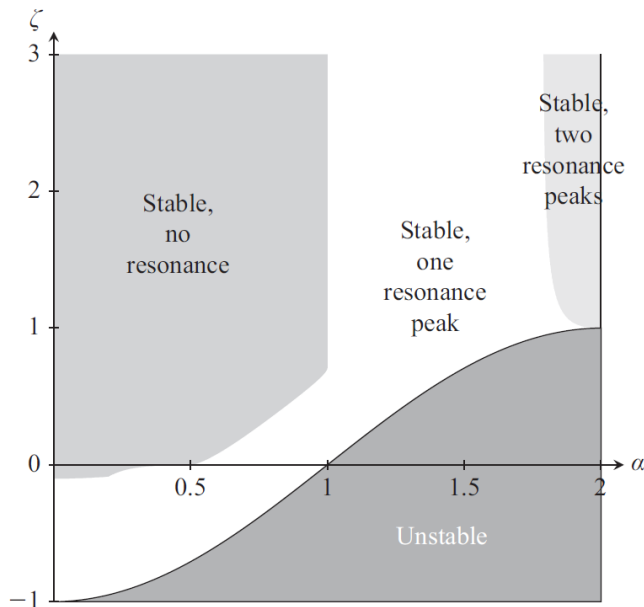
$$20 \log_{10} |G(j\omega)| \approx 20\alpha \log_{10} \omega - 20\alpha \log_{10} a \text{ dB}, \omega \gg a \quad (34.50)$$

$$\angle G(j\omega) \approx 90^\circ \alpha, \omega \gg a \quad (34.51)$$

The corresponding Bode diagram is that of $\frac{1}{(\frac{s}{a})^\alpha + 1}$, upside down.

Frequency response of $\frac{1}{(\frac{s}{a})^{2\alpha} + 2\zeta(\frac{s}{a})^\alpha + 1}$

• $\frac{1}{(\frac{s}{a})^{2\alpha} + 2\zeta(\frac{s}{a})^\alpha + 1}$, a second order polynomial on s^α in the denominator with


 Figure 34.3: Frequency behaviour of $\frac{1}{(\frac{s}{a})^{2\alpha} + 2\zeta(\frac{s}{a})^\alpha + 1}$.

$\alpha, a \in \mathbb{R}^+$:

$$G(j\omega) = \frac{1}{(\frac{j\omega}{a})^{2\alpha} + 2\zeta(\frac{j\omega}{a})^\alpha + 1} \quad (34.52)$$

$$20 \log_{10} |G(j\omega)| \approx 0 \text{ dB}, \quad \omega \ll a \quad (34.53)$$

$$\angle G(j\omega) \approx 0^\circ, \quad \omega \ll a \quad (34.54)$$

$$20 \log_{10} |G(j\omega)| \approx 40\alpha \log_{10} a - 40\alpha \log_{10} \omega \text{ dB}, \quad \omega \gg a \quad (34.55)$$

$$\angle G(j\omega) \approx -180^\circ \alpha, \quad \omega \gg a \quad (34.56)$$

While it is simple to find the conditions in which the gain has a resonance peak if $\alpha = 1$, when the order is fractional it is more expedient to find such conditions numerically. Figure 34.3 shows the combinations of values of α and ζ for which there are zero, one or two resonance frequencies (that is to say, for which the gain plot of the Bode diagram has zero, one or two local maxima).

- $(\frac{s}{a})^{2\alpha} + 2\zeta(\frac{s}{a})^\alpha + 1$, a second order polynomial on s^α in the numerator with $\alpha, a \in \mathbb{R}^+$: *Frequency response of*
 $(\frac{s}{a})^{2\alpha} + 2\zeta(\frac{s}{a})^\alpha + 1$

$$G(j\omega) = \left(\frac{j\omega}{a}\right)^{2\alpha} + 2\zeta\left(\frac{j\omega}{a}\right)^\alpha + 1 \quad (34.57)$$

$$20 \log_{10} |G(j\omega)| \approx 0 \text{ dB}, \quad \omega \ll a \quad (34.58)$$

$$\angle G(j\omega) \approx 0^\circ, \quad \omega \ll a \quad (34.59)$$

$$20 \log_{10} |G(j\omega)| \approx 40\alpha \log_{10} \omega - 40\alpha \log_{10} a \text{ dB}, \quad \omega \gg a \quad (34.60)$$

$$\angle G(j\omega) \approx 180^\circ \alpha, \quad \omega \gg a \quad (34.61)$$

The corresponding Bode diagram is that of $\frac{1}{(\frac{s}{a})^{2\alpha} + 2\zeta(\frac{s}{a})^\alpha + 1}$, upside down.

- In any of the cases above, if $a \in \mathbb{R}^-$, the gain behaviour remains the same, and the phase behaviour is symmetrical.

As to a non-commensurable fractional transfer function $G(s)$, if it can be written as a product of the terms above, its frequency response can be found in the same manner. Otherwise the only option is to compute $G(j\omega)$. *Frequency response of a non-commensurable fractional transfer function*

Example 34.2. Figure 34.4 shows the Bode diagram of

$$G(s) = \frac{s^{\frac{1}{2}} + 1}{\left(\left(\frac{s}{10^{-5}}\right)^{\frac{2}{3}} + 1\right) \left(\left(\frac{s}{10^5}\right)^{0.7} + 1\right)} \quad (34.62)$$

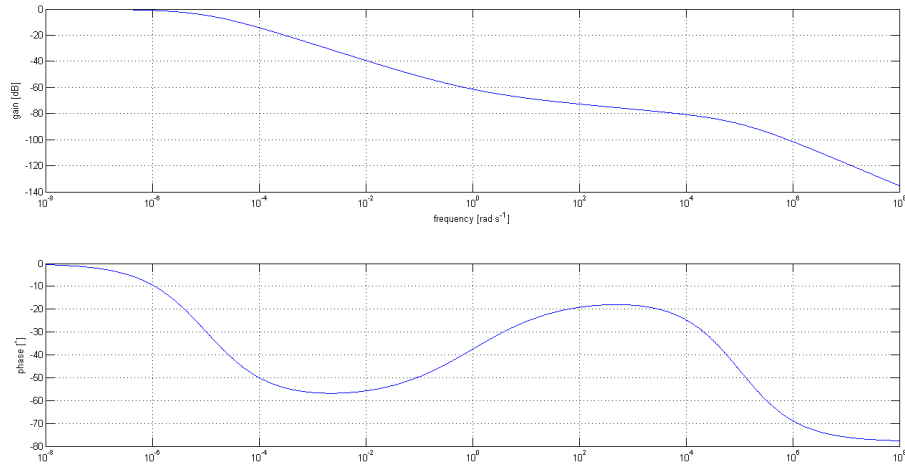


Figure 34.4: Bode diagram of (34.62), from Example 34.2.

which can be found from the frequency responses of its three constituent first order polynomials in fractional powers of s . Notice that:

- At $\omega = 10^{-5}$ rad/s, the phase goes down towards -60° , and the gain slopes down at -13.33 dB/decade.
- At $\omega = 1$ rad/s, the phase goes up 45° , towards -15° , and the gain bends up 10 dB/decade, ending with a still downwards slope of -3.33 dB/decade.
- At $\omega = 10^5$ rad/s, the phase goes down 63° , ending at -78° , and the gain bends down -14 dB/decade, ending with a downwards slope of -17.33 dB/decade.
- All this is in accord with the powers of s found in $G(s)$. □

34.3 Identification from the Bode diagram

Just as an integer model can be found from a Bode diagram by observation, so can a fractional model. The difference is that, with fractional models, all gain slopes and all phases are possible. If the model is implicit or commensurable, only a finite set of values will be found:

- gain slopes will have $k20\alpha$ dB/decade, $k \in \mathbb{Z}$;
- phases will have $k\alpha 90^\circ$, $k \in \mathbb{Z}$.

If the model is non-commensurable, any values can be found.

Fractional orders close to integers are hard to identify

Why this should not be a problem

When slopes are close to integer multiples of 20 dB/decade, and phases are close to integer multiples of 90° , it is difficult to distinguish a fractional order. And, after all, if that is the case, there may be no reason why the model should have a fractional order: the increase in model accuracy obtained with a fractional order may not compensate the difficulty of using a more difficult mathematical concept.

Example 34.3. Consider the frequency response in Figure 34.5:

- At low frequencies, the gain is constant and the phase is 0° . The plant behaves as a gain.
- At high frequencies, the gain has a slope of -10 dB and the phase is -45° . The plant behaves like $s^{-\frac{1}{2}}$.
- The asymptotes of the gain for low and high frequencies are shown in the figure, and intersect at 10 rad/s. It is also at that frequency that the phase is halfway between its low and high frequency values.

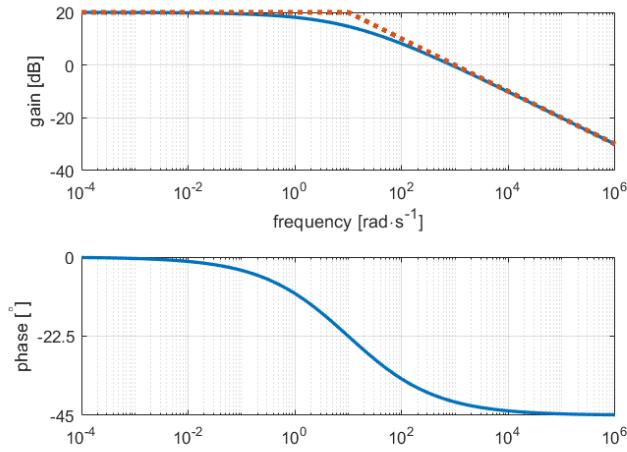


Figure 34.5: Frequency response of Example 34.3.

...ing for the corner Beware of the temptation of writing the model as
...ency

$$G(s) = \frac{k_1}{s^{\frac{1}{2}} + 10} \tag{34.63}$$

as if the plant were of integer order, and then finding the numerator k_1 from the value of the gain for low frequencies. As (34.30)–(34.31) show, the actual model should be

$$G(s) = \frac{k_2}{\left(\frac{s}{10}\right)^{\frac{1}{2}} + 1} \tag{34.64}$$

Since the order is not 1, (34.63) can never be the same as (34.64). The numerator k_2 is now found from the low frequency gain (20 dB for $\omega \approx 0$) as

$$20 = 20 \log_{10} \frac{k_2}{\left(\frac{0}{10}\right)^{\frac{1}{2}} + 1} \Rightarrow k_2 = 10 \tag{34.65}$$

Thus, $G(s) = \frac{10}{\left(\frac{s}{10}\right)^{\frac{1}{2}} + 1}$.

34.4 Levy's method extended

Levy's method can be extended to identify implicit and commensurable fractional transfer functions.

Theorem 34.2. Let a plant's frequency response $G(j\omega_p)$ be known at f frequencies, i.e. $p = 1, 2, \dots, f$. The implicit fractional model given by

Levy's method for implicit fractional order models

$$\hat{G}(s) = \hat{G}_i^\alpha(s) = \frac{N^\alpha(s)}{D^\alpha(s)} = \left(\frac{b_0 + b_1s + \dots + b_ms^m}{1 + a_1s + \dots + a_ns^n} \right)^\alpha = \left(\frac{\sum_{k=0}^m b_k s^k}{1 + \sum_{k=1}^n a_k s^k} \right)^\alpha \tag{34.66}$$

which minimises quadratic error

What Levy's method minimises

$$\varepsilon = \sum_{p=1}^f \left| G^{\frac{1}{\alpha}}(j\omega_p) D(j\omega_p) - N(j\omega_p) \right|^2 \tag{34.67}$$

is found identifying integer model $\hat{G}_i(s)$ from $G^{\frac{1}{\alpha}}(j\omega_p)$ using Levy's method.

Proof. This is a straightforward consequence of (34.18)–(34.19). □

Theorem 34.3. Let a plant's frequency response $G(j\omega_p)$ be known at f frequencies, i.e. $p = 1, 2, \dots, f$. The model

Levy's method for mensurate fractional models

$$\hat{G}(s) = \frac{N(s)}{D(s)} = \frac{\sum_{k=0}^m b_k s^{k\alpha}}{1 + \sum_{k=1}^n a_k s^{k\alpha}} \quad (34.68)$$

with parameters to be determined

$$\mathbf{b} = [b_0 \quad \dots \quad b_m]^T \quad (34.69)$$

$$\mathbf{a} = [a_1 \quad \dots \quad a_n]^T \quad (34.70)$$

What Levy's method minimises which minimises quadratic error

$$\begin{aligned} \varepsilon &= \sum_{p=1}^f \left| \left(G(j\omega_p) - \hat{G}(j\omega_p) \right) D(j\omega_p) \right|^2 \\ &= \sum_{p=1}^f \left| G(j\omega_p) D(j\omega_p) - N(j\omega_p) \right|^2 \end{aligned} \quad (34.71)$$

is found solving

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{e} \\ \mathbf{g} \end{bmatrix} \quad (34.72)$$

$$\mathbf{A} = \sum_{p=1}^f \mathbf{A}_p \quad (34.73)$$

$$\mathbf{B} = \sum_{p=1}^f \mathbf{B}_p \quad (34.74)$$

$$\mathbf{C} = \sum_{p=1}^f \mathbf{C}_p \quad (34.75)$$

$$\mathbf{D} = \sum_{p=1}^f \mathbf{D}_p \quad (34.76)$$

$$\mathbf{e} = \sum_{p=1}^f \mathbf{e}_p \quad (34.77)$$

$$\mathbf{g} = \sum_{p=1}^f \mathbf{g}_p \quad (34.78)$$

where the elements in line l and column c of matrixes \mathbf{A}_p , \mathbf{B}_p , \mathbf{C}_p and \mathbf{D}_p are given by

$$\mathbf{A}_{p;l,c} = -\Re [(j\omega_p)^{l\alpha}] \Re [(j\omega_p)^{c\alpha}] - \Im [(j\omega_p)^{l\alpha}] \Im [(j\omega_p)^{c\alpha}], \quad (34.79)$$

$$l = 0 \dots m \wedge c = 0 \dots m$$

$$\begin{aligned} \mathbf{B}_{p;l,c} &= \Re [(j\omega_p)^{l\alpha}] \Re [(j\omega_p)^{c\alpha}] \Re [G(j\omega_p)] + \Im [(j\omega_p)^{l\alpha}] \Re [(j\omega_p)^{c\alpha}] \Im [G(j\omega_p)] \\ &\quad - \Re [(j\omega_p)^{l\alpha}] \Im [(j\omega_p)^{c\alpha}] \Re [G(j\omega_p)] + \Im [(j\omega_p)^{l\alpha}] \Im [(j\omega_p)^{c\alpha}] \Re [G(j\omega_p)], \\ & \quad l = 0 \dots m \wedge c = 1 \dots n \end{aligned} \quad (34.80)$$

$$\begin{aligned} \mathbf{C}_{p;l,c} &= -\Re [(j\omega_p)^{l\alpha}] \Re [(j\omega_p)^{c\alpha}] \Re [G(j\omega_p)] + \Im [(j\omega_p)^{l\alpha}] \Re [(j\omega_p)^{c\alpha}] \Im [G(j\omega_p)] \\ &\quad - \Re [(j\omega_p)^{l\alpha}] \Im [(j\omega_p)^{c\alpha}] \Re [G(j\omega_p)] - \Im [(j\omega_p)^{l\alpha}] \Im [(j\omega_p)^{c\alpha}] \Re [G(j\omega_p)], \\ & \quad l = 1 \dots n \wedge c = 0 \dots m \end{aligned} \quad (34.81)$$

$$\begin{aligned} \mathbf{D}_{p;l,c} &= \left(\Re [G(j\omega_p)]^2 + \Im [G(j\omega_p)]^2 \right) \left(\Re [(j\omega_p)^{l\alpha}] \Re [(j\omega_p)^{c\alpha}] + \Im [(j\omega_p)^{l\alpha}] \Im [(j\omega_p)^{c\alpha}] \right), \\ & \quad l = 1 \dots n \wedge c = 1 \dots n \end{aligned} \quad (34.82)$$

and the elements of vectors \mathbf{e}_p and \mathbf{g}_p are given by

$$\mathbf{e}_{p;l,1} = -\Re [(j\omega_p)^{l\alpha}] \Re [G(j\omega_p)] - \Im [(j\omega_p)^{l\alpha}] \Im [G(j\omega_p)], \quad l = 0 \dots m \quad (34.83)$$

$$\mathbf{g}_{p;l,1} = -\Re [(j\omega_p)^{l\alpha}] \left(\Re [G(j\omega_p)]^2 + \Im [G(j\omega_p)]^2 \right), \quad l = 1 \dots n \quad (34.84)$$

Proof. The proof is similar to that of Theorem 32.1. Error (34.71) is given by

$$\begin{aligned} \varepsilon &= |GD - N|^2 \\ &= \left| \left(\Re[G] + j\Im[G] \right) \left(\Re[D] + j\Im[D] \right) - \left(\Re[N] + j\Im[N] \right) \right|^2 \\ &= \left| \left(\Re[G]\Re[D] - \Im[G]\Im[D] - \Re[N] \right) + j \left(\Re[G]\Im[D] + \Im[G]\Re[D] - \Im[N] \right) \right|^2 \\ &= \left(\Re[G]\Re[D] - \Im[G]\Im[D] - \Re[N] \right)^2 + \left(\Re[G]\Im[D] + \Im[G]\Re[D] - \Im[N] \right)^2 \end{aligned} \quad (34.85)$$

To minimise ε , we want its derivatives in order to coefficients a_k and b_k to be zero.

Let us suppose that there is only one frequency ω_p . First we find

$$\frac{d\Re[G]}{da_i} = \frac{d\Re[G]}{db_i} = \frac{d\Im[G]}{da_i} = \frac{d\Im[G]}{db_i} = 0 \quad (34.86)$$

$$\frac{d\Re[N]}{da_i} = 0 \quad (34.87)$$

$$\frac{d\Re[N]}{db_i} = \frac{d}{db_i} \sum_{k=0}^m b_k \Re [(j\omega_p)^{k\alpha}] = \Re [(j\omega_p)^{i\alpha}] \quad (34.88)$$

$$\frac{d\Im[N]}{da_i} = 0 \quad (34.89)$$

$$\frac{d\Im[N]}{db_i} = \frac{d}{db_i} \sum_{k=0}^m b_k \Im [(j\omega_p)^{k\alpha}] = \Im [(j\omega_p)^{i\alpha}] \quad (34.90)$$

$$\frac{d\Re[D]}{da_i} = \frac{d}{da_i} \left\{ 1 + \sum_{k=1}^n a_k \Re [(j\omega_p)^{k\alpha}] \right\} = \Re [(j\omega_p)^{i\alpha}] \quad (34.91)$$

$$\frac{d\Re[D]}{db_i} = 0 \quad (34.92)$$

$$\frac{d\Im[D]}{da_i} = \frac{d}{da_i} \left\{ 1 + \sum_{k=1}^n a_k \Im [(j\omega_p)^{k\alpha}] \right\} = \Im [(j\omega_p)^{i\alpha}] \quad (34.93)$$

$$\frac{d\Im[D]}{db_i} = 0 \quad (34.94)$$

(Notice that $j^{k\alpha}$ can now assume any value, and will, in the general case, have both a real part and an imaginary part is zero. This makes matrixes and vectors (34.79)–(34.84) more complicated than in the integer case.) Consequently,

$$\begin{aligned} \frac{\partial \varepsilon}{\partial a_i} &= 2 \left(\Re[G]\Re[D] - \Im[G]\Im[D] - \Re[N] \right) \left(\Re[G]\Re [(j\omega_p)^{i\alpha}] - \Im[G]\Im [(j\omega_p)^{i\alpha}] \right) \\ &\quad + 2 \left(\Re[G]\Im[D] + \Im[G]\Re[D] - \Im[N] \right) \left(\Re[G]\Im [(j\omega_p)^{i\alpha}] + \Im[G]\Re [(j\omega_p)^{i\alpha}] \right) \end{aligned} \quad (34.95)$$

$$\begin{aligned} \frac{\partial \varepsilon}{\partial b_i} &= 2 \left(\Re[G]\Re[D] - \Im[G]\Im[D] - \Re[N] \right) \left(-\Re [(j\omega_p)^{i\alpha}] \right) \\ &\quad + 2 \left(\Re[G]\Im[D] + \Im[G]\Re[D] - \Im[N] \right) \left(-\Im [(j\omega_p)^{i\alpha}] \right) \end{aligned} \quad (34.96)$$

Equaling to zero,

$$\begin{aligned} \frac{\partial \varepsilon}{\partial a_i} = 0 \Leftrightarrow & \Re[D] \left(\Im[G]^2 + \Re[G]^2 \right) \Re[(j\omega_p)^{i\alpha}] + \\ & \Im[D] \left(\Im[G]^2 + \Re[G]^2 \right) \Im[(j\omega_p)^{i\alpha}] + \\ & \Re[N] \left(\Im[G] \Im[(j\omega)^{i\alpha}] - \Re[G] \Re[(j\omega_p)^{i\alpha}] \right) + \\ & \Im[N] \left(-\Im[G] \Re[(j\omega_p)^{i\alpha}] - \Re[G] \Im[(j\omega_p)^{i\alpha}] \right) = 0 \end{aligned} \quad (34.97)$$

$$\begin{aligned} \frac{\partial \varepsilon}{\partial b_i} = 0 \Leftrightarrow & \left(\Re[G] \Re[D] - \Im[G] \Im[D] - \Re[N] \right) \Re[(j\omega_p)^{i\alpha}] + \\ & \left(\Re[G] \Im[D] + \Im[G] \Re[D] - \Im[N] \right) \Im[(j\omega_p)^{i\alpha}] = 0 \end{aligned} \quad (34.98)$$

The $m+1$ equations given by (34.97) and the n equations given by (34.98) form linear system (34.72), with matrixes and vectors defined by (34.79)–(34.84).

Considering f frequencies instead of only one, we arrive at (34.73)–(34.78). \square

Remark 34.4. Remarks on Theorem 32.1 apply here too. In particular, if the model should have $a_0 = 0$ in the denominator, the frequency response of s^α must be subtracted first, just as was done for the integer case. \square

Remark 34.5. Notice that in the fractional case not only n and m (which are integer) must be fixed in advance, but also α (which has the additional disadvantage of not being an integer). Reasonable ranges for these parameters may be found by inspection of the Bode diagram. \square

Glossary

Lo anterior se refiere a los idiomas del hemisferio austral. En los del hemisferio boreal (de cuya *Ursprache* hay muy pocos datos en el Onceno Tomo) la célula primordial no es el verbo, sino el adjetivo monosilábico. El sustantivo se forma por acumulación de adjetivos. No se dice *luna*: se dice *aéreo-claro sobre oscuro-redondo* o *anaranjado-tenué-del cielo* o cualquier otra agregación. En el caso elegido la masa de adjetivos corresponde a un objeto real; el hecho es puramente fortuito.

Jorge Luis BORGES (1899 — †1986), *Tlön, Uqbar, Orbis Tertius* (1940), II

commensurate fractional transfer function função de transferência fracionária comensurável

explicit fractional transfer function função de transferência fracionária explícita

implicit fractional transfer function função de transferência fracionária implícita

integer order transfer function função de transferência de ordem inteira

fractional order transfer function função de transferência de ordem fracionária

Exercises

1. Establish a correspondence between Bode diagrams a and b in Figure 34.6, polar plots A and B in Figure 34.7, and the following transfer functions:

$$G_1(s) = \frac{1}{s(s+1)^{1.5}(s+2)} \quad (34.99)$$

$$G_2(s) = \frac{64.47 + 12.46s}{0.598 + 39.96s^{1.25}} \quad (34.100)$$

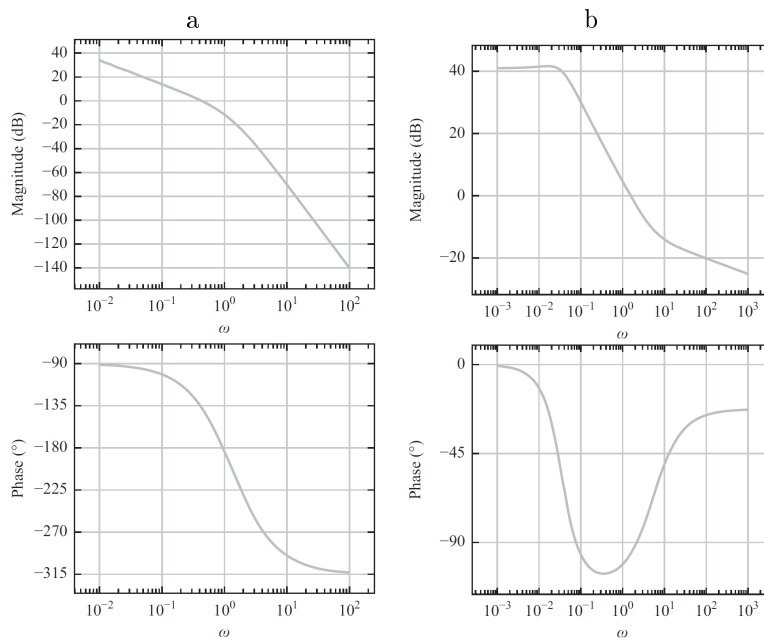


Figure 34.6: Bode diagrams of Exercise 1.

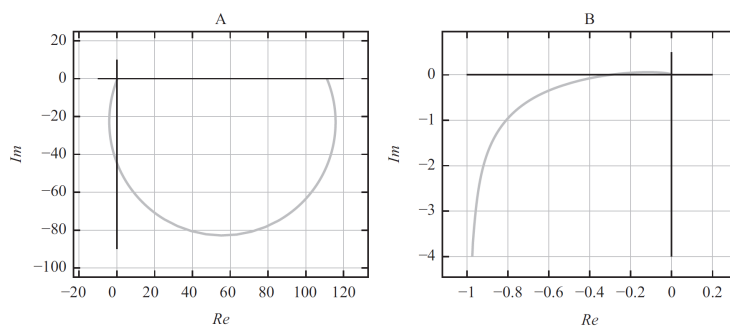


Figure 34.7: Polar plots of Exercise 1.

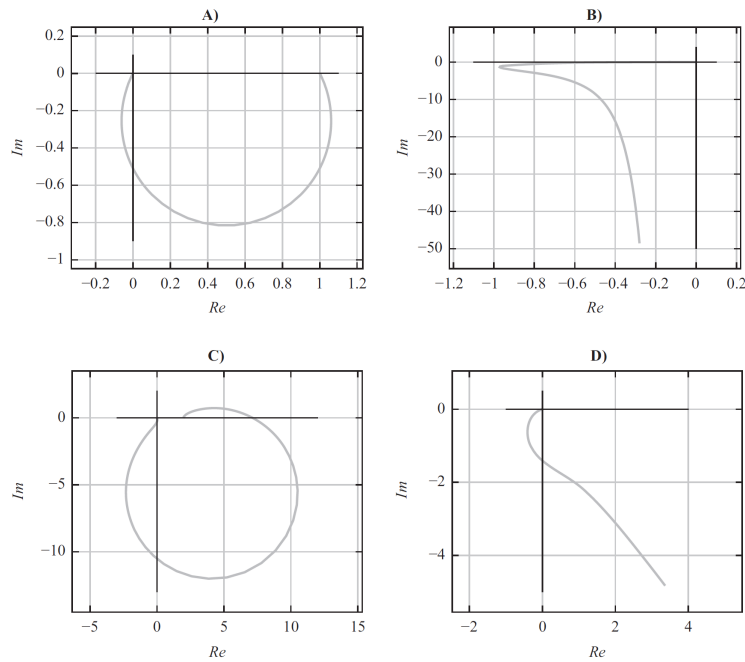


Figure 34.8: Polar plots of Exercise 2.

2. Establish a correspondence between polar plots A to D in Figure 34.8, Bode diagrams a to d in Figure 34.9, and the following transfer functions:

$$G_1(s) = \frac{1}{s^{1.3} + 1} \quad (34.101)$$

$$G_2(s) = \frac{1}{s(s^{1.3} + 1)} \quad (34.102)$$

$$G_3(s) = \frac{1}{s^{0.6}(s^{1.3} + 1)} \quad (34.103)$$

$$G_4(s) = \frac{s^{1.1} + 2}{s^{1.9} + 1} \quad (34.104)$$

3. Figure 34.10 shows the Bode diagram of a LTI system. Find its transfer function.
4. Figure 34.11 shows the Bode diagram of a LTI system. Find its transfer function.

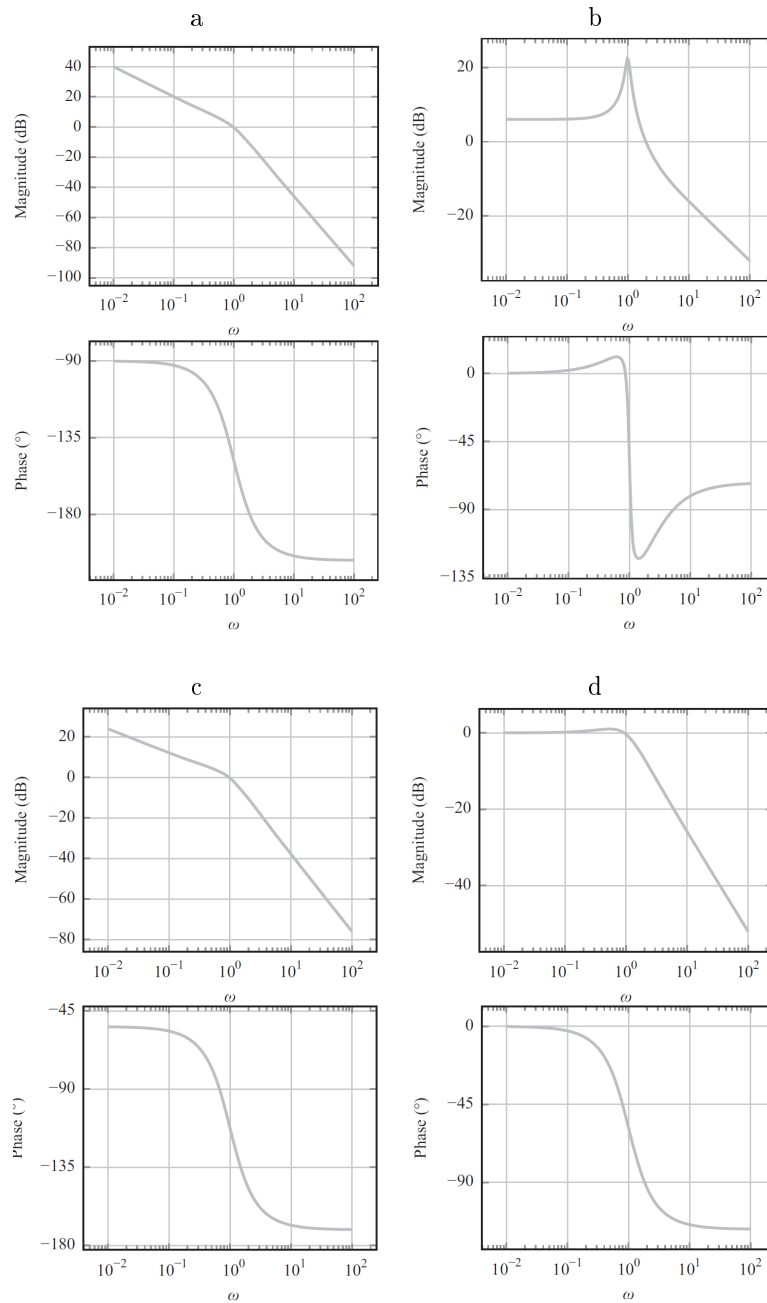


Figure 34.9: Bode diagrams of Exercise 2.

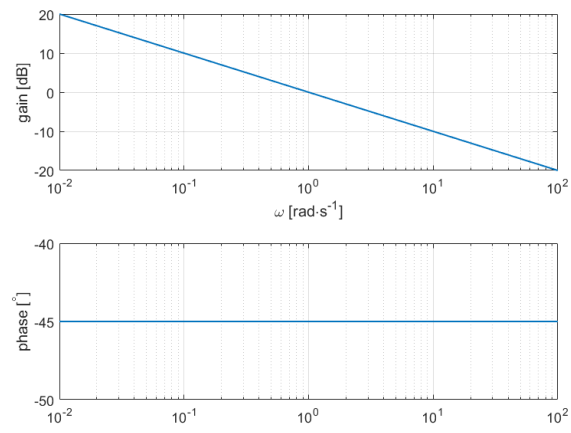


Figure 34.10: Bode diagram of Exercise 3.

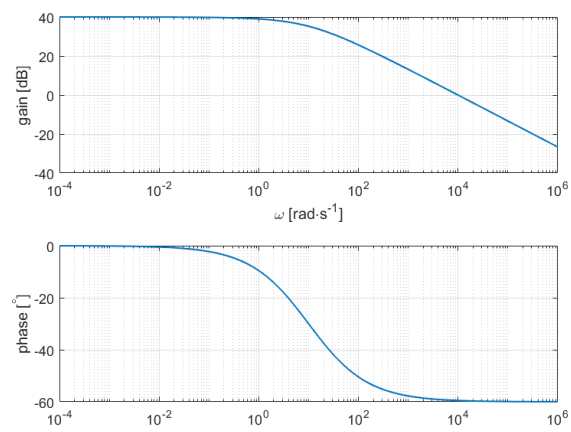


Figure 34.11: Bode diagram of Exercise 4.

Chapter 35

Fractional order derivatives

Students of mathematics early encounter the differential operators d/dx , d^2/dx^2 , d^3/dx^3 , etc., and some doubtless ponder whether it is necessary for the order of differentiation to be an integer. Why should there not be a $d^{1/2}/dx^{1/2}$ operator, for instance? Or d^{-1}/dx^{-1} or even $d^{\sqrt{2}}/dx^{\sqrt{2}}$? It is to these and related questions that the present work is addressed. It will come as no surprise to one versed in the calculus that the operator d^{-1}/dx^{-1} is nothing but an indefinite integral in disguise, but fractional orders of differentiation are more mysterious because they have no obvious geometric interpretation along the lines of the customary introduction to derivatives and integrals as slopes and areas. The reader who is prepared to dispense with a pictorial representation, however, will soon find that fractional order derivatives and integrals are just as tangible as those of integer order and that a new dimension in mathematics opens to him when the order q of the operator d^q/dx^q becomes an arbitrary parameter. Nor is this a sterile exercise in pure mathematics—many problems in the physical sciences can be expressed and solved succinctly by recourse to the fractional calculus.

Keith B. OLDHAM (1929 — ...), Jerome SPANIER (1930 — ...), *The Fractional Calculus: Theory and Applications of Differentiation and Integration to Arbitrary Order* (1974), Preface

In this chapter, fractional derivatives are introduced. They are at the origin of fractional order transfer functions, and generalise the concept of derivative to orders that do not need to be integer numbers. In fact, at the same time, they also generalise the concept of integral, since the Fundamental Theorem of Calculus shows we can identify $\frac{df(t)}{dt}$ with $\int f(t) dt$, $\frac{d^2f(t)}{dt^2}$ with $\int \int f(t) dt dt$, and so on.

35.1 Gamma function

Before generalising derivatives to arbitrary orders, it is necessary to generalise the factorial to arbitrary orders. This is done with the Gamma function.

Definition 35.1. For $x \in \mathbb{R}^+$, function Γ is defined as

Definition of Γ

$$\Gamma(x) = \int_0^{+\infty} e^{-y} y^{x-1} dy \quad \square \quad (35.1)$$

Lemma 35.1.

$$\Gamma(1) = \int_0^{+\infty} e^{-y} dy = [-e^{-y}]_0^{+\infty} = 1 \quad \square \quad (35.2)$$

Theorem 35.1.

$$\Gamma(x+1) = x\Gamma(x) \quad (35.3)$$

Proof.

$$\Gamma(x+1) = \int_0^{+\infty} e^{-y} y^x \, dy = \underbrace{[-e^{-y} y^x]_{y=0}^{+\infty}}_0 - \underbrace{\int_0^{+\infty} -e^{-y} x y^{x-1} \, dy}_{x\Gamma(x)} \quad (35.4)$$

$$\Gamma(n) = (n-1)!$$

Remark 35.1. From (35.2) and from (35.3) it can be easily seen that

$$\Gamma(2) = 1 \times \Gamma(1) = 1 \times 1 = 1 = 1! \quad (35.5)$$

$$\Gamma(3) = 2 \times \Gamma(2) = 2 \times 1 = 2 = 2! \quad (35.6)$$

$$\Gamma(4) = 3 \times \Gamma(3) = 3 \times 2 = 6 = 3! \quad (35.7)$$

$$\Gamma(5) = 4 \times \Gamma(4) = 4 \times 6 = 24 = 4! \quad (35.8)$$

⋮

$$\Gamma(n) = (n-1)!, n \in \mathbb{N} \quad (35.9)$$

It is this property that makes function Γ a generalisation of the factorial, in spite of the shift in the argument. \square

Theorem 35.2.

$$\lim_{x \rightarrow 0^+} \Gamma(x) = +\infty \quad (35.10)$$

Proof.

$$\Gamma(x) = \int_0^{+\infty} e^{-y} y^{x-1} \, dy > \int_0^1 e^{-y} y^{x-1} \, dy > \int_0^1 e^{-1} y^{x-1} \, dy = \frac{1}{e} \left[\frac{y^x}{x} \right]_{y=0}^1 = \frac{1}{ex} \quad (35.11)$$

and since $\lim_{x \rightarrow 0^+} \frac{1}{ex} = +\infty$ the result follows. \square

Iterating (35.3), we get

$$\Gamma(x+n) = \underbrace{(x+n-1) \underbrace{(x+n-2) \dots (x+1)}_{\Gamma(x+n-1)} x \Gamma(x)}_{\Gamma(x+n)} = \Gamma(x) \prod_{k=0}^{n-1} (x+k), \quad n \in \mathbb{N} \quad (35.12)$$

This expression allows defining the Γ function for negative arguments:

Definition 35.2. For $x \in \mathbb{R} \setminus \mathbb{Z}^-$, function Gamma is defined as

$$\Gamma(x) = \frac{\Gamma(x - \lfloor x \rfloor)}{-\lfloor x \rfloor - 1} \prod_{k=0}^{-\lfloor x \rfloor - 1} (x+k) \quad \square \quad (35.13)$$

In this way, by construction, (35.3) and (35.12) remain valid for all $x \in \mathbb{R}$. The evolution of $\Gamma(x)$ around $x = 0$ is shown in figure 35.1. From the figure it is clear that all non-positive integers are poles of function Γ . In other words:

Theorem 35.3.

$$\lim_{x \rightarrow n} \Gamma(x) = \infty, \quad n \in \mathbb{Z}_0^- \quad (35.14)$$

Proof. This is a consequence of (35.10) and (35.13). \square

Below we will need the following results:

Theorem 35.4. For any $n \in \mathbb{Z}$,

$$\Gamma(x)\Gamma(-x+1) = (-1)^n \Gamma(-x-n+1)\Gamma(x+n) \quad (35.15)$$

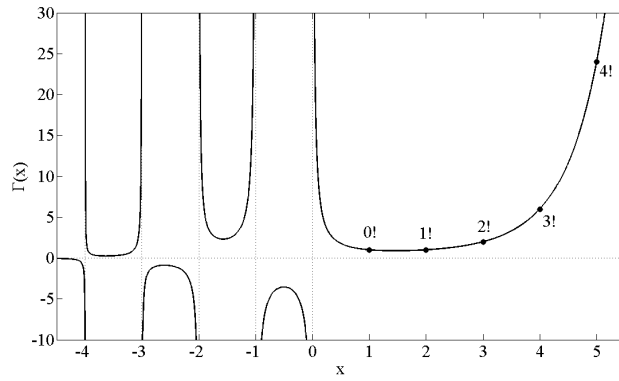


Figure 35.1: The Γ function.

Proof. If $n = 0$, the equality is obvious. For positive values of n , the equality is proved by mathematical induction. Using (35.3) twice, (35.15) is seen to hold for $n = 1$:

$$\Gamma(x)\Gamma(-x + 1) = \frac{\Gamma(x + 1)}{x} (-x\Gamma(-x)) = -\Gamma(-x)\Gamma(x + 1) \quad (35.16)$$

The inductive step is proved applying (35.16) to the right hand side of (35.15):

$$\begin{aligned} \Gamma(x)\Gamma(-x + 1) &= (-1)^n \Gamma(-x - n + 1)\Gamma(x + n) \\ &= (-1)^n [-\Gamma(-x - n)\Gamma(x + n + 1)] \\ &= (-1)^{n+1} \Gamma(-x - (n + 1) + 1)\Gamma(x + (n + 1)) \end{aligned} \quad (35.17)$$

For negative values of n , the equality is also proved by mathematical induction, in a similar manner. Using (35.3) twice, (35.15) is seen to hold for $n = -1$:

$$\Gamma(x)\Gamma(-x + 1) = (x - 1)\Gamma(x - 1) \frac{\Gamma(-x + 2)}{-x + 1} = -\Gamma(-x + 2)\Gamma(x - 1) \quad (35.18)$$

The inductive step is proved applying this to the right hand side of (35.15):

$$\begin{aligned} \Gamma(x)\Gamma(-x + 1) &= (-1)^n \Gamma(-x - n + 1)\Gamma(x + n) \\ &= (-1)^n [-\Gamma(-x - n + 2)\Gamma(x + n - 1)] \\ &= (-1)^{n+1} \Gamma(-x - (n - 1) + 1)\Gamma(x + (n - 1)) \end{aligned} \quad (35.19)$$

□

Corollary 35.1. From (35.12) and (35.15) we obtain

$$\prod_{k=0}^{n-1} (x + k) = \frac{\Gamma(x + n)}{\Gamma(x)} = (-1)^n \frac{\Gamma(-x + 1)}{\Gamma(-x - n + 1)} \quad (35.20)$$

$$\prod_{k=0}^{n-1} (x - k) = \prod_{k=0}^{n-1} (-1)(-x + k) = (-1)^n \frac{\Gamma(-x + n)}{\Gamma(-x)} = \frac{\Gamma(x + 1)}{\Gamma(x - n + 1)} \quad (35.21)$$

Thanks to this generalisation of the factorial, combinations of a things, b at a time, usually defined for integer non-negative arguments as

Combinations of a things, b at a time

$$\binom{a}{b} = \frac{a!}{b!(a - b)!}, \quad a, b \in \mathbb{Z}_0^+ \quad (35.22)$$

can be generalised as

$$\binom{a}{b} = \frac{\Gamma(a + 1)}{\Gamma(b + 1)\Gamma(a - b + 1)} \quad (35.23)$$

which makes sense if $a, b, a - b \in \mathbb{R} \setminus \mathbb{Z}^-$. Replacing (35.15) in (35.23), we obtain

$$\binom{a}{b} = \frac{(-1)^b \Gamma(b - a)}{\Gamma(b + 1)\Gamma(-a)} \quad (35.24)$$

which makes sense also if $a \in \mathbb{Z}^- \wedge b \in \mathbb{Z}_0^+$. (But notice that, if $b - a \in \mathbb{Z}^-$ or if $a = 0$, it may be that (35.23) makes sense, while (35.24) never does.) Furthermore,

- when $n \in \mathbb{Z}^- \wedge a \notin \mathbb{Z}^-$,

$$\lim_{b \rightarrow n} \binom{a}{b} = \frac{\overbrace{\Gamma(a+1)}^{\in \mathbb{R} \setminus \{0\}}}{\underbrace{\lim_{b \rightarrow n} \Gamma(b+1)}_{\infty} \underbrace{\lim_{b \rightarrow n} \Gamma(a-b+1)}_{\in \mathbb{R} \setminus \{0\}}} = 0 \quad (35.25)$$

- when $n \in \mathbb{Z}^- \wedge a \notin \mathbb{Z}^-$,

$$\lim_{a-b \rightarrow n} \binom{a}{b} = \frac{\overbrace{\Gamma(a+1)}^{\in \mathbb{R} \setminus \{0\}}}{\underbrace{\lim_{b \rightarrow a-n} \Gamma(b+1)}_{\in \mathbb{R} \setminus \{0\}} \underbrace{\lim_{a-b \rightarrow n} \Gamma(a-b+1)}_{\infty}} = 0 \quad (35.26)$$

- when $m, n \in \mathbb{Z}^- \wedge |m| > |n|$,

$$\lim_{(a,b) \rightarrow (m,n)} \binom{a}{b} = \frac{(-1)^b \overbrace{\lim_{b-a \rightarrow n-m \in \mathbb{N}} \Gamma(b-a)}^{\in \mathbb{R} \setminus \{0\}}}{\underbrace{\lim_{b \rightarrow n} \Gamma(b+1)}_{\infty} \underbrace{\lim_{a \rightarrow m} \Gamma(-a)}_{\in \mathbb{R} \setminus \{0\}}} = 0 \quad (35.27)$$

Putting all this together, combinations are defined as follows.

Definition 35.3.

$$\binom{a}{b} = \begin{cases} \frac{\Gamma(a+1)}{\Gamma(b+1)\Gamma(a-b+1)}, & \text{if } a, b, a-b \notin \mathbb{Z}^- \\ \frac{(-1)^b \Gamma(b-a)}{\Gamma(b+1)\Gamma(-a)}, & \text{if } a \in \mathbb{Z}^- \wedge b \in \mathbb{Z}_0^+ \\ 0, & \text{if } [(b \in \mathbb{Z}^- \vee b-a \in \mathbb{N}) \wedge a \notin \mathbb{Z}^-] \vee (a, b \in \mathbb{Z}^- \wedge |a| > |b|) \end{cases} \quad \square \quad (35.28)$$

Later in this chapter, we will need the following results:

Theorem 35.5.

$$\binom{a}{b} + \binom{a}{b-1} = \binom{a+1}{b} \quad (35.29)$$

for all $a, b \in \mathbb{R}$ for which the combinations above exist.

Proof. From (35.23),

$$\begin{aligned} \binom{a}{b} + \binom{a}{b-1} &= \frac{\Gamma(a+1)}{\Gamma(b+1)\Gamma(a-b+1)} + \frac{\Gamma(a+1)}{\Gamma(b)\Gamma(a-b+2)} \quad (35.30) \\ &= \frac{(a-b+1)\Gamma(a+1)}{\Gamma(b+1)\underbrace{(a-b+1)\Gamma(a-b+1)}_{\Gamma(a-b+2)}} + \frac{b\Gamma(a+1)}{\underbrace{b\Gamma(b)}_{\Gamma(b+1)}\Gamma(a-b+2)} \\ &= \frac{(a-b+1+b)\Gamma(a+1)}{\Gamma(b+1)\Gamma(a-b+2)} = \frac{\Gamma(a+2)}{\Gamma(b+1)\Gamma(a-b+2)} = \binom{a+1}{b} \end{aligned}$$

and, from (35.24),

$$\begin{aligned}
 \binom{a}{b} + \binom{a}{b-1} &= \frac{(-1)^b \Gamma(b-a)}{\Gamma(b+1)\Gamma(-a)} + \frac{(-1)^{b-1} \Gamma(b-1-a)}{\Gamma(b)\Gamma(-a)} \\
 &= \frac{(-1)^b \overbrace{(b-a-1)\Gamma(b-a-1)}^{\Gamma(b-a)}}{\Gamma(b+1)\Gamma(-a)} + \frac{-b(-1)^b \Gamma(b-1-a)}{\underbrace{b\Gamma(b)}_{\Gamma(b+1)} \Gamma(-a)} \\
 &= \frac{(-1)^b (b-a-1-b)\Gamma(b-a-1)}{\Gamma(b+1)\Gamma(-a)} = \frac{(-1)^b \Gamma(b-a-1)}{\Gamma(b+1)\Gamma(-a-1)} \\
 &= \binom{a+1}{b} \tag{35.31}
 \end{aligned}$$

It is easy to see that in (35.29) it is impossible to use simultaneously two different branches of (35.28). \square

Theorem 35.6.

$$\binom{a}{0} = 1 \tag{35.32}$$

$$\binom{a}{a} = 1 \tag{35.33}$$

for all $a \in \mathbb{R}$ for which the combinations above exist.

Proof. (35.32) is obtained from (35.23) because $\frac{\Gamma(a+1)}{\Gamma(1)\Gamma(a+1)} = 1$, and from (35.24) because $\frac{(-1)^0 \Gamma(-a)}{\Gamma(1)\Gamma(-a)} = 1$; the last branch of (35.28) never applies.

(35.33) is obtained from (35.23) because $\frac{\Gamma(a+1)}{\Gamma(a+1)\Gamma(a-a+1)} = 1$; the last two branches of (35.28) never apply. \square

35.2 Two apparently simple examples

Since fractional derivatives generalise to arbitrary orders both derivatives and integrals, it is expedient to denote them in a same way. So let us introduce a functional operator D , associated to an order $n \in \mathbb{Z}$, as follows:

Operator D

Definition 35.4.

$${}_c D_t^n f(t) = \begin{cases} \frac{d^n f(t)}{dt^n}, & \text{if } n \in \mathbb{N} \\ f(t), & \text{if } n = 0 \\ \int_c^t {}_c D_t^{n+1} f(t) dt, & \text{if } n \in \mathbb{Z}^- \end{cases} \tag{35.34}$$

$${}_t D_c^n f(t) = \begin{cases} (-1)^n \frac{d^n f(t)}{dt^n}, & \text{if } n \in \mathbb{N} \\ f(t), & \text{if } n = 0 \\ \int_t^c {}_t D_c^{n+1} f(t) dt, & \text{if } n \in \mathbb{Z}^- \end{cases} \tag{35.35}$$

Remark 35.2. The recursion used to define the $n \in \mathbb{Z}^-$ branches means that

$${}_c D_x^n f(t) = \underbrace{\int_c^x \cdots \int_c^x f(t) dt \cdots dt}_{|n| \text{ integrations}}, \quad n \in \mathbb{Z}^- \tag{35.36}$$

$${}_x D_c^n f(t) = \underbrace{\int_x^c \cdots \int_x^c f(t) dt \cdots dt}_{|n| \text{ integrations}}, \quad n \in \mathbb{Z}^- \tag{35.37}$$

Remark 35.3. When $n \in \mathbb{Z}_0^+$, operator D^n is local, and hence subscripts c and t are useless. Thus, for instance, ${}_0 D_t^2 f(t) = {}_{-\infty} D_t^2 f(t)$. But, when $n \in \mathbb{Z}^-$, the operator is no longer local; changing the value of c will, in general, change the result. \square

When D is local or non-local

Of course, $D^n f(t)$, $n \in \mathbb{N}$ only makes sense if f is n times differentiable, and ${}_c D_t^{-n} f(t)$, $n \in \mathbb{N}$ only makes sense if f is n times integrable.

Thanks to D , two cases in which it seems obvious what a fractional derivative should be can be easily presented.

Derivatives of $e^{\lambda t}$

Example 35.1. From

$${}_{-\infty} D_t^n f(t) = \frac{e^{\lambda t}}{\lambda^{-n}} = \lambda^n e^{\lambda t}, \quad n \in \mathbb{Z}^- \quad (35.38)$$

⋮

$$\int_{-\infty}^t \int_{-\infty}^t \int_{-\infty}^t f(t) dt dt dt = \frac{e^{\lambda t}}{\lambda^3} \quad (35.39)$$

$$\int_{-\infty}^t \int_{-\infty}^t f(t) dt dt = \frac{e^{\lambda t}}{\lambda^2} \quad (35.40)$$

$$\int_{-\infty}^t f(t) dt = \frac{e^{\lambda t}}{\lambda} \quad (35.41)$$

$$f(t) = e^{\lambda t}, \lambda \neq 0 \quad (35.42)$$

$$\frac{df(t)}{dt} = \lambda e^{\lambda t} \quad (35.43)$$

$$\frac{d^2 f(t)}{dt^2} = \lambda^2 e^{\lambda t} \quad (35.44)$$

$$\frac{d^3 f(t)}{dt^3} = \lambda^3 e^{\lambda t} \quad (35.45)$$

⋮

$$D^n f(t) = \lambda^n e^{\lambda t}, \quad n \in \mathbb{N} \quad (35.46)$$

we are tempted to write

$${}_{-\infty} D_t^\alpha e^{\lambda t} = \lambda^\alpha e^{\lambda t}, \lambda > 0 \quad (35.47)$$

even when $\alpha \notin \mathbb{Z}$. Notice that λ should now be positive to prevent the appearance of complex quantities. \square

Derivatives of t^λ

Example 35.2. From

$${}_0 D_t^n g(t) = \frac{t^{\lambda-n}}{-n-1} = \frac{\Gamma(\lambda+1)}{\Gamma(\lambda-n+1)} t^{\lambda-n}, \quad n \in \mathbb{Z}^- \quad (35.48)$$

⋮

$$\int_0^t \int_0^t \int_0^t g(t) dt dt dt = \frac{t^{\lambda+3}}{(\lambda+1)(\lambda+2)(\lambda+3)} \quad (35.49)$$

$$\int_0^t \int_0^t g(t) dt dt = \frac{t^{\lambda+2}}{(\lambda+1)(\lambda+2)} \quad (35.50)$$

$$\int_0^t g(t) dt = \frac{t^{\lambda+1}}{\lambda+1} \quad (35.51)$$

$$g(t) = t^\lambda, \quad t \in \mathbb{R}^+, \lambda \notin \mathbb{Z}^- \quad (35.52)$$

$$\frac{dg(t)}{dt} = \lambda t^{\lambda-1} \quad (35.53)$$

$$\frac{d^2 g(t)}{dt^2} = \lambda(\lambda-1)t^{\lambda-2} \quad (35.54)$$

$$\frac{d^3 g(t)}{dt^3} = \lambda(\lambda-1)(\lambda-2)t^{\lambda-3} \quad (35.55)$$

⋮

$$D^n g(t) = t^{\lambda-n} \prod_{k=0}^{n-1} (\lambda-k) = \frac{\Gamma(\lambda+1)}{\Gamma(\lambda-n+1)} t^{\lambda-n}, \quad n \in \mathbb{N}, n \leq \lambda \quad (35.56)$$

(where the last equalities in (35.48) and (35.56) are a result of (35.20)–(35.21)) we are tempted to write

$${}_0D_t^\alpha t^\lambda = \frac{\Gamma(\lambda + 1)}{\Gamma(\lambda - \alpha + 1)} t^{\lambda - \alpha}, \quad t \in \mathbb{R}^+, \lambda \notin \mathbb{Z} \quad (35.57)$$

also when $\alpha \notin \mathbb{Z}$. □

Remark 35.4. Notice how in these two examples of apparently simple generalisation of D to non-integer orders the lower limit of integration is in one case $c = -\infty$ and in the other $c = 0$. □

35.3 The Grünwald-Letnikoff definition of fractional derivatives

To generalise derivatives to non-integer orders, let us first remember what they are for integer ones.

Definition 35.5. The derivative of function $f(t)$ is given by

Definition of derivative

$$D^1 f(t) = \frac{df(t)}{dt} = \lim_{h \rightarrow 0} \frac{f(t) - f(t - h)}{h} \quad (35.58)$$

Notice that h in (35.58) can be positive or negative. Restricting h to positive values, we will have the left-side derivative; restricting h to negative values, we will have the right-side derivative. If f is differentiable at t , the left-side and right-side derivatives coincide. □

Theorem 35.7. The derivative of order $n \in \mathbb{N}$ of function $f(t)$ is given by

Derivative of order n

$$D^n f(t) = \frac{d^n f(t)}{dt^n} = \lim_{h \rightarrow 0} \frac{\sum_{k=0}^n (-1)^k \binom{n}{k} f(t - kh)}{h^n} \quad (35.59)$$

Proof. This is proved by mathematical induction. For $n = 1$, (35.59) becomes

$$D^1 f(t) = \lim_{h \rightarrow 0} \frac{(-1)^0 \binom{1}{0} f(t - 0) + (-1)^1 \binom{1}{1} f(t - h)}{h^1} \quad (35.60)$$

which is equal to (35.58). The inductive step is proved as follows:

$$\begin{aligned}
DD^n f(t) &= \lim_{h \rightarrow 0} \frac{\sum_{k=0}^n (-1)^k \binom{n}{k} f(t - kh) - \sum_{k=0}^n (-1)^k \binom{n}{k} f(t - kh - h)}{h^n} \\
&= \lim_{h \rightarrow 0} \frac{\sum_{k=0}^n (-1)^k \binom{n}{k} f(t - kh) - \sum_{k=1}^{n+1} (-1)^{k-1} \binom{n}{k-1} f(t - kh)}{h^{n+1}} \\
&= \lim_{h \rightarrow 0} \frac{\overbrace{(-1)^0 \binom{n}{0} f(t - 0h)}^1 + \sum_{k=1}^n (-1)^k \binom{n}{k} f(t - kh) + \sum_{k=1}^n (-1)^k \binom{n}{k-1} f(t - kh) + \underbrace{(-1)^{n+1} \binom{n}{n} f(t - (n+1)h)}_1}{h^{n+1}} \\
&= \lim_{h \rightarrow 0} \frac{\overbrace{(-1)^0 \binom{n+1}{0} f(t - 0h)}^1 + \sum_{k=1}^n (-1)^k \binom{n+1}{k} f(t - kh) + \underbrace{(-1)^{n+1} \binom{n+1}{n+1} f(t - (n+1)h)}_1}{h^{n+1}} \\
&= \lim_{h \rightarrow 0} \frac{\sum_{k=0}^{n+1} (-1)^k \binom{n+1}{k} f(t - kh)}{h^{n+1}} \tag{35.61}
\end{aligned}$$

where we used (35.29) and (35.32)–(35.33). \square

Since (35.28) generalises combinations for non-integer arguments, we are easily tempted to generalise (35.59) for an order $\alpha \in \mathbb{R}$ as

$$D^\alpha f(t) = \lim_{h \rightarrow 0} \frac{\sum_{k=0}^? (-1)^k \binom{\alpha}{k} f(t - kh)}{h^\alpha} \tag{35.62}$$

and all that is left is to know what the upper limit of the summation should be. In fact, for integer orders,

- the upper limit in (35.59) is n ;
- it might as well be $+\infty$ that the definition would not change. All the extra terms would be zero, thanks to the first branch of (35.28) and to (35.10).

For non-integer orders, it is unclear what it should be, but, since we want $D^{-1}f(t) = \int f(t) dt$, we make $\alpha = -1$ in (35.62); then, using (35.24) we get

$$\begin{aligned}
D^{-1}f(t) &= \lim_{h \rightarrow 0} \frac{\sum_{k=0}^? (-1)^k \binom{-1}{k} f(t - kh)}{h^{-1}} \\
&= \lim_{h \rightarrow 0} h \sum_{k=0}^? (-1)^k \frac{(-1)^k \Gamma(k+1)}{\Gamma(k+1)\Gamma(1)} f(t - kh) \tag{35.63}
\end{aligned}$$

If this is to be a Riemann integral $\int_c^t f(t) dt = {}_c D_t^{-1} f(t)$, the upper limit of the summation must be $\lfloor \frac{t-c}{h} \rfloor$, and h must be restricted to positive values. Thus, we arrive at

Grünwald-Letnikov defini-
of fractional deriva-

Definition 35.6.

$${}_c D_t^\alpha f(t) = \lim_{h \rightarrow 0^+} \frac{\sum_{k=0}^{\lfloor \frac{t-c}{h} \rfloor} (-1)^k \binom{\alpha}{k} f(t - kh)}{h^\alpha} \quad \square \quad (35.64)$$

Remark 35.5. If we make $\alpha = 0$ in (35.64) or in (35.68), we get

$${}_c D_t^0 f(t) = \lim_{h \rightarrow 0^+} \frac{\overbrace{(-1)^0}^1 \overbrace{\binom{0}{0}}^1 f(t-0) + \underbrace{\sum_{k=1}^{\lfloor \frac{t-c}{h} \rfloor} (-1)^k \binom{0}{k}}_{h^0} f(t - kh)}{1} = f(t) \quad (35.65)$$

and so we will obtain $f(t)$ back. □

Remark 35.6. Only when $\alpha \in \mathbb{N}$ does the summation have a finite number of terms. In other words, $\alpha \in \mathbb{N}$ is the only case in which (35.64) does not depend on c . That is to say, D is a non-local operator (depending on what happens to $f(t)$ between the integration limits c and t), except for the case of natural order derivatives ($\frac{d}{dt}$, $\frac{d^2}{dt^2}$, $\frac{d^3}{dt^3}$ and so on) and the case $\alpha = 0$. In this respect, fractional derivatives look like the integrals, not the derivatives, we are used to from Calculus; and this irrespective of the sign of α . □

D is non-local in the general case

But D is local for positive integer orders

Theorem 35.8. D is a linear operator:

D is a linear operator

$${}_c D_t^\alpha [af(t) + bg(t)] = a {}_c D_t^\alpha f(t) + b {}_c D_t^\alpha g(t), \quad a, b \in \mathbb{R} \quad (35.66)$$

Proof.

$$\begin{aligned} {}_c D_t^\alpha (af(t) + bg(t)) &= \lim_{h \rightarrow 0^+} \frac{\sum_{k=0}^{\lfloor \frac{t-c}{h} \rfloor} (-1)^k \binom{\alpha}{k} (af(t) + bg(t))}{h^\alpha} \\ &= a \lim_{h \rightarrow 0^+} \frac{\sum_{k=0}^{\lfloor \frac{t-c}{h} \rfloor} (-1)^k \binom{\alpha}{k} f(t)}{h^\alpha} + b \lim_{h \rightarrow 0^+} \frac{\sum_{k=0}^{\lfloor \frac{t-c}{h} \rfloor} (-1)^k \binom{\alpha}{k} g(t)}{h^\alpha} \\ &= a {}_c D_t^\alpha f(t) + b {}_c D_t^\alpha g(t) \end{aligned} \quad (35.67)$$

□

Remark 35.7. If the order of the terminals is reversed, as in (35.35), then (35.64) is replaced by

$${}_t D_c^\alpha f(t) = \lim_{h \rightarrow 0^+} \frac{\sum_{k=0}^{\lfloor \frac{c-t}{h} \rfloor} (-1)^k \binom{\alpha}{k} f(t + kh)}{h^\alpha} \quad \square \quad (35.68)$$

35.4 The Riemann-Liouville definition of fractional derivatives

While the Grünwald-Letnikov definition is the most straightforward definition of fractional derivatives, there is an alternative definition which is sometimes useful. It is based upon two results for integer order derivatives that are extended for the non-integer case:

Theorem 35.9. If all the derivatives exist, the equality

Law of exponents

$${}_c D_t^m {}_c D_t^n f(t) = {}_c D_t^{m+n} f(t) \quad (35.69)$$

holds in each of the three following cases:

$$m, n \in \mathbb{Z}_0^+ \quad (35.70)$$

$$m, n \in \mathbb{Z}_0^- \quad (35.71)$$

$$m \in \mathbb{Z}^+ \wedge n \in \mathbb{Z}^- \quad (35.72)$$

Proof. The first two cases wherein (35.69) holds are obvious consequences of definition (35.34). The third case can be easily proven by mathematical induction from the fact that differentiation is the left inverse operator of integration, that is to say, $D^1 {}_c D_t^{-1} f(t) = f(t)$. \square

Remark 35.8. If none of (35.70)–(35.72) holds, that is to say, if $m \in \mathbb{Z}^- \wedge n \in \mathbb{Z}^+$, then, when calculating ${}_c D_t^m {}_c D_t^n f(t)$, integration constants appear, related to initial conditions at $t = c$. This means that (35.69) will only hold if all the integration constants are equal to zero. \square

Cauchy's formula

Theorem 35.10. The indefinite integral of order $n \in \mathbb{N}$ of function $f(t)$ is given by

$${}_c D_x^{-n} f(t) = \overbrace{\int_c^x \cdots \int_c^x f(t) dt \cdots dt}^{n \text{ integrations}} = \int_c^x \frac{(x-t)^{n-1}}{(n-1)!} f(t) dt \quad (35.73)$$

$${}_x D_c^{-n} f(t) = \underbrace{\int_x^c \cdots \int_x^c f(t) dt \cdots dt}_{n \text{ integrations}} = \int_x^c \frac{(t-x)^{n-1}}{(n-1)!} f(t) dt \quad (35.74)$$

Proof. For $n = 1$, both (35.73) and (35.74) are trivial. The proof proceeds by mathematical induction and is based upon Dirichlet's equality for a function of two variables x_1 and x_2 :

$$\int_c^x \int_c^{x_1} f(x_1, x_2) dx_2 dx_1 = \int_c^x \int_{x_2}^x f(x_1, x_2) dx_1 dx_2 \quad (35.75)$$

If f does not depend on x_1 , but on x_2 alone (see the integration area in figure 35.2),

$$\begin{aligned} \int_c^x \int_c^{x_1} f(x_2) dx_2 dx_1 &= \int_c^x \int_{x_2}^x f(x_2) dx_1 dx_2 \\ &= \int_c^x f(x_2) \int_{x_2}^x dx_1 dx_2 = \int_c^x f(x_2)(x-x_2) dx_2 \end{aligned} \quad (35.76)$$

Clearly, this is a particular case of (35.73), when $n = 2$. The inductive step is proved applying it to (35.73):

$$\begin{aligned} \int_c^x {}_c D_x^{-n} f(t) dx &= \int_c^x \int_c^x \frac{(x-t)^{n-1}}{(n-1)!} f(t) dt dx \\ &= \int_c^x \int_t^x \frac{(x-t)^{n-1}}{(n-1)!} f(t) dx dt \\ &= \int_c^x f(t) \int_t^x \frac{(x-t)^{n-1}}{(n-1)!} dx dt \\ &= \int_c^x f(t) \left[\frac{(x-t)^n}{n!} \right]_{x=t}^x dt \\ &= \int_c^x f(t) \frac{(x-t)^n}{n!} dt \end{aligned} \quad (35.77)$$

When the indefinite limit of integration comes first, (35.76) becomes (see the integration area in figure 35.2)

$$\begin{aligned} \int_x^c \int_{x_1}^c f(x_2) dx_2 dx_1 &= \int_x^c \int_x^{x_2} f(x_2) dx_1 dx_2 \\ &= \int_x^c f(x_2) \int_x^{x_2} dx_1 dx_2 = \int_x^c f(x_2)(x_2-x) dx_2 \end{aligned} \quad (35.78)$$

Clearly, this is a particular case of (35.74), when $n = 2$. The inductive step is proved applying it to (35.74):

$$\begin{aligned}
 \int_x^c {}_x D_c^{-n} f(t) dx &= \int_x^c \int_x^c \frac{(t-x)^{n-1}}{(n-1)!} f(t) dt dx \\
 &= \int_x^c \int_x^t \frac{(t-x)^{n-1}}{(n-1)!} f(t) dx dt \\
 &= \int_x^c f(t) \int_x^t \frac{(t-x)^{n-1}}{(n-1)!} dx dt \\
 &= \int_x^c f(t) \left[-\frac{(t-x)^n}{n!} \right]_{x=x}^t dt \\
 &= \int_x^c f(t) \frac{(t-x)^n}{n!} dt \quad \square \tag{35.79}
 \end{aligned}$$

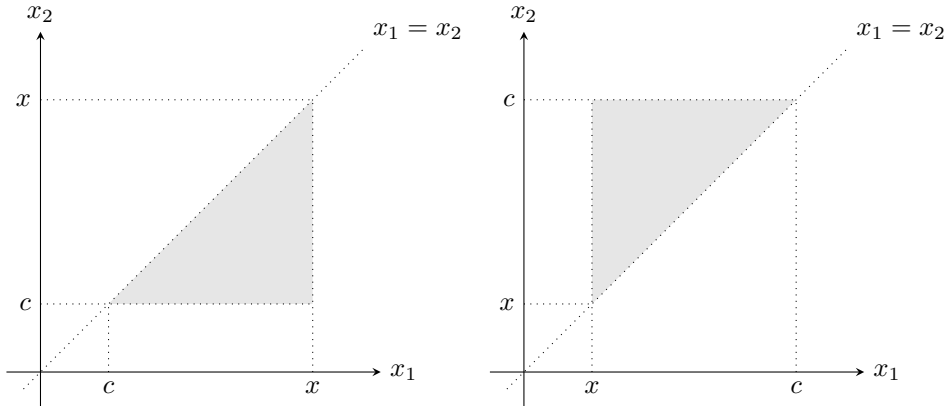


Figure 35.2: Left: integration area of (35.76); right: integration area of (35.78).

Generalising the law of exponents (35.69) for positive values of α , and Cauchy's formula (35.73)–(35.74) for negative values of α , we arrive at the following:

Definition 35.7 (Riemann-Liouville fractional derivatives).

$${}_c D_t^\alpha f(t) = \begin{cases} \int_c^t \frac{(t-\tau)^{-\alpha-1}}{\Gamma(-\alpha)} f(\tau) d\tau, & \text{if } \alpha \in \mathbb{R}^- \\ f(t), & \text{if } \alpha = 0 \\ \frac{d^{[\alpha]}}{dt^{[\alpha]}} {}_c D_t^{\alpha-[\alpha]} f(t), & \text{if } \alpha \in \mathbb{R}^+ \end{cases} \tag{35.80}$$

$${}_t D_c^\alpha f(t) = \begin{cases} \int_t^c \frac{(\tau-t)^{-\alpha-1}}{\Gamma(-\alpha)} f(\tau) d\tau, & \text{if } \alpha \in \mathbb{R}^- \\ f(t), & \text{if } \alpha = 0 \\ (-1)^{[\alpha]} \frac{d^{[\alpha]}}{dt^{[\alpha]}} {}_t D_c^{\alpha-[\alpha]} f(t), & \text{if } \alpha \in \mathbb{R}^+ \end{cases} \quad \square \tag{35.81}$$

Remark 35.9. Notice that, if $\alpha \in \mathbb{Z}$, (35.80) reduces to (35.34) and (35.81) reduces to (35.35); in particular, if $\alpha \in \mathbb{N}$, we will have

$${}_c D_t^\alpha f(t) = \frac{d^\alpha f(t)}{dt^\alpha} \tag{35.82}$$

$${}_t D_c^\alpha f(t) = (-1)^\alpha \frac{d^\alpha f(t)}{dt^\alpha} \tag{35.83}$$

The Riemann-Liouville and Grünwald-Letnikov definitions of fractional derivatives give the same result, provided that the function satisfies the conditions for the application of *both* the definitions. We will not prove this result, which is very difficult to arrive at. You can, however, implement numerically both definitions, and check that the results are the same (up to the numerical error resulting from the approximations, of course).

Theorem 35.11. If $f(t)$ has $\max\{0, [\alpha]\}$ continuous derivatives, and $D^{\max\{0, [\alpha]\}} f(t)$ is integrable, then ${}_c D_t^\alpha f(t)$ exists according to both the Riemann-Liouville and Grünwald-Letnikov definitions, which provide the same result. \square

35.5 Properties of fractional derivatives

It is because of the following result that differential equations with fractional derivatives originate fractional transfer functions.

$\mathcal{L}[D^\alpha f(t)]$

Theorem 35.12. The Laplace transform of D is

$$\mathcal{L}[{}_0D_t^\alpha f(t)] = \begin{cases} s^\alpha F(s), & \text{if } \alpha \in \mathbb{R}^- \\ F(s), & \text{if } \alpha = 0 \\ s^\alpha F(s) - \sum_{k=0}^{[\alpha]-1} s^k {}_0D_t^{\alpha-k-1} f(0), & \text{if } \alpha \in \mathbb{R}^+ \end{cases} \quad (35.84)$$

Proof. This is easily proved using the Riemann-Liouville definition. The result is trivial for $\alpha = 0$. For $\alpha < 0$,

$$\mathcal{L}[{}_0D_t^\alpha f(t)] = \mathcal{L}\left[\frac{1}{\Gamma(-\alpha)} \int_0^t (t-\tau)^{-\alpha-1} f(\tau) d\tau\right] \quad (35.85)$$

By (2.78) this is equal to

$$\mathcal{L}[{}_0D_t^\alpha f(t)] = \frac{1}{\Gamma(-\alpha)} \mathcal{L}[t^{-\alpha-1}] \mathcal{L}[f(t)] \quad (35.86)$$

Using the Laplace transform of the power function

$$\mathcal{L}[t^\lambda] = \frac{\Gamma(\lambda+1)}{s^{\lambda+1}}, \quad \lambda > -1 \quad (35.87)$$

this becomes

$$\mathcal{L}[{}_0D_t^\alpha f(t)] = \frac{1}{\Gamma(-\alpha)} \frac{\Gamma(-\alpha)}{s^{-\alpha}} \mathcal{L}[f(t)] = s^\alpha \mathcal{L}[f(t)] \quad (35.88)$$

For $\alpha > 0$,

$$\mathcal{L}[{}_0D_t^\alpha f(t)] = \mathcal{L}[D^{[\alpha]} {}_0D_t^{\alpha-[\alpha]} f(t)] \quad (35.89)$$

According to (2.45) and (2.47), this becomes

$$\mathcal{L}[{}_0D_t^\alpha f(t)] = s^{[\alpha]} s^{\alpha-[\alpha]} F(s) - \sum_{k=0}^{[\alpha]-1} s^k D^{[\alpha]-k-1} {}_0D_t^{\alpha-[\alpha]} f(0) \quad (35.90)$$

which is the expression in (35.84). \square

Theorem 35.13. When input

$$u(t) = A \sin(\omega t) \quad (35.91)$$

is applied to a stable fractional system $G(s)$, the output, after the transient regime has passed away, is

$$y(t) = |G(j\omega)| A \sin(\omega t + \angle G(j\omega)) \quad (35.92)$$

That is to say, the frequency response of a fractional transfer function may be found replacing s with $j\omega$, as is the case for integer transfer functions; i.e. it can be obtained evaluating $G(s)$ at the positive imaginary semiaxis, $s = j\omega$, $\omega \in \mathbb{R}^+$.

Proof. The proof is similar to that of Theorem 10.4. The differences are that stability conditions and vanishing terms in the response are as we will study below in Chapter 36. \square

How Riemann integrals are recovered with the GL definition

Let us take a final look at the Grünwald-Letnikov definition of D^α and show that, if $\alpha \in \mathbb{Z}^-$, the result is in fact equal to a Riemann integral. It is worthwhile to repeat that, though the upper limit of the summation in (35.64) is diverging to $+\infty$, if $\alpha \in \mathbb{N}$ then all terms with $k > \alpha$ will be zero, and thus (35.64) reduces to (35.59) when $h > 0$. In other words, (35.64) will be a right derivative of f . If f is differentiable, the right and left side derivatives will be equal, and no problem arises from restricting h to positive values.

As to higher order integrals, we need the following result.

Lemma 35.2.

$${}_c D_x^{-n} f(t) = \lim_{h \rightarrow 0^+} \sum_{k=0}^{\lfloor \frac{x-c}{h} \rfloor} h \frac{(kh)^{n-1}}{(n-1)!} f(x - kh) \quad (35.93)$$

Proof. Apply the definition of the Riemann integral

$$\int_c^x f(t) dt = \lim_{h \rightarrow 0^+} \sum_{k=0}^{\lfloor \frac{x-c}{h} \rfloor} h f(x - kh) \quad (35.94)$$

to (35.73). \square

We can now show using the Grünwald-Letnikov definition that ${}_c D_t^{-n} f(t)$ is in fact equal to

$$\underbrace{\int_c^x \cdots \int_c^x f(t) dt \cdots dt}_{n \text{ integrations}} \quad (35.95)$$

If we make $\alpha = -n < -1$, $n \in \mathbb{N}$ in (35.64), we get, according to (35.24) and (35.20),

$$\begin{aligned} {}_c D_t^{-n} f(t) &= \lim_{h \rightarrow 0^+} h^n \sum_{k=0}^{\lfloor \frac{t-c}{h} \rfloor} (-1)^k \frac{(-1)^k \Gamma(k+n)}{\Gamma(k+1)\Gamma(n)} f(t - kh) \\ &= \lim_{h \rightarrow 0^+} \sum_{k=0}^{\lfloor \frac{t-c}{h} \rfloor} h^n \frac{\prod_{i=1}^{n-1} (k+i)}{(n-1)!} f(t - kh) \end{aligned} \quad (35.96)$$

Whatever the value of $n \in \mathbb{N}$,

$$\prod_{i=1}^{n-1} (k+i) = k^{n-1} + \sum_{i=0}^{n-2} a_i k^i \quad (35.97)$$

where $a_i \in \mathbb{N}$. Thus (35.96) becomes

$$\begin{aligned} {}_c D_t^{-n} f(t) &= \lim_{h \rightarrow 0^+} \sum_{k=0}^{\lfloor \frac{t-c}{h} \rfloor} h \frac{(hk)^{n-1}}{(n-1)!} f(t - kh) \\ &\quad + \sum_{i=0}^{n-2} \left(\lim_{h \rightarrow 0^+} \frac{i! a_i h^{n-1-i}}{(n-1)!} \sum_{k=0}^{\lfloor \frac{t-c}{h} \rfloor} h \frac{(hk)^i}{i!} f(t - kh) \right) \end{aligned} \quad (35.98)$$

According to (35.93), the limits inside the summation with index i would be the integrals of f of order $i+1$ if it were not for the fraction $\frac{i! a_i h^{n-1-i}}{(n-1)!}$. Since $h^{n-1-i} \rightarrow 0$, these fractions converge to zero and thus the entire summation is equal to zero: all that is left is the first limit, which, again according to (35.93), is indeed the n th order integral of f , as we might expect.

35.6 Applications of fractional derivatives

Fractional derivatives have many applications in practice. They can be used for controller design, for solving physical problems (as in Exercise 3), and they can model physical systems.

Example 35.3. The stress-strain model for an elastic material is

$$\sigma(t) = E\varepsilon(t) \quad (35.99)$$

where σ is the stress, ε is the strain, and E is Young's modulus. For a viscous material,

$$\sigma(t) = \eta \frac{d\varepsilon(t)}{dt} \quad (35.100)$$

where η is viscosity.

Inbetween elasticity and viscosity lies viscoelasticity. The behaviour of the *Viscoelasticity* material may be modelled as

$$\sigma(t) = E\tau^{\frac{1}{2}} {}_0D_t^{\frac{1}{2}} \varepsilon(t) \quad (35.101)$$

where τ is a time constant. Actually the fractional order can change according to whether the viscoelastic behaviour is closer to elasticity or viscosity. \square

Diffusion

Heat diffusion and mass diffusion are also areas of application of fractional order models, as in Example 36.5 of Chapter 36.

Glossary

And behold, when ye shall come unto me, ye shall write them and shall seal them up, that no one can interpret them; for ye shall write them in a language that they cannot be read. And behold, these two stones will I give unto thee, and ye shall seal them up also, with the things which ye shall write. For behold, the language which ye shall write, I have confounded; wherefore I will cause in my own due time that these stones shall magnify to the eyes of men, these things which ye shall write.

Joseph SMITH Jr. (1805 — †1844), *The Book of Mormon* (1830), Book of Ether, iii 22-24

combinations of a things, b at a time combinações de a , b a b

cycloid cicloide

fractional derivative derivada fracionária

tautochrone curve curva tautócrona

Exercises

1. Consider a commensurable transfer function $G(s)$ given by

$$G(s) = \frac{\sum_{k=0}^m b_k s^{k\alpha}}{\sum_{k=0}^n a_k s^{k\alpha}} \quad (35.102)$$

and consider an integer transfer function

$$\tilde{G}(s) = \frac{\sum_{k=0}^m b_k s^k}{\sum_{k=0}^n a_k s^k} \quad (35.103)$$

built with the same numerator and denominator coefficients. Show that the frequency response $G(j\omega)$:

- (a) can be obtained calculating $\tilde{G}(j^\alpha \omega^\alpha)$;
- (b) can be obtained evaluating $\tilde{G}(s)$ over the ray $s = j^\alpha \omega^\alpha$, $\omega \in \mathbb{R}^+$, shown in figure 35.3.

2. Prove (35.87) from the definition of \mathcal{L} .

Tautochrone curve

3. Consider a mass m at rest on a frictionless curve surface defined by curve $y(x)$, as seen in figure 35.4. For convenience, the final point of the curve is set to $(x, y) = (0, 0)$. The initial height of the mass is h and it slides down to the end of the curve, under the influence of gravity acceleration g , with an ever increasing velocity V . Our objective is to find the shape

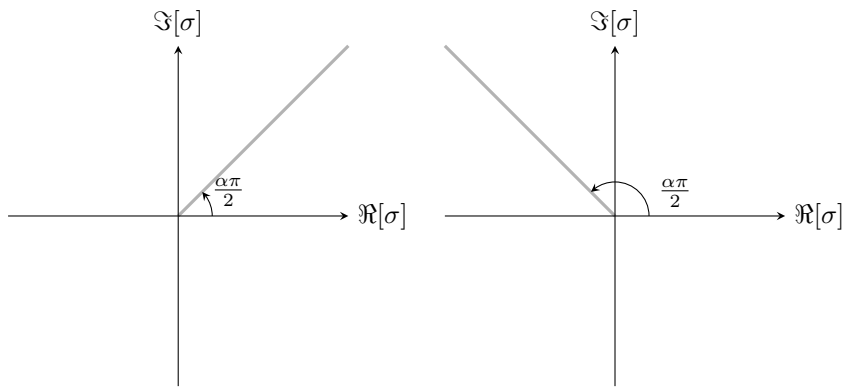


Figure 35.3: Ray where $\tilde{G}(s)$ is evaluated to obtain $G(j\omega)$; left: $0 < \alpha < 1$; right: $1 < \alpha < 2$.

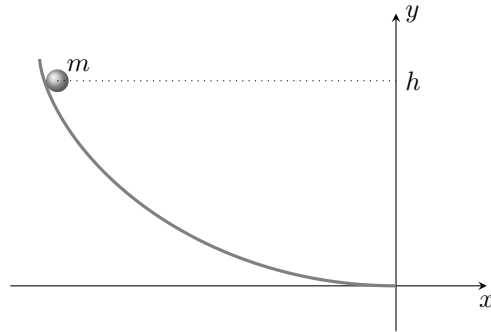


Figure 35.4: The tautochrone curve, the cycloid.

of the curve such that the time T that the mass takes to reach the bottom of the curve be constant, irrespective of the value of h . Such curve is thereby called tautochrone (from the Greek words ταὐτός, same, and χρόνος, time).

- (a) Find an expression for the total mechanical energy of the mass at any height y .
 (b) Show that

$$V = \sqrt{2g(h-y)} \quad (35.104)$$

at any height y .

- (c) Let ℓ be the distance travelled along the curve by the mass, from height y to the end of the curve, given by

$$\ell(y) = \int_0^y \sqrt{1 + \left(\frac{dx}{dy}\right)^2} dy \quad (35.105)$$

Show that $T(h)$ is given by

$$T(h) = \int_0^h \frac{1}{\sqrt{2g(h-y)}} \frac{d\ell(y)}{dy} dy \quad (35.106)$$

- (d) If the curve is tautochrone, $T(h)$ does not depend on h . Knowing that

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \quad (35.107)$$

show from the Riemann-Liouville definition of D that

$${}_0D_y^{-\frac{1}{2}} \frac{d\ell(y)}{dy} = \sqrt{\frac{2gT^2}{\pi}} \quad (35.108)$$

- (e) Apply a $\frac{1}{2}$ order derivative to both sides of the last result and show that

$$\frac{d\ell(y)}{dy} = \underbrace{\frac{\sqrt{2gT}}{\pi}}_a y^{-\frac{1}{2}} \quad (35.109)$$

Take into account that, according to the law of exponents,

$${}_0D_t^{\frac{1}{2}}{}_0D_t^{-\frac{1}{2}}f(t) = f(t) \quad (35.110)$$

(f) From this last result, show that

$$x(y) = \int_0^y \sqrt{\frac{a^2}{v} - 1} \, dv \quad (35.111)$$

(g) Show that this last result is equivalent to parametrisation

$$x(\theta) = r\theta + r \sin \theta \quad (35.112)$$

$$y(\theta) = r - r \cos \theta \quad (35.113)$$

which is that of a cycloid, the curve described by a circle rolling on a horizontal straight line, with radius $r = \frac{a^2}{2} = \frac{gT^2}{\pi^2}$.

Chapter 36

Time responses of fractional order systems

After studying the frequency responses of fractional order transfer functions, and the fractional derivatives that originate them, this chapter addresses their responses in time.

36.1 The Mittag-Leffler function

Just as we needed to generalise the factorial with the Gamma function, we now need to generalise the exponential function with a more general function, called Mittag-Leffler function.

Definition 36.1 (Mittag-Leffler functions). The one-parameter and the two-parameter Mittag-Leffler functions are defined as

$$E_{\alpha}(t) = \sum_{k=0}^{+\infty} \frac{t^k}{\Gamma(\alpha k + 1)} = E_{\alpha,1}(t), \quad \alpha > 0 \quad (36.1)$$

$$E_{\alpha,\beta}(t) = \sum_{k=0}^{+\infty} \frac{t^k}{\Gamma(\alpha k + \beta)}, \quad \alpha, \beta > 0 \quad (36.2)$$

The two-parameter Mittag-Leffler function will be referred to below simply as the Mittag-Leffler function. \square

Some particular values of these functions include

$$E_1(t) = E_{1,1}(t) = \sum_{k=0}^{+\infty} \frac{t^k}{\Gamma(k+1)} = \sum_{k=0}^{+\infty} \frac{t^k}{k!} = e^t \quad (36.3)$$

$$E_1(at) = E_{1,1}(at) = e^{at} \quad (36.4)$$

$$E_2(t) = E_{2,1}(t^2) = \sum_{k=0}^{+\infty} \frac{t^{2k}}{\Gamma(2k+1)} = \sum_{k=0}^{+\infty} \frac{t^{2k}}{(2k)!} = \cosh(t) \quad (36.5)$$

$$E_{2,2}(t^2) = \sum_{k=0}^{+\infty} \frac{t^{2k}}{\Gamma(2k+2)} = \frac{1}{t} \sum_{k=0}^{+\infty} \frac{t^{2k+1}}{(2k+1)!} = \frac{\sinh(t)}{t} \quad (36.6)$$

$$t^{\beta-1} E_{1,\beta}(0) = t^{\beta-1} \sum_{k=0}^{+\infty} \frac{0^k}{\Gamma(k+\beta)} = \frac{t^{\beta-1}}{\Gamma(\beta)} \quad (36.7)$$

Lemma 36.1. The (integer) derivatives of $\frac{1}{1 \mp t}$ are given by

$$D^k \frac{1}{1 \mp t} = \frac{k!(\pm 1)^k}{(1 \mp t)^{k+1}}, \quad k \in \mathbb{Z}_0^+ \quad (36.8)$$

Proof. This is proved by mathematical induction. For $k = 0$ the equality is obvious, and the inductive step is proved as follows:

$$\frac{d}{dt} \frac{k!(\pm 1)^k}{(1 \mp t)^{k+1}} = (-k-1)k!(\pm 1)^k (1 \mp t)^{-k-2} = \frac{(k+1)!(\pm 1)^{k+1}}{(1 \mp t)^{k+2}} \square \quad (36.9)$$

Corollary 36.1. The MacLaurin series of $\frac{1}{1 \mp t}$ is $\sum_{k=0}^{+\infty} (\pm t)^k$.

Proof. When $t = 0$ the derivative of order k reduces to $k!(\pm 1)^k$, and replacing this in

$$f(t) = \sum_{k=0}^{+\infty} \frac{t^k}{k!} \frac{d^k}{dt^k} f(0) \quad (36.10)$$

we obtain

$$\sum_{k=0}^{+\infty} \frac{k!}{k!} (\pm 1)^k t^k \quad (36.11)$$

The series converges for $|t| < 1$; otherwise its terms do not converge to 0 when $k \rightarrow +\infty$. \square

Theorem 36.1. The Laplace transform of $t^{\alpha k + \beta - 1} \frac{d^k E_{\alpha, \beta}(\pm at^\alpha)}{d(\pm at^\alpha)^k}$, $k \in \mathbb{Z}_0^+$ is

$$\mathcal{L} \left[t^{\alpha k + \beta - 1} \frac{d^k E_{\alpha, \beta}(\pm at^\alpha)}{d(\pm at^\alpha)^k} \right] = \frac{k! s^{\alpha - \beta}}{(s^\alpha \mp a)^{k+1}} \quad (36.12)$$

Proof. First we notice that

$$\begin{aligned} \int_0^{+\infty} e^{-t} t^{\beta-1} E_{\alpha, \beta}(\pm z t^\alpha) dt &= \int_0^{+\infty} e^{-t} t^{\beta-1} \sum_{k=0}^{+\infty} \frac{(\pm z)^k t^{\alpha k}}{\Gamma(\alpha k + \beta)} dt \\ &= \sum_{k=0}^{+\infty} \frac{(\pm z)^k}{\Gamma(\alpha k + \beta)} \underbrace{\int_0^{+\infty} e^{-t} t^{\alpha k + \beta - 1} dt}_{\Gamma(\alpha k + \beta)} \\ &= \frac{1}{1 \mp z} \end{aligned} \quad (36.13)$$

Differentiating the rightmost and the leftmost members of (36.13) $k \in \mathbb{Z}_0^+$ times,

$$\begin{aligned} \frac{k!(\pm 1)^k}{(1 \mp z)^{k+1}} &= \frac{d^k}{dz^k} \int_0^{+\infty} e^{-t} t^{\beta-1} E_{\alpha, \beta}(\pm z t^\alpha) dt \\ &= \int_0^{+\infty} e^{-t} t^{\beta-1} (\pm t^\alpha)^k \frac{d^k}{d(\pm z t^\alpha)^k} E_{\alpha, \beta}(\pm z t^\alpha) dt \end{aligned} \quad (36.14)$$

We now replace t with st (and thus dt with $s dt$) and get

$$\frac{k!(\pm 1)^k}{(1 \mp z)^{k+1}} = \int_0^{+\infty} e^{-st} s^{\beta-1} t^{\beta-1} (\pm 1)^k s^{\alpha k} t^{\alpha k} \frac{d^k E_{\alpha, \beta}(\pm z s^\alpha t^\alpha)}{d(\pm z s^\alpha t^\alpha)^k} s dt \quad (36.15)$$

Rearranging the terms and replacing zs^α with a (and thus z with $\frac{a}{s^\alpha}$),

$$\frac{k!}{s^\beta s^{\alpha k} (1 \mp \frac{a}{s^\alpha})^{k+1}} = \int_0^{+\infty} e^{-st} t^{\alpha k + \beta - 1} \frac{d^k E_{\alpha, \beta}(\pm at^\alpha)}{d(\pm at^\alpha)^k} dt \quad (36.16)$$

By comparison with (2.1), the right hand member can be seen to be the Laplace transform in (36.12). The left hand member is $\frac{k! s^{-\beta} s^\alpha}{s^{\alpha(k+1)} (1 \mp \frac{a}{s^\alpha})^{k+1}} = \frac{k! s^{\alpha - \beta}}{(s^\alpha \mp a)^{k+1}}$. \square

Corollary 36.2. Making $k = 0$ in (36.12),

$$\mathcal{L} [t^{\beta-1} E_{\alpha, \beta}(\pm at^\alpha)] = \frac{s^{\alpha - \beta}}{s^\alpha \mp a} \quad \square \quad (36.17)$$

Corollary 36.3. Making $\alpha = \beta$ in (36.17),

$$\mathcal{L} [t^{\alpha-1} E_{\alpha, \alpha}(\pm at^\alpha)] = \frac{1}{s^\alpha \mp a} \quad \square \quad (36.18)$$

Corollary 36.4. Making $\alpha = 1$ in (36.17),

$$\mathcal{L} [t^{\beta-1} E_{1, \beta}(\pm at)] = \frac{s^{1-\beta}}{s \mp a} \quad \square \quad (36.19)$$

Corollary 36.5. Making $a = 0$ in (36.19),

$$\mathcal{L} [t^{\beta-1} E_{1,\beta}(0)] = \mathcal{L} \left[\frac{t^{\beta-1}}{\Gamma(\beta)} \right] = \frac{1}{s^\beta} \quad \square \quad (36.20)$$

Approximation of E for t

The inverse Laplace transforms corresponding to the relations above become easier to evaluate for $|t| \rightarrow +\infty$ using the following approximations, given here without proof:

$$E_{\alpha,\beta}(t) \approx \frac{1}{\alpha} t^{\frac{1-\beta}{\alpha}} e^{t^{1/\alpha}} \quad (36.21)$$

$$\frac{d^n E_{\alpha,\beta}(t)}{dt^n} \approx \sum_{k=0}^n \frac{a_{k,n}(\alpha, \beta)}{\alpha^{n+1}} t^{\frac{1-(2n-k)\alpha-\beta+2(n-k)}{\alpha}} e^{t^{1/\alpha}} \quad (36.22)$$

Inverse Laplace transforms can be found in the usual way.

Example 36.1. Suppose we want to calculate the inverse Laplace transform of $\frac{4s^{1/2} - 1}{s + s^{1/2} - 2}$. We apply a partial fraction expansion, and then (36.18):

$$\begin{aligned} \mathcal{L}^{-1} \left[\frac{4s^{1/2} - 1}{s + s^{1/2} - 2} \right] &= \mathcal{L}^{-1} \left[\frac{1}{s^{1/2} - 1} + \frac{3}{s^{1/2} + 2} \right] \\ &= t^{-\frac{1}{2}} E_{\frac{1}{2}, \frac{1}{2}}(t^{\frac{1}{2}}) + 3t^{-\frac{1}{2}} E_{\frac{1}{2}, \frac{1}{2}}(-2t^{\frac{1}{2}}) \end{aligned} \quad (36.23)$$

Example 36.2. Suppose we want to calculate the inverse Laplace transform of

$$\frac{s + s^{2/3} - \sqrt{2}s^{1/3} - \sqrt{2}}{s^{4/3} - \sqrt{2}s - \sqrt{2}s^{1/3} + 2} = \frac{1}{s^{1/3} - \sqrt{2}} + \frac{s^{1/3}}{s - \sqrt{2}} \quad (36.24)$$

We will solve this in two ways.

- If we are aware of the equality above, we can apply (36.18)–(36.19) to the right-hand side of (36.24) and get

$$\mathcal{L}^{-1} \left[\frac{1}{s^{1/3} - \sqrt{2}} + \frac{s^{1/3}}{s - \sqrt{2}} \right] = t^{-\frac{2}{3}} E_{\frac{1}{3}, \frac{1}{3}}(\sqrt{2}t^{\frac{1}{3}}) + t^{-\frac{1}{3}} E_{1, \frac{2}{3}}(\sqrt{2}t) \quad (36.25)$$

- But if we simply apply a partial fraction expansion to the left-hand side of (36.24), we are led to

$$\begin{aligned} \mathcal{L}^{-1} \left[\frac{s + s^{2/3} - \sqrt{2}s^{1/3} - \sqrt{2}}{s^{4/3} - \sqrt{2}s - \sqrt{2}s^{1/3} + 2} \right] &= \\ \mathcal{L}^{-1} \left[\frac{1}{s^{1/3} - \sqrt{2}} + \frac{1}{3 \times 2^{1/6}} + \frac{1}{3\sqrt{3} 2^{1/6}} \left(-\frac{\sqrt{3}}{2} - \frac{3}{2}j \right) + \frac{1}{3\sqrt{3} 2^{1/6}} \left(-\frac{\sqrt{3}}{2} + \frac{3}{2}j \right) \right] &= \\ t^{-\frac{2}{3}} E_{\frac{1}{3}, \frac{1}{3}}(\sqrt{2}t^{\frac{1}{3}}) + \frac{1}{3 \times 2^{\frac{1}{6}}} t^{-\frac{2}{3}} E_{\frac{1}{3}, \frac{1}{3}}(2^{\frac{1}{6}}t^{\frac{1}{3}}) + & \\ \frac{1}{3\sqrt{3} 2^{\frac{1}{6}}} \left(-\frac{\sqrt{3}}{2} - \frac{3}{2}j \right) t^{-\frac{2}{3}} E_{\frac{1}{3}, \frac{1}{3}} \left(2^{-\frac{5}{6}}(-1 + \sqrt{3})t^{\frac{1}{3}} \right) + & \\ \frac{1}{3\sqrt{3} 2^{\frac{1}{6}}} \left(-\frac{\sqrt{3}}{2} + \frac{3}{2}j \right) t^{-\frac{2}{3}} E_{\frac{1}{3}, \frac{1}{3}} \left(2^{-\frac{5}{6}}(-1 - \sqrt{3})t^{\frac{1}{3}} \right) & \end{aligned} \quad (36.26)$$

This far more complicated expression, however, turns out to be equal to (36.25): its imaginary parts cancel out. While it is not trivial to prove this analytically, the reader can easily check numerically that it is so. \square

36.2 Time responses of simple fractional order transfer functions

The Laplace transforms above can be used to find the following impulse, unit *Impulse, step and ramp responses*

step and unit ramp responses:

$$\mathcal{L}^{-1} \left[\frac{1}{s^\alpha} \mathcal{L} [\delta(t)] \right] = \frac{t^{\alpha-1}}{\Gamma(\alpha)} \quad (36.27)$$

$$\mathcal{L}^{-1} \left[\frac{1}{s^\alpha} \mathcal{L} [H(t)] \right] = \frac{t^\alpha}{\Gamma(\alpha+1)} \quad (36.28)$$

$$\mathcal{L}^{-1} \left[\frac{1}{s^\alpha} \mathcal{L} [t] \right] = \frac{t^{\alpha+1}}{\Gamma(\alpha+2)} \quad (36.29)$$

$$\mathcal{L}^{-1} \left[\frac{1}{s^\alpha \mp a} \mathcal{L} [\delta(t)] \right] = t^{\alpha-1} E_{\alpha,\alpha}(\pm at^\alpha) \quad (36.30)$$

$$\mathcal{L}^{-1} \left[\frac{1}{s^\alpha \mp a} \mathcal{L} [H(t)] \right] = t^\alpha E_{\alpha,\alpha+1}(\pm at^\alpha) \quad (36.31)$$

$$\mathcal{L}^{-1} \left[\frac{1}{s^\alpha \mp a} \mathcal{L} [t] \right] = t^{\alpha+1} E_{\alpha,\alpha+2}(\pm at^\alpha) \quad (36.32)$$

$$\mathcal{L}^{-1} \left[\frac{1}{(s^\alpha \mp a)^{k+1}} \mathcal{L} [\delta(t)] \right] = \frac{t^{\alpha(k+1)-1}}{\Gamma(k+1)} \frac{d^k E_{\alpha,\alpha}(\pm at^\alpha)}{d(\pm at^\alpha)^k}, \quad k \in \mathbb{Z}_0^+ \quad (36.33)$$

$$\mathcal{L}^{-1} \left[\frac{1}{(s^\alpha \mp a)^{k+1}} \mathcal{L} [H(t)] \right] = \frac{t^{\alpha(k+1)}}{\Gamma(k+1)} \frac{d^k E_{\alpha,\alpha+1}(\pm at^\alpha)}{d(\pm at^\alpha)^k}, \quad k \in \mathbb{Z}_0^+ \quad (36.34)$$

$$\mathcal{L}^{-1} \left[\frac{1}{(s^\alpha \mp a)^{k+1}} \mathcal{L} [t] \right] = \frac{t^{\alpha(k+1)+1}}{\Gamma(k+1)} \frac{d^k E_{\alpha,\alpha+2}(\pm at^\alpha)}{d(\pm at^\alpha)^k}, \quad k \in \mathbb{Z}_0^+ \quad (36.35)$$

Because of (36.21), when $t \rightarrow +\infty$, the following approximations can be used instead of (36.30)-(36.32):

$$\begin{aligned} \mathcal{L}^{-1} \left[\frac{1}{s^\alpha \mp a} \mathcal{L} [\delta(t)] \right] &\approx \frac{t^{\alpha-1}}{\alpha} (\pm at^\alpha)^{\frac{1-\alpha}{\alpha}} e^{ta^{1/\alpha}} \\ &= \frac{1}{(\pm a)^{\frac{\alpha-1}{\alpha}} \alpha} e^{ta^{1/\alpha}} \end{aligned} \quad (36.36)$$

$$\begin{aligned} \mathcal{L}^{-1} \left[\frac{1}{s^\alpha \mp a} \mathcal{L} [H(t)] \right] &\approx \frac{t^\alpha}{\alpha} \frac{1}{\pm at^\alpha} e^{ta^{1/\alpha}} \\ &= \frac{1}{\pm a \alpha} e^{ta^{1/\alpha}} \end{aligned} \quad (36.37)$$

$$\begin{aligned} \mathcal{L}^{-1} \left[\frac{1}{s^\alpha \mp a} \mathcal{L} [t] \right] &\approx \frac{t^{\alpha+1}}{\alpha} (\pm at^\alpha)^{\frac{-1-\alpha}{\alpha}} e^{ta^{1/\alpha}} \\ &= \frac{1}{(\pm a)^{\frac{\alpha+1}{\alpha}} \alpha} e^{ta^{1/\alpha}} \quad \square \end{aligned} \quad (36.38)$$

(36.22) can likewise be used to approximate (36.33)-(36.35).

Time responses of other fractional transfer functions can be found from these using the convolution theorem (10.25), normally evaluated numerically.

Example 36.3. The impulse response of $G(s) = \frac{2}{(s^{\frac{1}{2}} + 1)(s^{\frac{1}{3}} + 4)}$ can be found from

$$\mathcal{L}^{-1} \left[\frac{1}{s^{\frac{1}{2}} + 1} \right] = t^{-\frac{1}{2}} E_{\frac{1}{2}, \frac{1}{2}}(-t^{\frac{1}{2}}) \quad (36.39)$$

$$\mathcal{L}^{-1} \left[\frac{1}{s^{\frac{1}{3}} + 4} \right] = t^{-\frac{2}{3}} E_{\frac{1}{3}, \frac{1}{3}}(-4t^{\frac{1}{3}}) \quad (36.40)$$

and is equal to

$$\mathcal{L}^{-1} \left[\frac{2}{(s^{\frac{1}{2}} + 1)(s^{\frac{1}{3}} + 4)} \mathcal{L} [\delta(t)] \right] = 2 \int_0^t \frac{E_{\frac{1}{2}, \frac{1}{2}}(-\sqrt{t-\tau}) E_{\frac{1}{3}, \frac{1}{3}}(-4\sqrt[3]{\tau})}{\sqrt{t-\tau} \sqrt[3]{\tau^2}} d\tau \quad \square \quad (36.41)$$

36.3 Stability of fractional systems

Just as an integer transfer function is stable when all its poles lie in the left-hand complex half-plane, a similar condition must be verified for the stability of fractional transfer functions.

Theorem 36.2. System $G(s) = \frac{N(s)}{D(s)}$ is stable if

$$\forall s : D(s) = 0, |\angle s| > \frac{\pi}{2} \quad (36.42)$$

restricting $\angle s$ to $[-\pi, +\pi]$ rad. \square

We will not prove this theorem in the general case; only in the form it takes for the particular case of commensurable transfer functions:

Corollary 36.6. Let $\sigma_k, k = 1 \dots n$ be the roots of the polynomial

Matignon's theorem

$$A(\sigma) = \sum_{k=0}^n a_k \sigma^k \quad (36.43)$$

built with the denominator coefficients of transfer function

$$G(s) = \frac{\sum_{k=0}^m b_k s^{k\alpha}}{\sum_{k=0}^n a_k s^{k\alpha}} \quad (36.44)$$

Then $G(s)$ is stable if and only if

$$|\angle \sigma_k| > \alpha \frac{\pi}{2}, \forall k \quad (36.45)$$

restricting $\angle \sigma_k$ to $[-\pi, +\pi]$ rad.

Proof. If $A(\sigma)$ has no roots with multiplicity higher than one, $G(s)$ can be written as a partial fraction expansion:

$$G(s) = \sum_{k=1}^n \frac{\rho_k}{s^\alpha - \sigma_k} \quad (36.46)$$

Applying (36.30), it is seen that the impulse response of (36.46) is given by

$$y(t) = \sum_{k=1}^n \rho_k t^{\alpha-1} E_{\alpha,\alpha}(\sigma_k t^\alpha) \quad (36.47)$$

The asymptotic behaviour (36.21) shows that for t large enough this will become

$$y(t) \approx \sum_{k=1}^n \rho_k t^{\alpha-1} \frac{1}{\alpha} (\sigma_k t^\alpha)^{\frac{1-\alpha}{\alpha}} e^{(\sigma_k t^\alpha)^{\frac{1}{\alpha}}} = \sum_{k=1}^n \frac{\rho_k}{\alpha} \sigma_k^{\frac{1-\alpha}{\alpha}} e^{t \sigma_k^{1/\alpha}} \quad (36.48)$$

This response will tend to zero if $\Re[\sigma_k^{1/\alpha}] < 0$. Since

$$\Re[\sigma_k^{1/\alpha}] = \Re\left[(|\sigma_k| e^{j\angle \sigma_k})^{1/\alpha} \right] = \Re\left[|\sigma_k|^{1/\alpha} \left(\cos \frac{\angle \sigma_k}{\alpha} + j \sin \frac{\angle \sigma_k}{\alpha} \right) \right] \quad (36.49)$$

the condition above will be satisfied if $\cos \frac{\angle \sigma_k}{\alpha} < 0 \Leftrightarrow |\frac{\angle \sigma_k}{\alpha}| > \frac{\pi}{2}$. Since α is positive, the result follows.

If $A(\sigma)$ has roots with multiplicity higher than one, (36.46) is replaced with

$$G(s) = \sum_{k=1}^{n_d} \sum_{q=1}^{m_k} \frac{\rho_{k,q}}{(s^\alpha - \sigma_k)^q} \quad (36.50)$$

where n_d is the number of different roots and m_k is the multiplicity of root σ_k (which means that $\sum_{k=1}^{n_d} m_k = n$). Thus (36.47) is replaced by

$$y(t) = \sum_{k=1}^{n_d} \sum_{q=1}^{m_k} \rho_{k,q} \frac{t^{\alpha q-1}}{\Gamma(q)} \frac{d^{q-1} E_{\alpha,\alpha}(\sigma_k t^\alpha)}{d(\sigma_k t^\alpha)^{q-1}} \quad (36.51)$$

and, according to (36.22), the asymptotic response (36.48) will become

$$\begin{aligned}
 y(t) &\approx \sum_{k=1}^{n_d} \sum_{q=1}^{m_k} \rho_{k,q} \frac{t^{\alpha q-1}}{\Gamma(q)} \sum_{r=0}^{q-1} \frac{a_{r,q-1}(\alpha, \alpha)}{\alpha^q} (\sigma_k t^\alpha)^{\frac{1-(2q-r-3)\alpha+2(q-r-2)}{\alpha}} e^{(\sigma_k t^\alpha)^{1/\alpha}} \\
 &= \sum_{k=1}^{n_d} \sum_{q=1}^{m_k} \sum_{r=0}^{q-1} \frac{\rho_{k,q} a_{r,q-1}(\alpha, \alpha)}{\alpha^q \Gamma(q)} \sigma_k^{\frac{1-(2q-r-3)\alpha+2(q-r-2)}{\alpha}} t^{-(q-r-3)\alpha+2(q-r-2)} e^{t \sigma_k^{1/\alpha}}
 \end{aligned}
 \tag{36.52}$$

Since the exponential tends to zero faster than the power function tends to infinity, all terms will again tend to zero if $\Re[\sigma_k^{1/\alpha}] < 0$. \square

The regions where the σ may lie in the case of a stable commensurate transfer function are shown in figure 36.1.

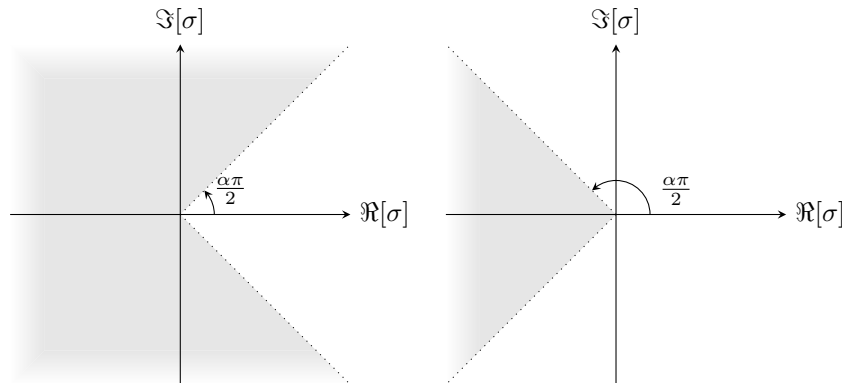


Figure 36.1: In grey: regions of the complex plane where the roots of $A(\sigma)$ may lie in the case of a stable commensurate transfer function, according to Matignon’s theorem; left: $0 < \alpha < 1$; right: $1 < \alpha < 2$.

Remark 36.1. For integer systems we reach the usual criterion of stability: all poles must have a negative real part. \square

Orders $\alpha > 2$ make plants unstable

Orders $\alpha > 2$ make the system necessarily unstable. \square

Example 36.4. Figure 36.2 shows four applications of Matignon’s theorem. \square

36.4 Identification from time responses

Identification from time responses is done as for the integer case, when using a model continuous in time. The Mittag-Leffler function cannot be inverted as easily as the exponential, which means that numerical solutions, using the Nelder-Mead simplex search method or metaheuristic methods, are the usual solution.

Example 36.5. In a few cases, analytical calculations to identify a fractional model are possible. Consider the heat equation

$$\frac{\partial T(x, t)}{\partial t} = d \frac{\partial^2 T(x, t)}{\partial x^2}
 \tag{36.53}$$

which describes the transmission of heat by conduction along one dimension x , as seen in figure 36.3 for the case of a semi-infinite body with a plane surface at $x = 0$. $T(x, t)$ is the temperature at point x and at time instant t ; d is a parameter called diffusivity. Assume a uniform initial temperature T_0 in the body at time $t = 0$ and a constant surface temperature T_{sf} . Using variable change $\theta = T - T_0$, (36.53) and the border conditions can be rewritten as

$$\frac{\partial \theta(x, t)}{\partial t} = d \frac{\partial^2 \theta(x, t)}{\partial x^2}
 \tag{36.54}$$

$$\theta(x, 0) = 0
 \tag{36.55}$$

$$\theta(0, t) = T_{sf} - T_0
 \tag{36.56}$$

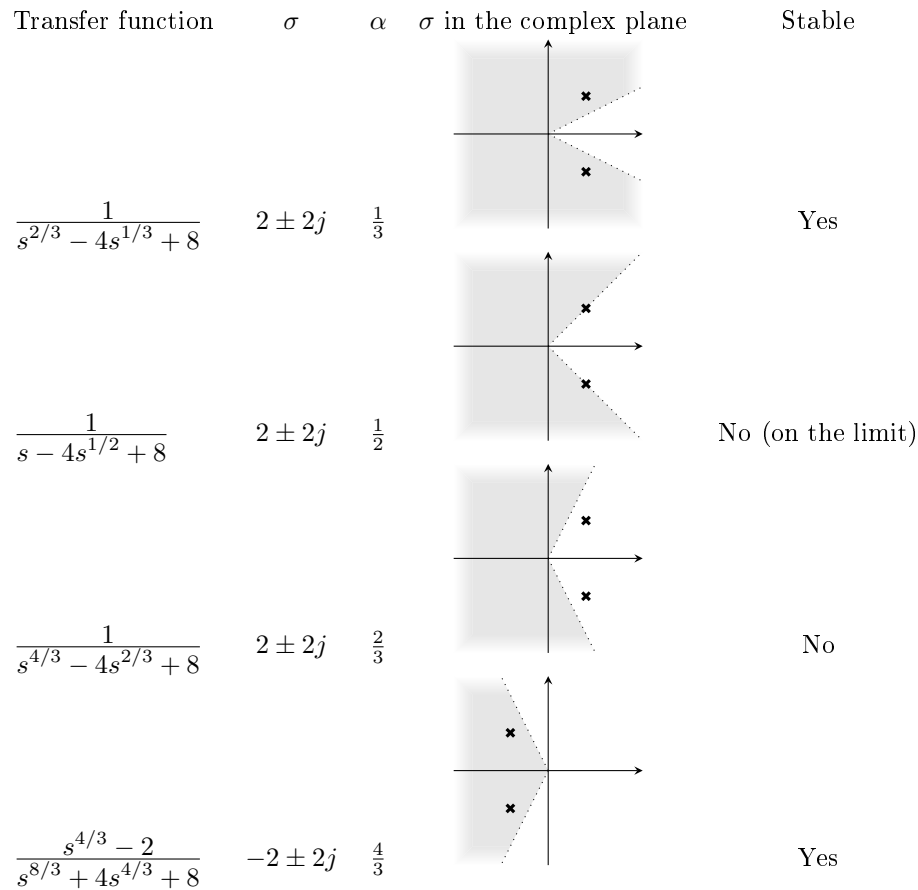


Figure 36.2: Stability of four plants verified by Matignon’s theorem

Applying the Laplace transform to the time derivative,

$$s\Theta(x, s) = d \frac{\partial^2 \Theta(x, s)}{\partial x^2} \Rightarrow \Theta = k_1 e^{x\sqrt{\frac{s}{d}}} + k_2 e^{-x\sqrt{\frac{s}{d}}} \tag{36.57}$$

for indeed

$$\frac{\partial \Theta(x, s)}{\partial x} = k_1 \sqrt{\frac{s}{d}} e^{x\sqrt{\frac{s}{d}}} - k_2 \sqrt{\frac{s}{d}} e^{-x\sqrt{\frac{s}{d}}} \tag{36.58}$$

$$\frac{\partial^2 \Theta(x, s)}{\partial x^2} = k_1 \frac{s}{d} e^{x\sqrt{\frac{s}{d}}} + k_2 \frac{s}{d} e^{-x\sqrt{\frac{s}{d}}} \tag{36.59}$$

Because $\left| \lim_{x \rightarrow -\infty} \Theta(x, s) \right| < +\infty$ (which means that the temperature inside the body cannot grow infinitely), it is necessary that $k_2 = 0$. Hence the solution must verify

$$\Theta = k_1 e^{x\sqrt{\frac{s}{d}}} \tag{36.60}$$

where $\Theta = \mathcal{L}[\theta]$.

An estimate of the age of the Earth estimate was arrived at by Lord Kelvin *Age of the Earth* from the heat equation, and redone by Heaviside using fractional derivatives in a manner similar to the following. From (36.60) we get

$$\frac{\partial \Theta(x, s)}{\partial x} = k_1 \sqrt{\frac{s}{d}} e^{x\sqrt{\frac{s}{d}}} = s^{\frac{1}{2}} \frac{1}{\sqrt{d}} \Theta(x, s) \tag{36.61}$$

Using an inverse Laplace transform (with zero initial conditions) and making $x = 0$ we conclude that the temperature gradient at the surface must verify

$$\frac{\partial \theta(0, t)}{\partial x} = \frac{1}{\sqrt{d}} {}_0 D_t^{\frac{1}{2}} \theta(0, t) \tag{36.62}$$

We now assume that the radius of the Earth is large enough for the planet to be compared to a semi-infinite solid; that its temperature, when it was formed,

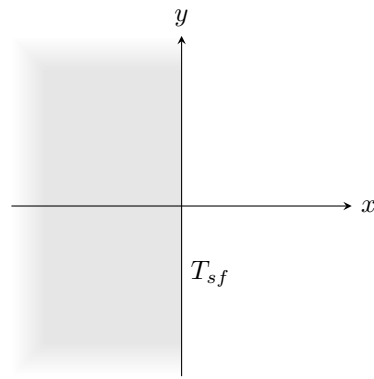


Figure 36.3: The heat conduction problem.

was $\theta(0,0) = 3900$ °C, which is the temperature of molten rock; and that its surface has a constant temperature of 0 °C. It immediately follows from (36.75), proved in Exercise 7, that an expression for the temperature gradient at the surface when the Earth was formed is

$$\frac{\partial\theta(0,0)}{\partial x} = \frac{1}{\sqrt{d}} \frac{3900 \text{ °C}}{\sqrt{\pi t}} \quad (36.63)$$

Assuming that the expression found is valid in the future, knowing that $d = 1.178 \times 10^{-6}$ m²/s and that nowadays the temperature gradient at the surface of the Earth is 1 °C for every 27.43 m (as measured in the 19th century), we can now estimate the age of the Earth: rather than making $t = 0$ in (36.63), substituting numerical values and solving in order to t we get $t = 3.0923 \times 10^{15}$ s = 98×10^6 years.

Actually, the currently accepted age for the Earth is 4.54×10^9 years: Lord Kelvin's estimate is faulty because it does not take into account heat production due to radioactive decay, unsuspected at the time. Of course, as the Earth is heated from inside, it takes longer to cool down than it otherwise would. □

Example 36.6. Fractional order models in heat conduction can be experimentally verified using an apparatus depicted in Figure 36.4, where the length of the rod is very large in comparison with its diameter. Using one such apparatus, and turning the heater on and off repeatedly, a sequence of step responses was obtained, also shown in the Figure. The response is a piecewise function; while exponentials can be fitted to the successive pieces, the fit is far more accurate using the Mittag-Leffler function with $\alpha = 0.5$. □

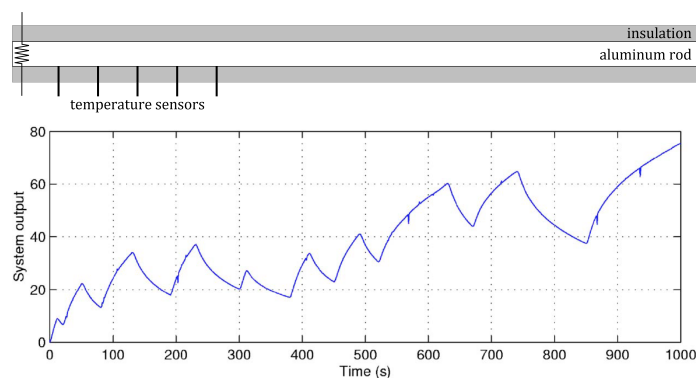


Figure 36.4: Top: experimental apparatus to measure heat conduction in a semi-infinite solid. Bottom: experimental response to a sequence of steps, obtained turning the heater on and off (source: DOI 10.1115/1.2833910).

36.5 Final comments about models of fractional order

Many dynamic systems can be modelled accurately using integer derivatives only. Fractional order models should be used as an alternative in identification only if either

- the nature of the system is such that a fractional order model can be expected (presence of diffusion, or viscoelasticity, or another phenomenon known to lead to such models); or
- the experimental data clearly corresponds to the behaviour of a fractional order plant.

Even in such cases, it is possible that a mathematically simpler integer order model can achieve enough accuracy to be preferred. Fractional order models should be chosen over integer order ones when they are simpler for a similar accuracy, or only they achieve the necessary accuracy. In this respect, fractional order models can be compared with integer order models as in Figure 36.5.



Figure 36.5: Left: integer order models. Right: fractional order models. (Source: Wikimedia. Credit for the idea in this Figure goes to Professors YangQuan Chen, Igor Podlubny and Blas Vinagre.)

Exercises

1. Use Matignon's theorem to find whether the following transfer functions are stable.

$$G_1(s) = \frac{1}{s - 6s^{1/2} + 18} \quad (36.64)$$

$$G_2(s) = \frac{1}{s^{4/3} - 6s^{2/3} + 18} \quad (36.65)$$

$$G_3(s) = \frac{s^{4/3} - 2}{s^{8/3} + 6s^{4/3} + 18} \quad (36.66)$$

$$G_4(s) = \frac{1}{s^{2/3} - 6s^{1/3} + 18} \quad (36.67)$$

Table 36.1: Unit step response of Exercise 6.

t	0	1	2	3	4	5	6	7	8	9	10
$y(t)$	0.0000	1.0000	1.4142	1.7321	2.0000	2.2361	2.4495	2.6458	2.8284	3.0000	3.1623

- Explain whether or not the plant from Exercise 4 of Chapter 34 is stable.
- Find analytically the impulse and step responses of a plant with transfer function $G(s) = \frac{10}{s^{\frac{1}{2}} + 100}$.
- Show that the impulse response of $G(s) = \frac{3}{(s^{\frac{1}{2}} + 1)(s^{\frac{1}{3}} + 11)}$ is

$$g(t) = 3 \int_0^t \frac{E_{\frac{1}{2}, \frac{1}{2}}(-\sqrt{t-\tau}) E_{\frac{1}{3}, \frac{1}{3}}(-11\sqrt[3]{\tau})}{\sqrt{t-\tau} \sqrt[3]{\tau^2}} d\tau \quad (36.68)$$

Hint: use inverse Laplace transforms and the convolution theorem.

- Find the analytic expression of the output of the plant from Exercise 3 of Chapter 34, when the input is a ramp with slope -10 .
- The unit step response of a LTI system is tabulated in Table 36.1. Sketch a plot of this unit step response, find the transfer function of this plant, and plot its Bode diagram. *Hint:* notice that the tabulated value for 2 s is $\sqrt{2}$. Can you recognise the tabulated values for 1, 4 and 9 s? What about the other time instants?
- Prove that

$${}_0D_t^{\alpha} t^{\lambda} = \frac{\Gamma(\lambda + 1)}{\Gamma(\lambda - \alpha + 1)} t^{\lambda - \alpha}, \quad \lambda > -1 \quad (36.69)$$

in the following way:

- Show from the definition of the Laplace transform that

$$\mathcal{L}[t^{\lambda}] = \frac{\Gamma(\lambda + 1)}{s^{\lambda + 1}}, \quad \lambda > -1 \quad (36.70)$$

- Show from (35.84) that, when $\alpha < 0$,

$$\mathcal{L}[{}_0D_t^{\alpha} t^{\lambda}] = s^{\alpha} \frac{\Gamma(\lambda + 1)}{s^{\lambda + 1}} = s^{\alpha - \lambda - 1} \Gamma(\lambda + 1), \quad \lambda > -1 \quad (36.71)$$

- Use (36.70) again to prove (36.69) for $\alpha < 0$.
- Use this result for $\alpha < 0$ together with the Riemann-Liouville definition (35.80) to show that, if $\alpha > 0$,

$${}_0D_t^{\alpha} t^{\lambda} = \frac{d^{[\alpha]}}{dt^{[\alpha]}} {}_0D_t^{\alpha - [\alpha]} t^{\lambda} = \frac{d^{[\alpha]}}{dt^{[\alpha]}} \frac{\Gamma(\lambda + 1)}{\Gamma(\lambda - \alpha + [\alpha] + 1)} t^{\lambda - \alpha + [\alpha]} \quad (36.72)$$

- Use that result together with (35.56) to obtain

$${}_0D_t^{\alpha} t^{\lambda} = \frac{\Gamma(\lambda + 1)}{\Gamma(\lambda - \alpha + [\alpha] + 1)} \frac{\Gamma(\lambda - \alpha + [\alpha] + 1)}{\Gamma(\lambda - \alpha + [\alpha] + 1 - [\alpha])} t^{\lambda - \alpha + [\alpha] - [\alpha]} \quad (36.73)$$

- Having proved (36.69) for both positive and negative values of α , show as a corollary that, making $\lambda = 0$,

$${}_0D_t^{\alpha} k = \frac{k}{\Gamma(1 - \alpha)} t^{-\alpha} \quad (36.74)$$

- Knowing that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, show as well, using (36.74), that

$${}_0D_t^{\frac{1}{2}} k = \frac{k}{\sqrt{t\pi}} \quad (36.75)$$

Part VIII

Stochastic systems

Or du hasard il n'est point de science.
 S'il en estoit, on auroit tort
 De l'appeler hasard, ni fortune, ni sort,
 Toutes choses tres-incertaines.

Jean de LA FONTAINE (1621 — 1695), *Fables* (1668), 2 XIII (L'astrologue qui se laisse tomber dans un puits)

In this part of the lecture notes:

Chapter 37 introduces stochastic processes and systems, and the tools to characterise them.

Chapter 38 introduces the important concept of spectral density; it is yet another tool to characterise stochastic processes and systems, but is so important that it deserves a chapter on its own.

Chapter 39 is about the identification of stochastic models.

Chapter 40 presents design methods for filters for stochastic systems.

Chapter 41 describes different types of models for digital stochastic systems and how to identify them from data.

Chapter 42 concerns the design of controllers for stochastic systems.

Here is what you need to know beforehand to follow these chapters:

- Probability and Statistics, up to the usual level of an undergraduate course on the subject;
- The Laplace and Fourier transforms, from Chapter 2;
- Transfer functions, from Sections 4.1 and 4.2 of Chapter 4;
- System theory, from Part II;
- Filters, from Sections 12.2 and 12.3 of Chapter 12;
- Discrete transfer functions, from Chapter 25, Sections 26.1 to 26.5 of Chapter 26, and Sections 27.1, 27.2 and 27.6 of Chapter 27;
- System identification, from Chapters 30 to 32.

Chapter 37

Stochastic processes and systems

Ah, the evening when I took those seventy glden to the gaming table was a memorable one for me. I began by staking ten glden upon passe. For passe I had always had a sort of predilection, yet I lost my stake upon it. This left me with sixty glden in silver. After a moment's thought I selected zero—beginning by staking five glden at a time. Twice I lost, but the third round suddenly brought up the desired coup. I could almost have died with joy as I received my one hundred and seventy-five glden. Indeed, I have been less pleased when, in former times, I have won a hundred thousand glden. Losing no time, I staked another hundred glden upon the red, and won; two hundred upon the red, and won; four hundred upon the black, and won; eight hundred upon manque, and won. Thus, with the addition of the remainder of my original capital, I found myself possessed, within five minutes, of seventeen hundred glden. Ah, at such moments one forgets both oneself and one's former failures! This I had gained by risking my very life. I had dared so to risk, and behold, again I was a member of mankind!

I went and hired a room, I shut myself up in it, and sat counting my money until three o'clock in the morning. To think that when I awoke on the morrow, I was no lacquey! I decided to leave at once for Homburg. There I should neither have to serve as a footman nor to lie in prison. Half an hour before starting, I went and ventured a couple of stakes—no more; with the result that, in all, I lost fifteen hundred florins. Nevertheless, I proceeded to Homburg, and have now been there for a month.

Fyodor DOSTOYEVSKY (1821 — †1881), *The Gambler* (1866), XVII (transl. C. J. Hogarth)

Many systems have inputs with values that cannot be measured, but which can be described as random variables with a known distribution. Such systems are called **stochastic systems**.

Stochastic system

37.1 Stochastic processes

Definition 37.1. A **stochastic process** is a process that generates a signal $X(t)$ that depends on time t and is in each time instant a random variable. While we may want to make a distinction between the particular random variable corresponding to time t which is $X(t)$, the signal $X(t)$ that assumes those successive values, and the process $\{X(t)\}$ that may originate different signals consisting of random values $X(t)$, the process itself is often denoted by $X(t)$ as well. \square

Stochastic process

Remark 37.1. A more general definition of a stochastic process as a mathematical entity can be given, but the above suffices for our purpose, which is the study of stochastic systems in time and the identification of models for such systems. \square

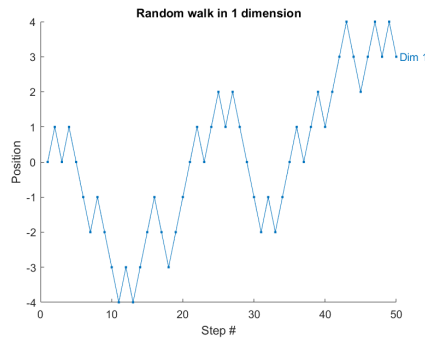


Figure 37.1: One dimensional random walk, as a function of time.

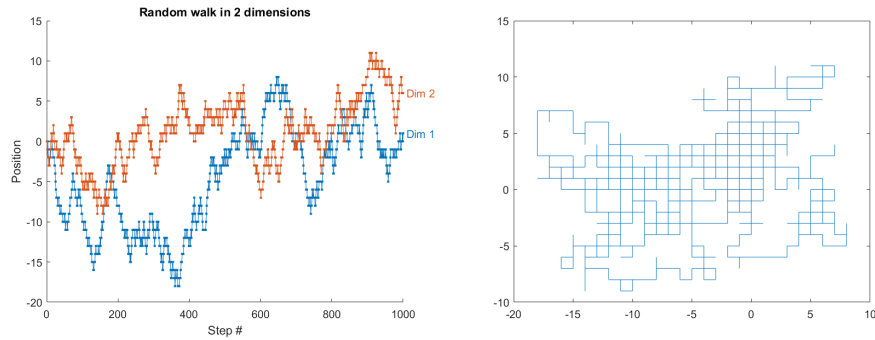


Figure 37.2: Two dimensional random walk. Left: the x and y coordinates as functions of time. Right: the movement on the plane, beginning at $(0, 0)$ and following the random walks in the left plot.

Continuous and discrete stochastic processes

A stochastic process is either

- **continuous**, if $t \in \mathbb{R}_0^+$;
- **discrete**, if $t = kT_s$, where $T_s > 0$ is the sample time and $k \in \mathbb{Z}_0^+$ is the sample. In this case, sample k corresponding to value $X(kT_s)$ is usually denoted as X_k .

Random walk

Example 37.1. A **random walk** is a discrete stochastic process given by

$$X_0 = 0 \quad (37.1)$$

$$X_{k+1} = X_k + x_k, \quad k = 0, 1, 2, 3 \dots \quad (37.2)$$

$$P(x_k = +1) = \frac{1}{2} \quad (37.3)$$

$$P(x_k = -1) = \frac{1}{2} \quad (37.4)$$

More precisely, if the probabilities of variation x_k being -1 and $+1$ are equal as in (37.3)–(37.4), we have an **unbiased random walk**. When these probabilities are different, we have a **biased random walk**.

Figure 37.1 shows an unbiased random walk with 50 instants. Two random walks can be used to define the movement of a particle on a plane, as shown in Figure 37.2 for 1000 instants. Three random walks define a movement in three-dimensional space, which can be used to model the random movement of a particle in suspension in a fluid. This movement is called **Brownian motion**.

Notice that a random walk only assumes integer values. Just as time samples can correspond to multiples of an arbitrary sample time T_s , these integer values assumed by the random walk can correspond to multiples of some unit of length. \square

Brownian motion

Example 37.2. Brownian motion can be better modelled replacing the unbiased random walk by a stochastic process which is also discrete in time but continuous in amplitude:

- in each time instant, the particle travels a distance d which is a random number, following a normal distribution with zero mean and variance σ^2 ;

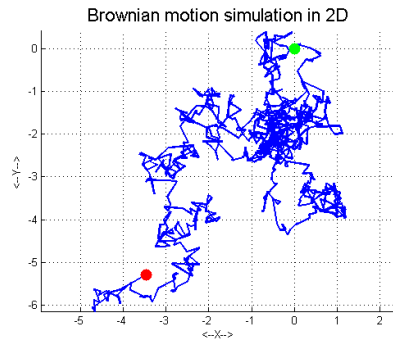


Figure 37.3: Simulation of Brownian motion as in Example 37.2, beginning at $(0, 0)$, for 1000 time samples.

- the direction in which this movement takes place is uniformly distributed. If the movement takes place on a plane, there is only one angle θ , uniformly distributed in $[0, 2\pi[$ (or any other geometrically equivalent interval of angles). In three-dimensional space, two angles are needed.

Notice that a travel with distance d and angle θ is the same as a travel with distance $-d$ and angle $\theta + \pi$. Figure 37.3 shows one such simulation of Brownian motion. \square

Definition 37.2. The random value $X(t)$ assumed by a stochastic process is characterised by a probability distribution, given either by

Probability distribution

- the **probability mass function** (PMF) $f_X(x)$, if the variable assumes values in a discrete set, in which situation each of the N possible outcomes x_k , $k = 1, 2, \dots, N$ has a probability $P(X(t) = x_k)$ which verifies

PMF

$$0 \leq P(X(t) = x_k) \leq 1, \quad \forall k = 1, 2, \dots, N \quad (37.5)$$

$$\sum_{k=1}^N P(X(t) = x_k) = 1 \quad (37.6)$$

$$f_X(x) = P(X(t) = x) \quad (37.7)$$

- the **probability distribution function** (PDF) $f_X(x)$, if the variable assumes values in a continuous set, in which situation each possible outcome x has a probability

PDF

$$P(X(t) = x) = 0, \quad \forall x \quad (37.8)$$

since there are infinitely many outcomes. The PDF is defined so that

$$f_X(x) \geq 0, \quad \forall x \quad (37.9)$$

$$P(x_1 \leq X(t) \leq x_2) = \int_{x_1}^{x_2} f_X(x) dx, \quad x_1 < x_2 \quad (37.10)$$

$$\int_{-\infty}^{+\infty} f_X(x) dx = 1 \quad \square \quad (37.11)$$

Definition 37.3. The **cumulative distribution function** (CDF) of a probability distribution is a function $F_X(x)$ such that

CDF

$$P(X(t) \leq x) = F_X(x) \quad \square \quad (37.12)$$

Notice that:

- For a PMF $f_X(x)$, the CDF is given by the sum of all probabilities up to x , and will be discontinuous at every possible outcome x_k .

$$F_X(x) = \sum_{k=1}^N f_X(x_k), \quad x_1, x_2, \dots, x_k \leq x < x_{k+1}, x_{k+2}, \dots, x_N \quad (37.13)$$

- For a PDF $f_X(x)$, as a consequence of (37.10), the CDF is its indefinite integral.

$$F_X(x) = \int_{-\infty}^x f_X(x) dx \quad (37.14)$$

$$f_X(x) = \frac{dF_X(x)}{dx} \quad (37.15)$$

Stationarity

A stochastic process may have a time-varying probability distribution, in which case it is said to be **non-stationary**. A stochastic process with a probability distribution that remains the same is **stationary**.

Ergodicity

If it is possible to characterise the probability distribution of a stochastic process from measurements, the process is said to be **ergodic**. Obviously, if the probability distribution keeps changing with time, the statistical properties of the probability distribution cannot be found from measurements, since measurements of different distributions are mixed up, and the process is not ergodic. On the other hand, it is possible to devise stationary processes that are *not* ergodic, as the one in the next example. In other words, stationarity is a necessary, but not sufficient, condition for ergodicity. In what follows we assume ergodic — and hence stationary — stochastic processes; at least, it is assumed that changes in the probability distribution are slow enough for this to be possible, in which case we have a **quasi-stationary** stochastic process.

Only stationary processes can be ergodic

We assume stationary, ergodic stochastic processes

Example 37.3. Consider a stochastic process given by

$$X(t) = A \sin(t + \theta) \quad (37.16)$$

where θ is a random constant uniformly distributed in $[0, 2\pi]$, and A is a random constant with a normal distribution. $X(t)$ will be in each instance a sinusoid, and is stationary because A and θ are constants and thus do not have a distribution varying with time; in fact, because these are constants, $X(t)$ will even be deterministic, i.e. we can predict future values without any uncertainty, if there is no noise in the measurements of past values of $X(t)$. We can find the values of A and θ from measurements of a particular instance of $X(t)$, but since they are random they will not have the same value in other instances. Consequently, knowing A and θ from one instance of $X(t)$ will tell us nothing about the values they will have when $X(t)$ takes place again (say, after stopping and restarting). Thus, this stochastic process, even though stationary and even deterministic, is not ergodic. \square

Remark 37.2. Examples 37.1 and 37.2 show MATLAB simulations, that use **pseudo-random number generators**. As you surely know, the outputs of such algorithms are not really random, but follow a specified distribution.

MATLAB pseudo-random number generators

Among others, the following MATLAB commands are useful for this purpose:

- `rand(a,b)` returns an $a \times b$ matrix of numbers uniformly distributed in $[0, 1]$;
- `randi(n,a,b)` returns an $a \times b$ matrix of natural numbers uniformly distributed in $\{1, 2, \dots, n\}$;
- `rand(a,b)` returns an $a \times b$ matrix of numbers following a standard normal distribution (i.e. with mean 0 and standard deviation 1);
- all these functions, if arguments `a` and `b` are not provided, return only one number. \square

37.2 Characterisation of stochastic processes

Several statistical properties are of interest when studying probability distributions and stochastic processes:

- expected value;
- variance;
- standard deviation;

- autocorrelation;
- autocorrelation coefficient;
- autocovariance.

For each of them, we will:

- define it for a stochastic process;
- provide a numerical approximation of the definition to estimate it from a signal that measures a particular instance of the stochastic process. These expressions can be applied to any signal, not just in the presence of a stochastic process.

Definition 37.4. The **expected value** $E[X(t)]$ of a stochastic process $X(t)$ is *Expected value* the centroid of its distribution:

$$E[X(t)] = \int_{-\infty}^{+\infty} x f_X(x) dx, \text{ for a PDF} \quad (37.17)$$

$$E[X(t)] = \sum_{k=1}^N x_k f_X(x_k), \text{ for a PMF} \quad \square \quad (37.18)$$

To estimate the expected value of a stochastic process from the values observed up to time t , the **mean** of the observed values $\overline{X}(t)$ is used. *Mean*

Definition 37.5. The mean of a signal continuous in time, for measurements from $t = 0$ to $t = t_{\text{final}}$, is

$$\overline{X}(t) = \frac{1}{t_{\text{final}}} \int_0^{t_{\text{final}}} x(t) dt \quad \square \quad (37.19)$$

Definition 37.6. The mean of a signal sampled in time with sampling time T_s , for $n + 1$ measurements from $t = 0$ to $t = T_s n$, is

$$\overline{X}(t) = \frac{1}{n + 1} \sum_{k=0}^n X_k \quad \square \quad (37.20)$$

The discrete time case is more often found in practice.

Remark 37.3. Since

$$\lim_{t \rightarrow +\infty} \overline{X}(t) = E[X(t)] \quad (37.21)$$

it is usual to employ $E[X(t)]$ and $\overline{X}(t)$ interchangeably. \square

Definition 37.7. The **variance** σ_X^2 is the average of the squares of the deviations: *Variance*

$$\sigma_X^2 = E \left[(X(t) - E[X(t)])^2 \right] \quad \square \quad (37.22)$$

Theorem 37.1. For a stationary stochastic process,

$$\begin{aligned} \sigma_X^2 &= E \left[(X(t) - E[X(t)])^2 \right] \\ &= E \left[X^2(t) - 2X(t)E[X(t)] + (E[X(t)])^2 \right] \\ &= E \left[X^2(t) - \underbrace{E[X(t) 2E[X(t)]]}_{2E[X(t)]E[X(t)]} + \underbrace{E[(E[X(t)])^2]}_{(E[X(t)])^2} \right] \\ &= E \left[X^2(t) - (E[X(t)])^2 \right] = \overline{X^2(t)} - \overline{X(t)}^2 \quad \square \end{aligned} \quad (37.23)$$

Corollary 37.1.

$$\overline{X^2(t)} = \overline{X(t)}^2 + \sigma_X^2 \quad \square \quad (37.24)$$

The variance of a stochastic process is estimated from observed values applying (37.19)–(37.20) to (37.23), i.e. approximating the variance of the distribution by the variance of the observed values.

- In the continuous case, for measurements from $t = 0$ to $t = t_{\text{final}}$:

$$\sigma_X^2 = \frac{1}{t_{\text{final}}} \int_0^{t_{\text{final}}} x^2(t) dt - \left(\frac{1}{t_{\text{final}}} \int_0^{t_{\text{final}}} x(t) dt \right)^2 \quad (37.25)$$

- In discrete time with sampling time T_s , for $n+1$ measurements from $t = 0$ to $t = T_s n$:

$$\sigma_X^2 = \frac{1}{n+1} \sum_{k=0}^n X_k^2 - \left(\frac{1}{n+1} \sum_{k=0}^n X_k \right)^2 \quad (37.26)$$

Standard deviation

Definition 37.8. The **standard deviation** σ_X is the square root of the variance:

$$\sigma_X = \sqrt{\sigma_X^2} \quad \square \quad (37.27)$$

Example 37.4. The more samples of a process we have, the better our estimation of its mean and variance is. Consider a uniform distribution X in $[0, 1]$, which has, applying (37.17) and (37.23),

$$f_X(x) = \begin{cases} 1, & \text{if } 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases} \quad (37.28)$$

$$E[X] = \int_0^1 x dx = \frac{1}{2} \quad (37.29)$$

$$\sigma_X^2 = \int_0^1 x^2 dx - \left(\frac{1}{2} \right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12} \quad (37.30)$$

MATLAB's mean and var

commands In three instances of 10 samples of this stochastic process, the mean and the variance are

```
>> for k=1:4, X=rand(1,10); disp(['mean ' num2str(mean(X))...
', variance ' num2str(var(X))]), end
mean 0.62386, variance 0.11961
mean 0.66021, variance 0.10946
mean 0.58731, variance 0.086137
mean 0.39782, variance 0.13008
```

Means and variances do not stray much from the expected values, but are much closer to them with 100 samples:

```
>> for k=1:4, X=rand(1,100); disp(['mean ' num2str(mean(X))...
', variance ' num2str(var(X))]), end
mean 0.49027, variance 0.080127
mean 0.46436, variance 0.071762
mean 0.49628, variance 0.084483
mean 0.49745, variance 0.078337
```

And even more with 1000:

```
>> for k=1:4, X=rand(1,1000); disp(['mean ' num2str(mean(X))...
', variance ' num2str(var(X))]), end
mean 0.50516, variance 0.081755
mean 0.48485, variance 0.082201
mean 0.49836, variance 0.082948
mean 0.4916, variance 0.083498
```

□

Remark 37.4. Notice how the estimates of the mean and the variance are random variables themselves. That is why, in the previous example, several instances of each were found for the same number of samples. If they are updated in real time, their time-changing values will be stochastic processes too. □

Definition 37.9. The **autocorrelation** $R_X(t_1, t_2)$ of a stochastic process depends on two time instants t_1 and t_2 and is given by

$$R_X(t_1, t_2) = E[X(t_1)X(t_2)] \quad \square \quad (37.31)$$

Autocorrelation has several important properties.

Lemma 37.1. Autocorrelation is an even function:

$$R_X(t_1, t_2) = E[X(t_1)X(t_2)] = E[X(t_2)X(t_1)] = R_X(t_2, t_1) \quad \square \quad (37.32)$$

Remark 37.5. For a stationary stochastic process,

$$R_X(t_1, t_2) = R_X(t_1 + \Delta t, t_2 + \Delta t), \quad \forall t_1, t_2, \Delta t \quad (37.33)$$

Since Δt can take any value, we can make $\Delta t = -t_1$ and conclude that

$$R_X(t_1, t_2) = R_X(0, t_2 - t_1), \quad \forall t_1, t_2 \quad (37.34)$$

In other words, for a stationary process the autocorrelation only depends on the time interval between two instants $\tau = t_2 - t_1$:

$$R_X(\tau) = E[X(t)X(t + \tau)] \quad \square \quad (37.35)$$

The following result is a consequence of (37.35).

Theorem 37.2. Autocorrelation as a function of τ is an even function: $R_X(\tau) = R_X(-\tau)$

$$R_X(\tau) = E[X(t)X(t + \tau)] = E[X(\underbrace{t + \tau}_{t'})X(t)] = E[X(t')X(t' - \tau)] = R_X(-\tau) \quad \square \quad (37.36)$$

Remark 37.6. (37.36) can also be seen as a consequence of (37.32) and (37.34): since $R_X(t_1, t_2)$ in fact only depends on $\tau = t_2 - t_1$, then

$$R_X(\tau) = R_X(t_2 - t_1) = R_X(t_1, t_2) = R_X(t_2, t_1) = R_X(t_1 - t_2) = R_X(-\tau) \quad \square \quad (37.37)$$

Theorem 37.3. For $\tau = 0$,

$$R_X(0) = E[X(t)X(t)] = \overline{X^2(t)} = \overline{X(t)}^2 + \sigma_X^2 \quad (37.38)$$

The last equality is taken from (37.24). \square

To estimate the autocorrelation, the expected value in (37.35) is estimated using a mean, as in (37.19)–(37.20):

- In the continuous case, for measurements from $t = 0$ to $t = t_{\text{final}}$:

$$R_X(\tau) = \frac{1}{t_{\text{final}} - \tau} \int_0^{t_{\text{final}} - \tau} x(t)x(t + \tau) dt \quad (37.39)$$

Notice that we cannot integrate from 0 to t_{final} , as in (37.19) or (37.25), because the second measurement would be out of range. Values of τ so large that they are close to t_{final} will lead to bad estimations of $R_X(\tau)$, because measurements will be few. In practice it is often a good idea to make $0 \leq \tau \leq \frac{1}{3}t_{\text{final}}$ or $0 \leq \tau \leq \frac{1}{2}t_{\text{final}}$. Because $R_X(\tau)$ is an even function, we will know its values for $-\frac{1}{3}t_{\text{final}} \leq \tau \leq \frac{1}{3}t_{\text{final}}$ or $-\frac{1}{2}t_{\text{final}} \leq \tau \leq \frac{1}{2}t_{\text{final}}$.

- In discrete time with sampling time T_s , for $n + 1$ measurements from $t = 0$ to $t = T_s n$:

$$R_X(\nu) = \frac{1}{n - \nu + 1} \sum_{k=0}^{n-\nu} X_k X_{k+\nu} \quad (37.40)$$

Notice that in the discrete case it is usual to give the autocorrelation depending on the number of samples ν between t_1 and t_2 , rather than the time $\tau = T_s \nu$.

Remark 37.7. If $\tau \ll t_{\text{final}}$, or $\nu \ll n$, (37.39)-(37.40) can be approximated as *Parzen (biased) estimator of $R_X(\tau)$*

$$R_X(\tau) = \frac{1}{t_{\text{final}}} \int_0^{t_{\text{final}}-\tau} x(t)x(t+\tau) dt \quad (37.41)$$

$$R_X(\nu) = \frac{1}{n+1} \sum_{k=0}^{n-\nu} X_k X_{k+\nu} \quad (37.42)$$

These biased approximations with a constant denominator are the so-called Parzen estimator of the autocorrelation. \square

Example 37.5. The autocorrelation can be numerically found from (37.40) as follows:

```
Ts = 0.01; tfinal = 100;
t = 0 : Ts : tfinal;
x = exp(0.05*t);
N = length(t)-1; % there are N+1 points; the first is labelled 0
M = floor(N*.5); % number of points for which Rx is found
Rx = zeros(1,M);
for k = 0 : M-1
    Rx(k+1) = sum( x(1:N-k+1).*x(1+k:N+1) ) / (N-k+1);
end
figure,subplot(2,1,1),plot(t,x),xlabel('t'),ylabel('x')
subplot(2,1,2),plot((0:M-1)*Ts,Rx),xlabel('t'),ylabel('R_x'),xlim([0,tfinal])
```

The result of the example above, for $X(t) = e^{0.05t}$, is shown in Figure 37.4 together with that of $X(t) = e^{-0.05t}$. Values are given only for $\tau \geq 0$, but remember that the function is even. Notice how the largest value is that for $R_X(0)$.

Figure 37.5 shows the result for two random signals. There are negative values of $R_X(\tau)$ now; no value has an absolute larger than $R_X(0)$.

Figure 37.6 shows the result for two periodic functions. In these cases $R_X(0)$ has the largest absolute value, which recurs with the period of the function. \square

This example suggests the following result.

Theorem 37.4. The maximum absolute value of the autocorrelation is found at $\tau = 0$:

$$|R_X(\tau)| \leq R_X(0), \quad \forall \tau \quad (37.43)$$

Proof. Since $(X(t+\tau) \pm X(t))^2 \geq 0$, its expected value is also non-negative:

$$\begin{aligned} E \left[(X(t+\tau) \pm X(t))^2 \right] &\geq 0 \\ \Rightarrow E \left[(X(t+\tau))^2 \pm 2X(t)X(t+\tau) + (X(t))^2 \right] &\geq 0 \\ \Rightarrow E \left[(X(t+\tau))^2 \right] \pm E [2X(t)X(t+\tau)] + E \left[(X(t))^2 \right] &\geq 0 \\ \Rightarrow \underbrace{E \left[(X(t+\tau))^2 \right]}_{R_X(0)} \pm \underbrace{E [2X(t)X(t+\tau)]}_{2R_X(\tau)} + \underbrace{E \left[(X(t))^2 \right]}_{R_X(0)} &\geq 0 \\ \Rightarrow 2R_X(0) \pm 2R_X(\tau) &\geq 0 \end{aligned} \quad (37.44)$$

where the last result was obtained applying (37.38). From these two inequalities,

$$\begin{cases} R_X(0) + R_X(\tau) \geq 0 \\ R_X(0) - R_X(\tau) \geq 0 \end{cases} \Rightarrow \begin{cases} R_X(\tau) \geq -R_X(0) \\ R_X(0) \geq R_X(\tau) \end{cases} \Rightarrow -R_X(0) \leq R_X(\tau) \leq R_X(0) \quad \square \quad (37.45)$$

This range can actually be narrowed down further, as seen below in (37.55). Meanwhile, it is possible to better understand why (37.43) is true with the following reasoning:

- $R_X(\tau)$ is the expected value of the product $X(t)X(t+\tau)$.

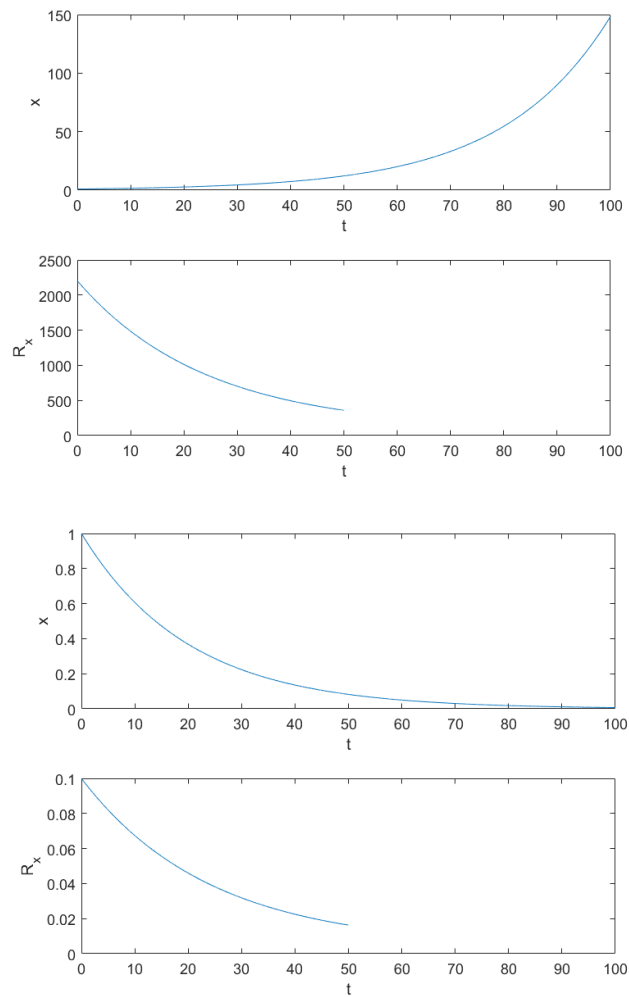


Figure 37.4: Top: autocorrelation of $e^{0.05t}$. Bottom: autocorrelation of $e^{-0.05t}$.

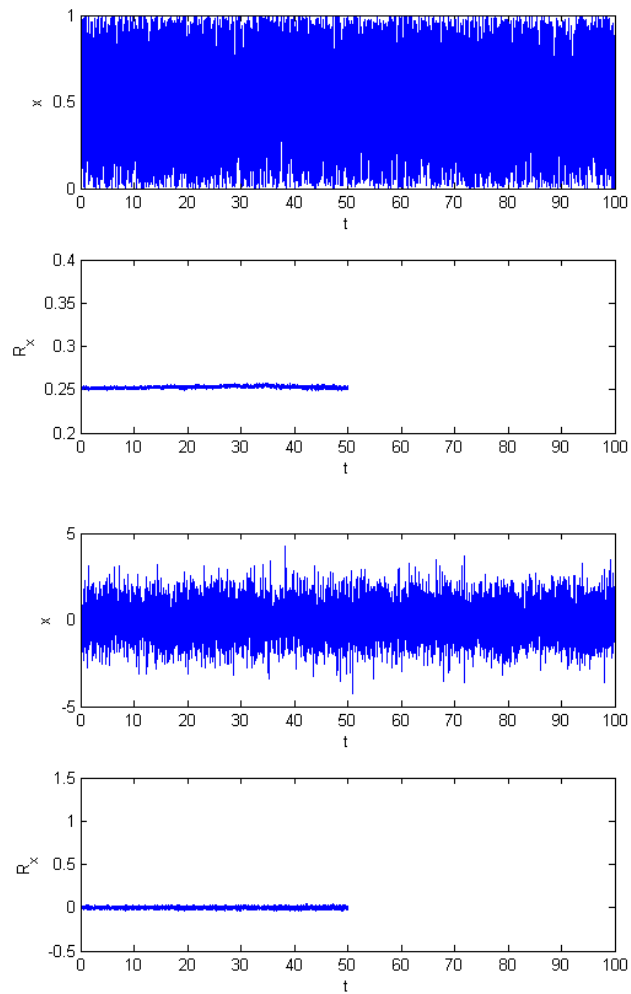


Figure 37.5: Top: autocorrelation of uniformly distributed values in $[0, 1]$. Bottom: autocorrelation of standard normally distributed values.

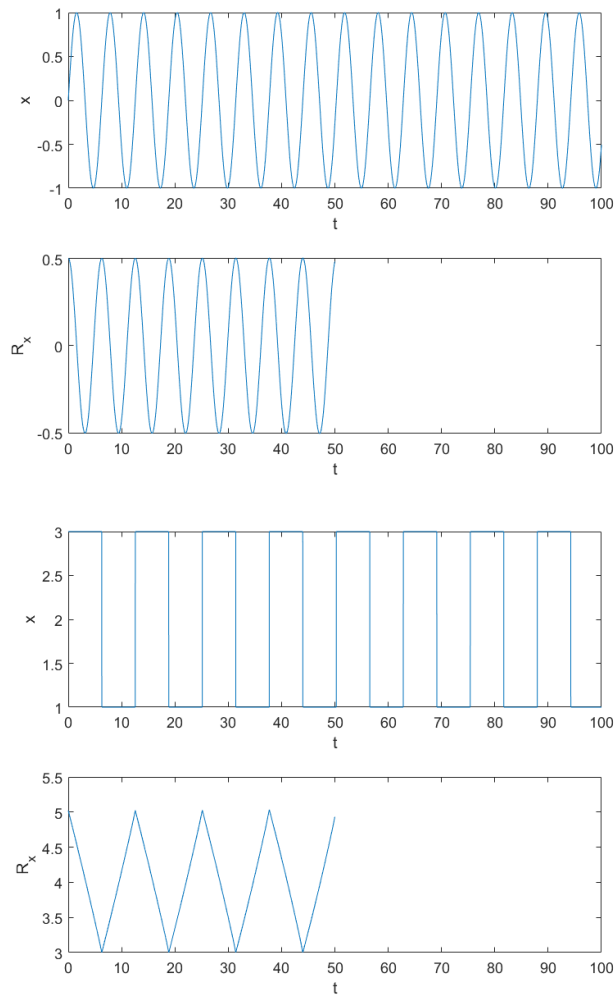


Figure 37.6: Top: autocorrelation of $\sin(t)$. Bottom: autocorrelation of a square wave with frequency 0.5 rad/s.

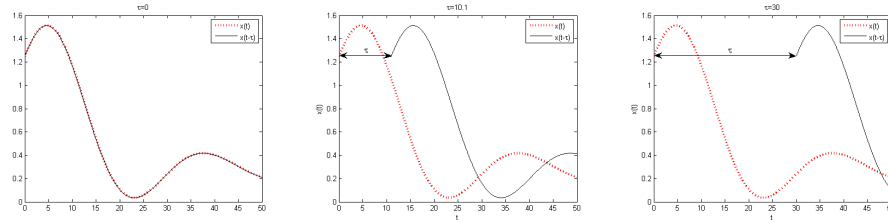


Figure 37.7: Nothing is more similar to a signal than itself. The figures show how the product $X(t)X(t + \tau)$, used to calculate the autocorrelation $R_X(\tau) = E[X(t)X(t + \tau)]$, is found.

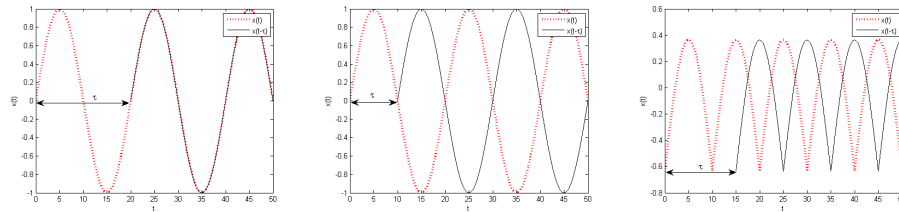


Figure 37.8: Left: how $R_X(0)$ recurs in the autocorrelation when τ is a multiple of the period. Centre: how $-R_X(0)$ recurs in the autocorrelation when a periodic signal is symmetrical. Right: periodic signal with zero mean for which it is impossible that $R_X(\tau) = -R_X(0)$.

- When $\tau = 0$, then $X(t)$ is multiplied by $X(t)$ itself. The result is positive everywhere. Furthermore, wherever $|X|$ is large, it is multiplied by a large value, which is itself.
- When $\tau \neq 0$, then $X(t)X(t + \tau)$ may be negative for some values of τ . Even where this is not so, large values of $|X|$ will be multiplied by smaller values. Thus, $|R_X(\tau)| < R_X(0)$. This is illustrated in Figure 37.7.
- The only exception of the above occurs when the signal is periodic and τ is such that local maxima and minima are again aligned. In that case, $R_X(0)$ is obtained again. This is illustrated in Figure 37.8.
- If the periodic signal is symmetric around zero, a time shift of half period will, so to say, turn the signal upside down. This happens for instance for $X(t) = \sin(t)$, since

$$\sin(t + (2k + 1)\pi) = -\sin(t), \quad k \in \mathbb{Z} \tag{37.46}$$

In cases such as this, $-R_X(0)$ recurs whenever $X(t)$ and $X(t + \tau)$ are in phase opposition.

The reasoning above also justifies the following results.

Theorem 37.5. If $X(t)$ is not periodic,

$$|R_X(\tau)| < R_X(0), \quad \forall \tau \neq 0 \tag{37.47}$$

If $X(t)$ is periodic with period T , then $R_X(\tau) = R_X(0)$ if and only if τ is a multiple of T , i.e. $\tau = kT$, $k \in \mathbb{Z}$.

If $X(t)$ is not only periodic but also symmetric, then $R_X(\tau) = -R_X(0)$ if and only if τ is a multiple of T plus half period, i.e. $\tau = (k + \frac{1}{2})T$, $k \in \mathbb{Z}$.

Proof. If $X(t)$ is periodic,

$$\begin{aligned} X(t + kT) &= X(t), \quad k \in \mathbb{Z} \\ \Rightarrow X(t)X(t + kT) &= X^2(t) \\ \Rightarrow \underbrace{R_X(kT)}_{E[X(t)X(t+kT)]} &= \underbrace{R_X(0)}_{E[X(t)X(t)]} \end{aligned} \tag{37.48}$$

Nothing is more similar to a signal than itself

Autocorrelation of periodic signals

If $X(t)$ is symmetric,

$$\begin{aligned} X\left(t + \left(k + \frac{1}{2}\right)T\right) &= -X(t), \quad k \in \mathbb{Z} \\ \Rightarrow X(t)X\left(t + \left(k + \frac{1}{2}\right)T\right) &= -X^2(t) \\ \Rightarrow \underbrace{R_X\left(\left(k + \frac{1}{2}\right)T\right)}_{E[X(t)X(t+(k+\frac{1}{2})T)]} &= \underbrace{-R_X(0)}_{-E[X^2(t)]} \end{aligned} \quad (37.49)$$

As to the necessity of the conditions, suppose that $R_X(\tau) = R_X(0)$ for some value of τ . This means that $X(t)X(t+\tau) = X^2(t) \Rightarrow X(t+\tau) = X(t)$, and, since this must be true for all values of t , we conclude that τ is a period of $X(t)$. The reasoning for the cases when $R_X(\tau) = -R_X(0)$ is similar. \square

Definition 37.10. The **autocovariance** $C_X(t_1, t_2)$ of a stochastic process is defined similarly to the autocorrelation, but using deviations from the mean: *Autocovariance*

$$C_X(t_1, t_2) = E\left[(X(t_1) - E[X])(X(t_2) - E[X])\right] \quad \square \quad (37.50)$$

Remark 37.8. In this definition it is assumed that X is ergodic, and hence stationary; and thus $E[X]$ is constant. This means that C_X will, just like R_X , depend only on $\tau = t_2 - t_1$:

$$C_X(\tau) = E\left[(X(t) - E[X])(X(t+\tau) - E[X])\right] \quad \square \quad (37.51)$$

Remark 37.9. The autocovariance $C_X(t_1, t_2)$ of $X(t)$ is the autocorrelation of $X(t) - E[X]$, which has zero mean. This is another way to show that it only depends on τ if $X(t)$ is ergodic. It also shows immediately that $C_X(\tau) = C_X(-\tau)$.

Furthermore, the code from Example 37.5 can be used to find the autocovariance, provided that the mean is subtracted from the signal. \square

Lemma 37.2.

Relation between autocorrelation and covariance

$$\begin{aligned} C_X(\tau) &= E\left[(X(t) - E[X])(X(t+\tau) - E[X])\right] \\ &= E\left[X(t)X(t+\tau) - X(t)E[X] - X(t+\tau)E[X] + E[X]^2\right] \\ &= E[X(t)X(t+\tau)] - E[X(t)E[X]] - E[X(t+\tau)E[X]] + E[E[X]^2] \\ &= \underbrace{E[X(t)X(t+\tau)]}_{R_X(\tau)} - \underbrace{E[X]E[X(t)]}_{E[X]^2} - \underbrace{E[X]E[X(t+\tau)]}_{E[X]^2} + E[X]^2 \\ &= R_X(\tau) - E[X]^2 \quad \square \end{aligned} \quad (37.52)$$

Corollary 37.2. If $\bar{X} = 0$ then $R_X(\tau) = C_X(\tau)$. \square

Corollary 37.3. From (37.52), (37.38) and (37.24),

$$\begin{aligned} C_X(0) &= R_X(0) - E[X]^2 \\ &= E[X^2] - E[X]^2 \\ &= E[X]^2 + \sigma_X^2 - E[X]^2 = \sigma_X^2 \quad \square \end{aligned} \quad (37.53)$$

Corollary 37.4. Since $C_X(\tau)$ is the autocorrelation of $X(t) - \bar{X}$, it verifies (37.43), and thus, from (37.53),

$$|C_X(\tau)| \leq C_X(0) = \sigma_X^2, \quad \forall \tau \quad \square \quad (37.54)$$

Theorem 37.6. The autocorrelation is limited by

Autocorrelation is bounded

$$\bar{X}^2 - \sigma_X^2 \leq R_X(\tau) \leq \bar{X}^2 + \sigma_X^2 = \bar{X}^2 = R_X(0), \quad \forall \tau \quad (37.55)$$

Proof. From (37.52) we know that

$$R_X(\tau) = C_X(\tau) + E[X]^2 \quad (37.56)$$

So, from (37.54),

$$\begin{aligned} -\sigma_X^2 &\leq C_X(\tau) \leq \sigma_X^2 \\ \Rightarrow -\sigma_X^2 + \bar{X}^2 &\leq \underbrace{C_X(\tau) + \bar{X}^2}_{R_X(\tau)} \leq \sigma_X^2 + \bar{X}^2 \end{aligned} \quad (37.57)$$

Concerning the upper limit, see (37.24) and (37.38). \square

Mean value of $R_X(\tau)$

Remark 37.10. This shows that if $X(t)$ has a mean value then $R_X(\tau)$ also has a mean value. This mean value is in fact \bar{X}^2 . This is clear from (37.55) for a constant signal (which has $\sigma_X^2 = 0$). \square

Autocorrelation coefficient

The limits in (37.55) show that autocorrelation can be normalised.

Definition 37.11. The **autocorrelation coefficient** $\rho_X(\tau)$, that verifies $-1 \leq \rho_X(\tau) \leq 1$, is given by

$$\rho_X(\tau) = \frac{R_X(\tau) - \bar{X}^2}{\sigma_X^2} = \frac{C_X(\tau)}{\sigma_X^2} \quad \square \quad (37.58)$$

Remark 37.11. Some authors call autocorrelation to the autocorrelation coefficient. \square

Autocorrelation and autocovariance matrices

Consider a signal in discrete time given by

$$\begin{aligned} \mathbf{X} &= [X_0 \ X_1 \ X_2 \ \cdots \ X_n]^T \\ &= [X(0) \ X(T_s) \ X(2T_s) \ \cdots \ X(nT_s)]^T \end{aligned} \quad (37.59)$$

We can arrange in matrices the values of its autocorrelation

$$\begin{aligned} \mathbf{R}_X &= \begin{bmatrix} R_X(0) & R_X(1) & R_X(2) & \cdots & R_X(n) \\ R_X(-1) & R_X(0) & R_X(1) & \cdots & R_X(n-1) \\ R_X(-2) & R_X(-1) & R_X(0) & \cdots & R_X(n-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R_X(-n) & R_X(-n+1) & R_X(-n+2) & \cdots & R_X(0) \end{bmatrix} \\ &= E \begin{bmatrix} X_0 X_0 & X_0 X_1 & X_0 X_2 & \cdots & X_0 X_n \\ X_1 X_0 & X_1 X_1 & X_1 X_2 & \cdots & X_1 X_n \\ X_2 X_0 & X_2 X_1 & X_2 X_2 & \cdots & X_2 X_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_n X_0 & X_n X_1 & X_n X_2 & \cdots & X_n X_n \end{bmatrix} \\ &= E \begin{bmatrix} X(0)X(0) & X(0)X(T_s) & X(0)X(2T_s) & \cdots & X(0)X(nT_s) \\ X(T_s)X(0) & X(T_s)X(T_s) & X(T_s)X(2T_s) & \cdots & X(T_s)X(nT_s) \\ X(2T_s)X(0) & X(2T_s)X(T_s) & X(2T_s)X(2T_s) & \cdots & X(2T_s)X(nT_s) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X(nT_s)X(0) & X(nT_s)X(T_s) & X(nT_s)X(2T_s) & \cdots & X(nT_s)X(nT_s) \end{bmatrix} \\ &= E [\mathbf{X}\mathbf{X}^T] \end{aligned} \quad (37.60)$$

and the values of its autocovariance

$$\begin{aligned}
 \mathbf{C}_X &= \begin{bmatrix} C_X(0) & C_X(1) & C_X(2) & \cdots & C_X(n) \\ C_X(-1) & C_X(0) & C_X(1) & \cdots & C_X(n-1) \\ C_X(-2) & C_X(-1) & C_X(0) & \cdots & C_X(n-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_X(-n) & C_X(-n+1) & C_X(-n+2) & \cdots & C_X(0) \end{bmatrix} \\
 &= \sigma_X^2 \begin{bmatrix} \rho_X(0) & \rho_X(1) & \rho_X(2) & \cdots & \rho_X(n) \\ \rho_X(-1) & \rho_X(0) & \rho_X(1) & \cdots & \rho_X(n-1) \\ \rho_X(-2) & \rho_X(-1) & \rho_X(0) & \cdots & \rho_X(n-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_X(-n) & \rho_X(-n+1) & \rho_X(-n+2) & \cdots & \rho_X(0) \end{bmatrix} \\
 &= E \begin{bmatrix} X_0 X_0 & X_0 X_1 & X_0 X_2 & \cdots & X_0 X_n \\ X_1 X_0 & X_1 X_1 & X_1 X_2 & \cdots & X_1 X_n \\ X_2 X_0 & X_2 X_1 & X_2 X_2 & \cdots & X_2 X_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_n X_0 & X_n X_1 & X_n X_2 & \cdots & X_n X_n \end{bmatrix} \\
 &= \begin{bmatrix} (X(0) - \bar{X})(X(0) - \bar{X}) & (X(0) - \bar{X})(X(T_s) - \bar{X}) & (X(0) - \bar{X})(X(2T_s) - \bar{X}) & \cdots & (X(0) - \bar{X})(X(nT_s) - \bar{X}) \\ (X(T_s) - \bar{X})(X(0) - \bar{X}) & (X(T_s) - \bar{X})(X(T_s) - \bar{X}) & (X(T_s) - \bar{X})(X(2T_s) - \bar{X}) & \cdots & (X(T_s) - \bar{X})(X(nT_s) - \bar{X}) \\ (X(2T_s) - \bar{X})(X(0) - \bar{X}) & (X(2T_s) - \bar{X})(X(T_s) - \bar{X}) & (X(2T_s) - \bar{X})(X(2T_s) - \bar{X}) & \cdots & (X(2T_s) - \bar{X})(X(nT_s) - \bar{X}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (X(nT_s) - \bar{X})(X(0) - \bar{X}) & (X(nT_s) - \bar{X})(X(T_s) - \bar{X}) & (X(nT_s) - \bar{X})(X(2T_s) - \bar{X}) & \cdots & (X(nT_s) - \bar{X})(X(nT_s) - \bar{X}) \end{bmatrix} \\
 &= E \left[(\mathbf{X} - \bar{\mathbf{X}}) (\mathbf{X} - \bar{\mathbf{X}})^T \right] \tag{37.61}
 \end{aligned}$$

Since $R_X(\tau) = R_X(-\tau)$ and $C_X(\tau) = C_X(-\tau)$, matrices \mathbf{R}_X and \mathbf{C}_X are symmetric. The farther from the main diagonal, the poorer the estimations will be, as already mentioned about $R_X(\tau)$. $\mathbf{R}_X = \mathbf{R}_X^T$
 $\mathbf{C}_X = \mathbf{C}_X^T$

37.3 Relations between stochastic processes

Since we are studying systems, we often have to study two stochastic processes simultaneously: an input and an output. In this situation, their **joint probability distribution** is needed. *Joint probability distribution*

Definition 37.12. The **joint probability distribution function** (joint PDF) *Joint PDF* $f_{XY}(x, y)$, of two variables that assume values in a continuous set is defined so that

$$f_{X,Y}(x, y) \geq 0, \forall x, y \tag{37.62}$$

$$\begin{aligned}
 P(x_1 \leq X(t) \leq x_2 \wedge y_1 \leq Y(t) \leq y_2) &= P(x_1 \leq X(t) \leq x_2 | y_1 \leq Y(t) \leq y_2) P(y_1 \leq Y(t) \leq y_2) \\
 &= P(y_1 \leq Y(t) \leq y_2 | x_1 \leq X(t) \leq x_2) P(x_1 \leq X(t) \leq x_2) \\
 &= \int_{y_1}^{y_2} \int_{x_1}^{x_2} f_{XY}(x, y) dx dy, \quad x_1 < x_2, \quad y_1 < y_2
 \end{aligned} \tag{37.63}$$

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y) dx dy = 1 \quad \square \tag{37.64}$$

Definition 37.13. The **joint probability mass function** (joint PMF) *Joint PMF* $f_{XY}(x, y)$ of two variables that assume values in a discrete set is defined so that

$$\begin{aligned}
 f_{XY}(x, y) &= P(X(t) = x \wedge Y(t) = y) \\
 &= P(X(t) = x | Y(t) = y) P(Y(t) = y) \\
 &= P(Y(t) = y | X(t) = x) P(X(t) = x)
 \end{aligned} \tag{37.65}$$

$$0 \leq f_{XY}(x, y) \leq 1, \forall x, y \tag{37.66}$$

$$\sum_x \sum_y f_{XY}(x, y) = 1 \quad \square \tag{37.67}$$

Definition 37.14. The **joint cumulative distribution function** (joint CDF) *Joint CDF* of two probability distributions is a function $F_{XY}(x, y)$ such that

$$\begin{aligned} F_{XY}(x, y) &= P(X(t) \leq x \wedge Y(t) \leq y) \\ &= P(X(t) \leq x \mid Y(t) \leq y) P(Y(t) \leq y) \\ &= P(Y(t) \leq y \mid X(t) \leq x) P(X(t) \leq x) \quad \square \end{aligned} \quad (37.68)$$

The joint CDF of a joint PDF is given by

$$F_{X,Y}(x, y) = \int_{-\infty}^y \int_{-\infty}^x f_{X,Y}(x, y) dx dy \quad (37.69)$$

$$f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y} \quad (37.70)$$

and the joint CDF of a joint PMF is given by

$$\begin{aligned} F_{X,Y}(x, y) &= \sum_{m=1}^{N_y} \sum_{k=1}^{N_x} f_{X,Y}(x_k, y_m), \quad (37.71) \\ & \quad x_1, x_2, \dots, x_k \leq x < x_{k+1}, x_{k+2}, \dots, x_{N_x} \\ & \quad y_1, y_2, \dots, y_m \leq y < y_{m+1}, y_{m+2}, \dots, y_{N_y} \end{aligned}$$

Just as we did when we had only one stochastic process, each of the following statistical properties will be

- defined for two stochastic processes;
- numerically approximated from the definition to be estimated from two signals that measure particular instances of the stochastic processes. Again, these expressions can be applied to any signals, not just in the presence of stochastic processes.

Correlation

Definition 37.15. The **cross-correlation**, or simply correlation, $R_{XY}(t_1, t_2)$ of two stochastic processes depends on two time instants t_1 and t_2 and is given by

$$R_{XY}(t_1, t_2) = E[X(t_1)Y(t_2)] \quad \square \quad (37.72)$$

Remark 37.12. The autocorrelation $R_X(t_1, t_2)$ is the correlation of a signal with itself $R_{XX}(t_1, t_2)$. \square

Remark 37.13. For stationary stochastic processes,

$$R_{XY}(t_1, t_2) = R_{XY}(t_1 + \Delta t, t_2 + \Delta t), \quad \forall t_1, t_2, \Delta t \quad (37.73)$$

Since Δt can take any value, we can make $\Delta t = -t_1$ and conclude that

$$R_{XY}(t_1, t_2) = R_{XY}(0, t_2 - t_1), \quad \forall t_1, t_2 \quad (37.74)$$

In other words, for stationary processes the correlation only depends on the time interval between two instants $\tau = t_2 - t_1$:

$$R_{XY}(\tau) = E[X(t)Y(t + \tau)], \quad \forall t \quad \square \quad (37.75)$$

While this is similar to what happens with the autocorrelation, the correlation, unlike the autocorrelation, is not an even function in the general case. Figure 37.9 illustrates why.

$$R_{XY}(\tau) = R_{YX}(-\tau)$$

Additionally, the order of the signals $X(t)$ and $Y(t)$ matters.

Theorem 37.7.

$$\begin{aligned} R_{XY}(\tau) &= E[X(t)Y(t + \tau)] = E[Y(\underbrace{t + \tau}_{t'})X(t)] = \\ &= E[Y(t')X(t' - \tau)] = R_{YX}(-\tau) \quad \square \end{aligned} \quad (37.76)$$

Theorem 37.8.

$$-\sqrt{R_X(0)R_Y(0)} \leq R_{XY}(\tau) \leq \sqrt{R_X(0)R_Y(0)}, \quad \forall \tau \quad (37.77)$$

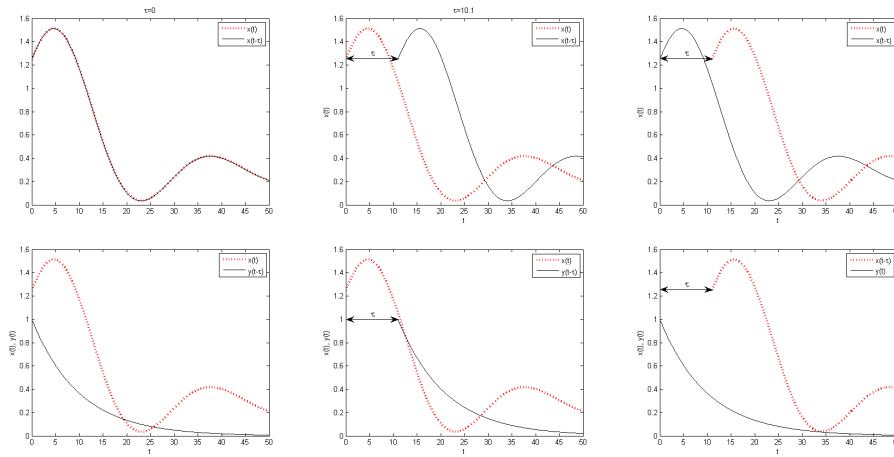


Figure 37.9: Top: in autocorrelation, a signal is correlated with itself, so it does not matter which of the two is shifted in time. Bottom: in the correlation of two different signals, the product depends on which signal is being shifted in time.

Proof. The proof differs from that of (37.43): a similar reasoning would lead to more conservative boundaries for $R_{XY}(\tau)$. This time it has to be argued that the expected value $E[x]$ is a norm, and consequently, the Cauchy-Schwartz inequality, which in the general case states that a norm verifies

$$|\langle \mathbf{u}, \mathbf{v} \rangle|^2 \leq \langle \mathbf{u}, \mathbf{u} \rangle \langle \mathbf{v}, \mathbf{v} \rangle \tag{37.78}$$

gives, for the expected value,

$$|E[ab]|^2 \leq E[a^2] E[b^2] \tag{37.79}$$

Consequently,

$$\begin{aligned} |E[X(t)Y(t+\tau)]|^2 &\leq E[(X(t))^2] E[(Y(t+\tau))^2] \\ &= \underbrace{E[X(t)X(t)]}_{R_X(t-t)=R_X(0)} \underbrace{E[Y(t+\tau)Y(t+\tau)]}_{R_Y(t+\tau-(t+\tau))=R_Y(0)} \end{aligned} \tag{37.80}$$

and the result follows immediately. \square

Once more, narrower limits can be found, given below in (37.93).

Remark 37.14. Notice that, while (37.43) shows that $R_X(\tau)$ has its maximum value at $\tau = 0$, (37.77) shows no such thing for $R_{XY}(\tau)$. Its maximum value can be anywhere. \square

Example 37.6. Let $X(t) = \sin(2\pi t)$ and $Y(t) = \cos(2\pi t)$, and remember that nothing is more similar to a signal than itself. Obviously, $R_{XY}(\tau)$ will have a maximum when the two sinusoids are in phase, i.e. for $\tau = \frac{3}{4} + 2k\pi$, $k \in \mathbb{Z}$. When $\tau = 0$, the sinusoids are not in phase, and there is no maximum. See Figure 37.10. \square

Estimating a correlation is similar to estimating an autocorrelation, but care must be taken since $R_{XY}(\tau)$ is not even. The same comments apply on how τ cannot get close to t_{final} , or ν close to n :

- In the continuous case, for measurements from $t = 0$ to $t = t_{\text{final}}$:

$$R_{XY}(\tau) = \frac{1}{t_{\text{final}} - \tau} \int_0^{t_{\text{final}} - \tau} x(t)y(t+\tau) dt, \quad \tau \geq 0 \tag{37.81}$$

$$R_{XY}(\tau) = \frac{1}{t_{\text{final}} - \tau} \int_{-\tau}^{t_{\text{final}}} x(t)y(t+\tau) dt, \quad \tau \leq 0 \tag{37.82}$$

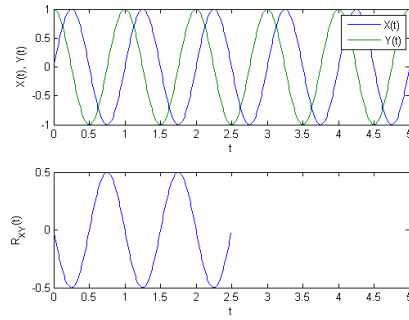


Figure 37.10: Correlation of $X(t) = \sin(2\pi t)$ and $Y(t) = \cos(2\pi t)$. When $\tau = \frac{3}{4}$, signals $X(t)$ and $Y(t + \frac{3}{4})$ are in phase, and their correlation has a maximum.

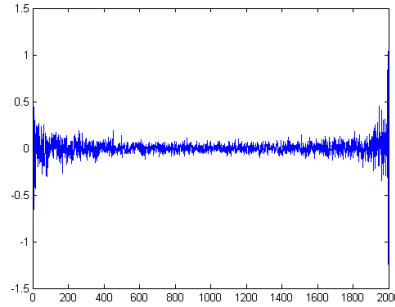


Figure 37.11: Correlation of two independent random signals with normal distribution.

- In discrete time with sampling time T_s , for $n+1$ measurements from $t = 0$ to $t = T_s n$:

$$R_{XY}(\nu) = \frac{1}{n - \nu + 1} \sum_{k=0}^{n-\nu} X_k Y_{k+\nu} \quad \nu \geq 0 \quad (37.83)$$

$$R_{XY}(\nu) = \frac{1}{n - \nu + 1} \sum_{k=-\nu}^n X_k Y_{k+\nu} \quad \nu \leq 0 \quad (37.84)$$

- If $\tau \ll t_{\text{anal}}$, or $\nu \ll n$, (37.81)–(37.84) can be approximated with biased estimations as in (37.41)–(37.42).
- Adapting the MATLAB code from Example 37.5 for the correlation of two signals is obvious.

MATLAB function *xcorr*

- The correlation can also be found in MATLAB using function `xcorr` (the x stands for *cross*), using option `'unbiased'`, as in the next example. Option `'biased'` returns the biased estimation as in (37.41)–(37.42). Of course, this function can also be used for the autocorrelation, if given the same signal twice.

Example 37.7. The correlation of two independent random signals with normal distribution can be found as

```
figure,plot(xcorr(randn(1,1000),randn(1,1000),'unbiased'))
```

and is shown in Figure 37.11. There are 1000 points in each signal; `xcorr` returns 1999 points and $\tau = 0$ is found halfway through, in the 1000th point. The result should be zero everywhere, and is in fact close to zero save for large values of $|\tau|$. That is because very few points are left to calculate the correlation; never forget that results can only be trusted when $|\tau| \ll t_{\text{anal}}$. \square

Covariance

Definition 37.16. The **covariance** $C_{XY}(t_1, t_2)$ of two stochastic processes is the correlation of their deviations from the mean:

$$C_{XY}(t_1, t_2) = E \left[(X(t_1) - E[X]) (Y(t_2) - E[Y]) \right] \quad \square \quad (37.85)$$

Remark 37.15. Since the covariance is a correlation, then

- for stationary processes it depends only on the time delay τ :

$$C_{XY}(\tau) = E\left[(X(t) - E[X])(Y(t + \tau) - E[Y])\right] \quad (37.86)$$

- it is not in general even, but

$$C_{XY}(\tau) = C_{YX}(-\tau) \quad (37.87)$$

- it is estimated numerically in the same way. □

Lemma 37.3.

Relation between correlation and covariance

$$\begin{aligned} C_{XY}(\tau) &= E\left[(X(t) - E[X])(Y(t + \tau) - E[Y])\right] \\ &= E[X(t)Y(t + \tau) - X(t)E[Y] - E[X]Y(t + \tau) + E[X]E[Y]] \\ &= E[X(t)Y(t + \tau)] - E[X(t)E[Y]] - E[E[X]Y(t + \tau)] + E[E[X]E[Y]] \\ &= \underbrace{E[X(t)Y(t + \tau)]}_{R_{XY}(\tau)} - E[X]E[Y] - E[X]E[Y] + E[X]E[Y] \\ &= R_{XY}(\tau) - E[X]E[Y] \end{aligned} \quad (37.88)$$

Corollary 37.5. If either $\bar{X} = 0$ or $\bar{Y} = 0$, then $C_{XY}(\tau) = R_{XY}(\tau)$. □

Covariance is bounded

Theorem 37.9.

$$-\sigma_X\sigma_Y \leq C_{XY}(\tau) \leq \sigma_X\sigma_Y, \quad \forall \tau \quad (37.89)$$

Proof. Again, the covariance is a correlation:

$$C_{XY}(\tau) = E\left[(X(t) - E[X])(Y(t + \tau) - E[Y])\right] = R_{(X-\bar{X})(Y-\bar{Y})}(\tau) \quad (37.90)$$

Thus, it verifies (37.77):

$$-\sqrt{R_{X-\bar{X}}(0)R_{Y-\bar{Y}}(0)} \leq \underbrace{R_{(X-\bar{X})(Y-\bar{Y})}(\tau)}_{C_{XY}(\tau)} \leq \sqrt{R_{X-\bar{X}}(0)R_{Y-\bar{Y}}(0)} \quad (37.91)$$

(37.38) shows that $R_{X-\bar{X}}(0) = \sigma_X^2$, so

$$-\sqrt{\sigma_X^2\sigma_Y^2} \leq C_{XY}(\tau) \leq \sqrt{\sigma_X^2\sigma_Y^2} \quad (37.92)$$

and the result follows immediately. □

Corollary 37.6. From (37.88) and (37.89),

Correlation is bounded

$$E[X]E[Y] - \sigma_X\sigma_Y \leq \underbrace{C_{XY}(\tau) + E[X]E[Y]}_{R_{XY}(\tau)} \leq E[X]E[Y] + \sigma_X\sigma_Y \quad \square \quad (37.93)$$

Correlation (and covariance) can be normalised just like the autocorrelation (and the autocovariance). *Correlation coefficient*

Definition 37.17. The **correlation coefficient** $\rho_{XY}(\tau)$, that verifies $-1 \leq \rho_{XY}(\tau) \leq 1$, is given by

$$\rho_{XY}(\tau) = \frac{R_{XY}(\tau) - \bar{X}\bar{Y}}{\sigma_X\sigma_Y} = \frac{C_{XY}(\tau)}{\sigma_X\sigma_Y} \quad \square \quad (37.94)$$

Remark 37.16. A scalar ρ_{XY} is sometimes given as correlation coefficient. This is understood to be $\rho_{XY}(0)$. □

The correlation coefficient is returned by MATLAB function `xcorr` using option `'coeff'`.

Matrices with the values of the correlation

Correlation and covariance matrices

$$\begin{aligned}
\mathbf{R}_{XY} &= \begin{bmatrix} R_{XY}(0) & R_{XY}(1) & R_{XY}(2) & \cdots & R_{XY}(n) \\ R_{XY}(-1) & R_{XY}(0) & R_{XY}(1) & \cdots & R_{XY}(n-1) \\ R_{XY}(-2) & R_{XY}(-1) & R_{XY}(0) & \cdots & R_{XY}(n-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R_{XY}(-n) & R_{XY}(-n+1) & R_{XY}(-n+2) & \cdots & R_{XY}(0) \end{bmatrix} \\
&= E \begin{bmatrix} X_0Y_0 & X_0Y_1 & X_0Y_2 & \cdots & X_0Y_n \\ X_1Y_0 & X_1Y_1 & X_1Y_2 & \cdots & X_1Y_n \\ X_2Y_0 & X_2Y_1 & X_2Y_2 & \cdots & X_2Y_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_nY_0 & X_nY_1 & X_nY_2 & \cdots & X_nY_n \end{bmatrix} \\
&= E \begin{bmatrix} X(0)Y(0) & X(0)Y(T_s) & X(0)Y(2T_s) & \cdots & X(0)Y(nT_s) \\ X(T_s)Y(0) & X(T_s)Y(T_s) & X(T_s)Y(2T_s) & \cdots & X(T_s)Y(nT_s) \\ X(2T_s)Y(0) & X(2T_s)Y(T_s) & X(2T_s)Y(2T_s) & \cdots & X(2T_s)Y(nT_s) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X(nT_s)Y(0) & X(nT_s)Y(T_s) & X(nT_s)Y(2T_s) & \cdots & X(nT_s)Y(nT_s) \end{bmatrix} \\
&= E [\mathbf{XY}^T] \tag{37.95}
\end{aligned}$$

and the covariance

$$\begin{aligned}
\mathbf{C}_{XY} &= \begin{bmatrix} C_{XY}(0) & C_{XY}(1) & C_{XY}(2) & \cdots & C_{XY}(n) \\ C_{XY}(-1) & C_{XY}(0) & C_{XY}(1) & \cdots & C_{XY}(n-1) \\ C_{XY}(-2) & C_{XY}(-1) & C_{XY}(0) & \cdots & C_{XY}(n-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{XY}(-n) & C_{XY}(-n+1) & C_{XY}(-n+2) & \cdots & C_{XY}(0) \end{bmatrix} \\
&= \sigma_X \sigma_Y \underbrace{\begin{bmatrix} \rho_{XY}(0) & \rho_{XY}(1) & \rho_{XY}(2) & \cdots & \rho_{XY}(n) \\ \rho_{XY}(-1) & \rho_{XY}(0) & \rho_{XY}(1) & \cdots & \rho_{XY}(n-1) \\ \rho_{XY}(-2) & \rho_{XY}(-1) & \rho_{XY}(0) & \cdots & \rho_{XY}(n-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{XY}(-n) & \rho_{XY}(-n+1) & \rho_{XY}(-n+2) & \cdots & \rho_{XY}(0) \end{bmatrix}}_{\rho_{XY}} \\
&= E \begin{bmatrix} (X_0 - \bar{X})(Y_0 - \bar{Y}) & (X_0 - \bar{X})(Y_1 - \bar{Y}) & (X_0 - \bar{X})(Y_2 - \bar{Y}) & \cdots & (X_0 - \bar{X})(Y_n - \bar{Y}) \\ (X_1 - \bar{X})(Y_0 - \bar{Y}) & (X_1 - \bar{X})(Y_1 - \bar{Y}) & (X_1 - \bar{X})(Y_2 - \bar{Y}) & \cdots & (X_1 - \bar{X})(Y_n - \bar{Y}) \\ (X_2 - \bar{X})(Y_0 - \bar{Y}) & (X_2 - \bar{X})(Y_1 - \bar{Y}) & (X_2 - \bar{X})(Y_2 - \bar{Y}) & \cdots & (X_2 - \bar{X})(Y_n - \bar{Y}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (X_n - \bar{X})(Y_0 - \bar{Y}) & (X_n - \bar{X})(Y_1 - \bar{Y}) & (X_n - \bar{X})(Y_2 - \bar{Y}) & \cdots & (X_n - \bar{X})(Y_n - \bar{Y}) \end{bmatrix} \\
&= E \begin{bmatrix} (X(0) - \bar{X})(Y(0) - \bar{Y}) & (X(0) - \bar{X})(Y(T_s) - \bar{Y}) & (X(0) - \bar{X})(Y(2T_s) - \bar{Y}) & \cdots & (X(0) - \bar{X})(Y(nT_s) - \bar{Y}) \\ (X(T_s) - \bar{X})(Y(0) - \bar{Y}) & (X(T_s) - \bar{X})(Y(T_s) - \bar{Y}) & (X(T_s) - \bar{X})(Y(2T_s) - \bar{Y}) & \cdots & (X(T_s) - \bar{X})(Y(nT_s) - \bar{Y}) \\ (X(2T_s) - \bar{X})(Y(0) - \bar{Y}) & (X(2T_s) - \bar{X})(Y(T_s) - \bar{Y}) & (X(2T_s) - \bar{X})(Y(2T_s) - \bar{Y}) & \cdots & (X(2T_s) - \bar{X})(Y(nT_s) - \bar{Y}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (X(nT_s) - \bar{X})(Y(0) - \bar{Y}) & (X(nT_s) - \bar{X})(Y(T_s) - \bar{Y}) & (X(nT_s) - \bar{X})(Y(2T_s) - \bar{Y}) & \cdots & (X(nT_s) - \bar{X})(Y(nT_s) - \bar{Y}) \end{bmatrix} \\
&= E [(\mathbf{X} - \bar{\mathbf{X}})(\mathbf{Y} - \bar{\mathbf{Y}})^T] \tag{37.96}
\end{aligned}$$

can be found as in the case of \mathbf{R}_X and \mathbf{C}_X . But, unlike them, matrices \mathbf{R}_{XY} and \mathbf{C}_{XY} are not symmetric. The matrix denoted by ρ_{XY} in (37.96) is the correlation coefficient matrix.

37.4 Operations with stochastic processes

Theorem 37.10. Let $X(t)$ and $Y(t)$ be stochastic processes with correlation coefficient $\rho_{XY} = \rho_{XY}(0)$. Then

$$E[X(t) + Y(t)] = E[X(t)] + E[Y(t)] \tag{37.97}$$

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\rho_{XY}\sigma_X\sigma_Y \tag{37.98}$$

$$E[X(t) - Y(t)] = E[X(t)] - E[Y(t)] \tag{37.99}$$

$$\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2 - 2\rho_{XY}\sigma_X\sigma_Y \tag{37.100}$$

$$E[X(t)Y(t)] = E[X(t)]E[Y(t)] + \rho_{XY}\sigma_X\sigma_Y \tag{37.101}$$

Proof. (37.97) and (37.99) have already been used and can be proved from the fact that the limit of the mean is the expected value, as stated in (37.21). Assuming continuous time,

$$\begin{aligned} E[X(t) \pm Y(t)] &= \lim_{t_{\text{final}} \rightarrow +\infty} \frac{1}{t_{\text{final}}} \int_0^{t_{\text{final}}} x(t) \pm y(t) dt \\ &= \lim_{t_{\text{final}} \rightarrow +\infty} \frac{1}{t_{\text{final}}} \int_0^{t_{\text{final}}} x(t) dt \pm \lim_{t_{\text{final}} \rightarrow +\infty} \frac{1}{t_{\text{final}}} \int_0^{t_{\text{final}}} y(t) dt = E[X(t)] \pm E[Y(t)] \end{aligned} \quad (37.102)$$

The discrete time case is similar.

(37.101) is an obvious consequence of definition (37.94) when $\tau = 0$:

$$\rho_{XY} = \frac{\overbrace{E[X(t)Y(t)]}^{R_{XY}(0)} - \bar{X}\bar{Y}}{\sigma_X\sigma_Y} \quad (37.103)$$

(37.98) and (37.100) are obtained from

$$\begin{aligned} E[(X(t) \pm Y(t))^2] &= E[(X(t))^2 \pm 2X(t)Y(t) + (Y(t))^2] \\ &= \underbrace{E[(X(t))^2]}_{E[X(t)]^2 + \sigma_X^2} \pm 2 \underbrace{E[X(t)Y(t)]}_{E[X(t)]E[Y(t)] + \rho_{XY}\sigma_X\sigma_Y} + \underbrace{E[(Y(t))^2]}_{E[Y(t)]^2 + \sigma_Y^2} \end{aligned} \quad (37.104)$$

where we made use of (37.24) and (37.101). Thus

$$\begin{aligned} E[(X(t) \pm Y(t))^2] &= E[X(t)]^2 + \sigma_X^2 \pm 2E[X(t)]E[Y(t)] \pm 2\rho_{XY}\sigma_X\sigma_Y + E[Y(t)]^2 + \sigma_Y^2 \\ &= \sigma_X^2 + \sigma_Y^2 \pm 2\rho_{XY}\sigma_X\sigma_Y + \underbrace{(E[X(t)] \pm E[Y(t)])^2}_{(E[X(t) \pm Y(t)])^2} \end{aligned} \quad (37.105)$$

Applying (37.24) once more,

$$E[(X(t) \pm Y(t))^2] = (E[X(t) \pm Y(t)])^2 + \underbrace{\sigma_X^2 + \sigma_Y^2 \pm 2\rho_{XY}\sigma_X\sigma_Y}_{\sigma_{X \pm Y}^2} \quad (37.106)$$

Example 37.8. The following commands show what happens for two uncorrelated signals with normal distribution:

```
>> X = randn(1,10000);
>> Y = randn(1,10000);
>> rho = xcorr(X,Y,0,'coeff')
rho =
0.0068
>> % this is the correlation coefficient for tau=0; it should be zero
>> meanX = mean(X), meanY = mean(Y) % these should be zero
meanX =
-0.0078
meanY =
0.0139
>> varX = var(X), varY = var(Y) % these should be one
varX =
1.0198
varY =
1.0156
>> S = X+Y;
>> mean(S), meanX+meanY % these should be the same
ans =
0.0061
ans =
0.0061
```

```

>> var(S), varX+varY+2*rho*sqrt(varX*varY) % these should be the same
ans =
2.0495
ans =
2.0493
>> D = X-Y;
>> mean(D), meanX-meanY % these should be the same
ans =
-0.0216
ans =
-0.0216
>> var(D), varX+varY-2*rho*sqrt(varX*varY) % these should be the same
ans =
2.0212
ans =
2.0214
>> P = X.*Y;
>> mean(P), meanX*meanY+rho*sqrt(varX*varY) % these should be the same
ans =
0.0070
ans =
0.0069

```

□

Example 37.9. Variables X and S of the previous example are correlated, since S was obtained from X .

```

>> rhoXS = xcorr(X,S,0,'coeff')
rhoXS =
0.7102
>> meanS = mean(S); varS = var(S);
>> S2 = X+S;
>> mean(S2), mean(X)+mean(S) % these should be the same
ans =
-0.0016
ans =
-0.0016
>> var(S2), varX+varS+2*rhoXS*sqrt(varX*varS) % these should be the same
ans =
5.1231
ans =
5.1229

```

□

Glossary

E o senhor extraterrestre
viu-se um pouco atrapalhado.
Quis falar mas disse “pi”,
estava mal sintonizado.

Carlos PAIÃO (1957 — †1988), *O Senhor Extraterrestre* (1982)

autocorrelation autocorrelação
autocorrelation coefficient coeficiente de autocorrelação
autocovariance autocovariância
bias viés
biased enviesado
Brownian motion movimento Browniano
correlation correlação
correlation coefficient coeficiente de correlação
covariance covariância
cross correlation correlação cruzada

cumulative distribution function função de distribuição acumulada
ergodicity ergodicidade
ergodic process processo ergódico
expected value valor esperado
joint probability distribution distribuição conjunta de probabilidades
mean média
probability distribution distribuição de probabilidades
probability distribution function função densidade de probabilidade
probability mass function função de probabilidade
pseudo-random number generator gerador de números pseudo-aleatórios
quasi-stationarity quase-estacionaridade
random walk passeio aleatório
standard deviation desvio padrão
stationarity estacionaridade
stochasticity estocasticidade
stochastic process processo estocástico
variance variância

Exercises

1. Question.

Chapter 38

Spectral density

But what was peculiar about it was its colour. It was an entirely new colour—not a new shade or combination, but a new primary colour, as vivid as blue, red, or yellow, but quite different. When he inquired, she told him that it was known as *ulfire*. Presently he met with a second new colour. This she designated *jale*. The sense-impressions caused in Maskull by these two additional primary colors can only be vaguely hinted at by analogy. Just as blue is delicate and mysterious, yellow clear and unsubtle, and red sanguine and passionate, so he felt *ulfire* to be wild and painful, and *jale* dreamlike, feverish, and voluptuous.

David LINDSAY (1876 — †1945), *A voyage to Arcturus* (1920), 6

In the last chapter, stochastic processes and systems were characterised using different statistical properties. These were often functions of time. Among them were the autocorrelation and the crosscorrelation; assuming ergodicity, they depended on a time difference.

The bilateral Fourier transforms of the autocorrelation and the crosscorrelation, which of course depend on frequency, turn out to be very important functions. To see why, first we must study the Fourier transform better.

38.1 The bilateral Fourier transform

Definition 38.1. The bilateral Fourier transform is the Fourier transform, as in (2.87), corresponding to the bilateral Laplace transform (2.2); that is to say,

$$\mathcal{F}[f(t)] = \int_{-\infty}^{+\infty} f(t)e^{-j\omega t} dt \quad \square \quad (38.1)$$

We will need an explicit expression for the inverse Fourier transform, which we already met in Example 2.25. To find it, we first need a result about function $\delta(t)$, which we met in Remark 10.2.

Integral form of $\delta(t)$

Lemma 38.1.

$$\int_{-\infty}^{+\infty} e^{j\omega t} d\omega = 2\pi\delta(t) \quad (38.2)$$

Proof.

$$\begin{aligned} \int_{-\infty}^{+\infty} e^{j\omega t} d\omega &= \int_{-\infty}^0 e^{j\omega t} d\omega + \int_0^{+\infty} e^{j\omega t} d\omega \\ &= \int_0^{+\infty} e^{-j\omega t} d\omega + \int_0^{+\infty} e^{j\omega t} d\omega \end{aligned} \quad (38.3)$$

These integrals have integrands which are limited but do not vanish at infinity, since

$$\int_0^{+\infty} e^{-j\omega t} d\omega + \int_0^{+\infty} e^{j\omega t} d\omega = \int_0^{+\infty} (\cos(-\omega t) + j \sin(-\omega t)) d\omega + \int_0^{+\infty} (\cos \omega t + j \sin \omega t) d\omega \quad (38.4)$$

It is possible to limit them with a subterfuge:

$$\begin{aligned}
\int_{-\infty}^{+\infty} e^{j\omega t} d\omega &= \lim_{\varepsilon \rightarrow 0^+} \left(\int_0^{+\infty} \underbrace{e^{-j\omega t}}_{\text{limited}} \underbrace{e^{-\omega\varepsilon}}_{\text{vanishes at } +\infty} d\omega + \int_0^{+\infty} \underbrace{e^{j\omega t}}_{\text{limited}} \underbrace{e^{-\omega\varepsilon}}_{\text{vanishes at } +\infty} d\omega \right) \\
&= \lim_{\varepsilon \rightarrow 0^+} \left(\int_0^{+\infty} e^{-j\omega t} e^{-j\omega(-j\varepsilon)} d\omega + \int_0^{+\infty} e^{j\omega t} e^{j\omega(j\varepsilon)} d\omega \right) \\
&= \lim_{\varepsilon \rightarrow 0^+} \left(\int_0^{+\infty} \underbrace{e^{-j(t-j\varepsilon)\omega}}_{\text{vanishes at } +\infty} d\omega + \int_0^{+\infty} \underbrace{e^{j(t+j\varepsilon)\omega}}_{\text{vanishes at } +\infty} d\omega \right) \\
&= \lim_{\varepsilon \rightarrow 0^+} \left(\left[\frac{1}{-j(t-j\varepsilon)} e^{-j(t-j\varepsilon)\omega} \right]_{\omega=0}^{+\infty} + \left[\frac{1}{j(t+j\varepsilon)} e^{j(t+j\varepsilon)\omega} \right]_{\omega=0}^{+\infty} \right) \\
&= \lim_{\varepsilon \rightarrow 0^+} \left(\frac{j}{t-j\varepsilon} \underbrace{(0 - e^0)}_{-1} + \frac{-j}{t+j\varepsilon} \underbrace{(0 - e^0)}_{-1} \right) \\
&= \lim_{\varepsilon \rightarrow 0^+} \left(\frac{j}{t+j\varepsilon} - \frac{j}{t-j\varepsilon} \right) = \lim_{\varepsilon \rightarrow 0^+} \frac{jt + \varepsilon - jt + \varepsilon}{t^2 + \varepsilon^2} = \lim_{\varepsilon \rightarrow 0^+} \frac{2\varepsilon}{t^2 + \varepsilon^2} \tag{38.5}
\end{aligned}$$

Consequently,

- if $t \neq 0$,

$$\int_{-\infty}^{+\infty} e^{j\omega t} d\omega = \lim_{\varepsilon \rightarrow 0^+} \frac{2\varepsilon}{t^2 + \varepsilon^2} = 0 \tag{38.6}$$

- if $t = 0$,

$$\int_{-\infty}^{+\infty} e^{j\omega t} d\omega = \lim_{\varepsilon \rightarrow 0^+} \frac{2\varepsilon}{\varepsilon^2} = \lim_{\varepsilon \rightarrow 0^+} \frac{2}{\varepsilon} = +\infty \tag{38.7}$$

This is the same as (10.15)–(10.16), so all that is left is to see if a relation similar to (10.17)–(10.18) holds. Thus we calculate

$$\int_{-\infty}^{+\infty} \frac{2\varepsilon}{t^2 + \varepsilon^2} dt = 2 \int_{-\infty}^{+\infty} \frac{1}{\left(\frac{t}{\varepsilon}\right)^2 + 1} \frac{dt}{\varepsilon} \tag{38.8}$$

Using variable change

$$\tau = \frac{t}{\varepsilon} \tag{38.9}$$

$$t = -\infty \Rightarrow \tau = -\infty \tag{38.10}$$

$$t = +\infty \Rightarrow \tau = +\infty \tag{38.11}$$

$$d\tau = \frac{dt}{\varepsilon} \tag{38.12}$$

we obtain

$$\int_{-\infty}^{+\infty} \frac{2\varepsilon}{t^2 + \varepsilon^2} dt = 2 \int_{-\infty}^{+\infty} \frac{1}{\tau^2 + 1} d\tau = 2 [\arctan \tau]_{\tau=-\infty}^{+\infty} = 2 \left(\frac{\pi}{2} - \left(-\frac{\pi}{2} \right) \right) = 2\pi \tag{38.13}$$

Since this integral is 2π larger than (10.17), the result follows. \square

We can now find an explicit expression for $\mathcal{F}^{-1}[F(s)]$.

Inverse Fourier transform **Theorem 38.1.** Let $f(t)$ be a function with bilateral Fourier transform $\mathcal{F}[f(t)] = F(s)$. Then

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(j\omega) e^{j\omega t} d\omega \tag{38.14}$$

Proof. This can be shown replacing the definition of the bilateral Fourier transform (38.1) in (38.14), and applying (38.2) when needed:

$$\begin{aligned}
 f(t) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(\tau) e^{-j\omega\tau} d\tau e^{j\omega t} d\omega \\
 &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(\tau) e^{j\omega(t-\tau)} d\omega d\tau \\
 &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} f(\tau) \underbrace{\int_{-\infty}^{+\infty} e^{j\omega(t-\tau)} d\omega}_{2\pi\delta(t-\tau)} d\tau \\
 &= \int_{-\infty}^{+\infty} f(\tau) \delta(t-\tau) d\tau
 \end{aligned} \tag{38.15}$$

We now apply (10.21), and, since $t - \tau = 0 \Leftrightarrow \tau = t$, conclude that this integral is in fact $f(t)$. \square

With the inverse Fourier transform, we can prove two important results, both known either as **Parserval's theorem** or **Plancherel theorem**.

Theorem 38.2. Let $f(t)$ be a function with bilateral Fourier transform $\mathcal{F}[f(t)] = F(j\omega)$. Then

$$\int_{-\infty}^{+\infty} |f(t)|^2 dt = \int_{-\infty}^{+\infty} f(t) \overline{f(t)} dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |F(j\omega)|^2 d\omega = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(j\omega) \overline{F(j\omega)} d\omega \tag{38.16}$$

Proof. The proof uses (38.14) to obtain the complex conjugate of $f(t)$:

$$\begin{aligned}
 \int_{-\infty}^{+\infty} f(t) \overline{f(t)} dt &= \int_{-\infty}^{+\infty} f(t) \underbrace{\frac{1}{2\pi} \int_{-\infty}^{+\infty} F(-j\omega) e^{-j\omega t} d\omega}_{\frac{1}{2\pi} \int_{-\infty}^{+\infty} F(j\omega) e^{j\omega t} d\omega} dt \\
 &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(t) F(-j\omega) e^{-j\omega t} dt d\omega \\
 &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(-j\omega) \underbrace{\int_{-\infty}^{+\infty} f(t) e^{-j\omega t} dt}_{F(j\omega)} d\omega \square
 \end{aligned} \tag{38.17}$$

Theorem 38.3. Let $f(t)$ and $g(t)$ be functions with bilateral Fourier transforms $\mathcal{F}[f(t)] = F(j\omega)$ and $\mathcal{F}[g(t)] = G(j\omega)$. Then *Generalised Parseval's theorem*

$$\int_{-\infty}^{+\infty} f(t) \overline{g(t)} dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(j\omega) \overline{G(j\omega)} d\omega \tag{38.18}$$

Proof. The proof uses (38.14) to obtain both $f(t)$ and the complex conjugate of $g(t)$:

$$\int_{-\infty}^{+\infty} f(t) \overline{g(t)} dt = \int_{-\infty}^{+\infty} \left(\frac{1}{2\pi} \int_{-\infty}^{+\infty} F(j\omega) e^{j\omega t} d\omega \right) \underbrace{\left(\frac{1}{2\pi} \int_{-\infty}^{+\infty} G(-j\Omega) e^{-j\Omega t} d\Omega \right)}_{\left(\frac{1}{2\pi} \int_{-\infty}^{+\infty} G(j\Omega) e^{j\Omega t} d\Omega \right)} dt \tag{38.19}$$

Now we need (38.2) switching variables t and ω :

$$\int_{-\infty}^{+\infty} e^{jt\omega} dt = 2\pi\delta(\omega) \tag{38.20}$$

Replacing this in (38.19),

$$\int_{-\infty}^{+\infty} f(t) \overline{g(t)} dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(j\omega) G(-j\Omega) \underbrace{\frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{j(\omega-\Omega)t} dt}_{\delta(\omega-\Omega)} d\Omega d\omega \tag{38.21}$$

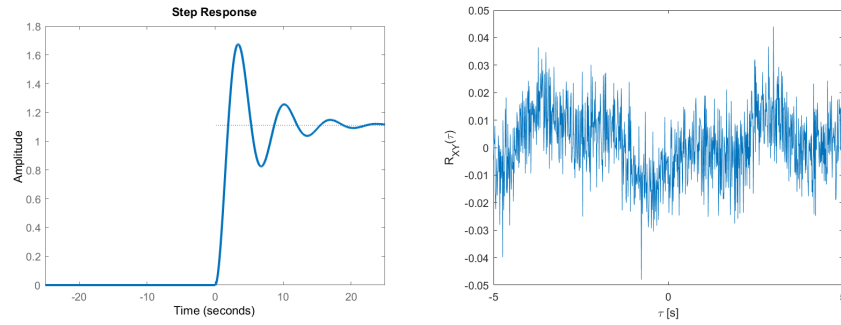


Figure 38.1: Left: a function in time which is zero when $t < 0$, for which a unilateral transform makes sense. Right: a cross-correlation of two signals, that exists for both $\tau > 0$ and $\tau < 0$, and for which only a bilateral transform makes sense.

Applying (10.21) and taking into account that $\omega - \Omega = 0 \Leftrightarrow \Omega = \omega$,

$$\int_{-\infty}^{+\infty} f(t)\overline{g(t)} dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(j\omega) \underbrace{\int_{-\infty}^{+\infty} G(-j\Omega)\delta(\omega - \Omega) d\Omega}_{G(-j\omega)} d\omega \quad (38.22)$$

Remark 38.1. (38.16) can be seen as a particular case of (38.18) for $f(t) = g(t)$. \square

38.2 Definition and properties of the spectral density

PSD **Definition 38.2.** The **power spectral density** (PSD) $S_X(j\omega)$ of a signal $X(t)$ is the bilateral Fourier transform of its autocorrelation:

$$\begin{aligned} S_X(j\omega) &= \mathcal{F} [R_X(\tau)] \\ &= \int_{-\infty}^{+\infty} R_X(\tau)e^{-j\omega\tau} d\tau \quad \square \end{aligned} \quad (38.23)$$

CSD **Definition 38.3.** The **cross-spectral density** (CSD) $S_{XY}(j\omega)$ of two signals $X(t)$ and $Y(t)$ is the bilateral Fourier transform of their correlation:

$$\begin{aligned} S_{XY}(j\omega) &= \mathcal{F} [R_{XY}(\tau)] \\ &= \int_{-\infty}^{+\infty} R_{XY}(\tau)e^{-j\omega\tau} d\tau \quad \square \end{aligned} \quad (38.24)$$

The reason why the bilateral transform is used for the PSD and the CSD, while we have used unilateral transforms until now, should be clear:

- We have been using unilateral transforms for time responses, that may not be defined for $t < 0$, or are equal to 0 for $t < 0$. See Figure 38.1.
- $R_X(\tau)$ and $R_{XY}(\tau)$ exist both for $\tau > 0$ and $\tau < 0$.
- $R_{XY}(\tau) \neq R_{XY}(-\tau)$ in the general case, and so values for both $\tau > 0$ and $\tau < 0$ must be considered when calculating the Fourier transform. See Figure 38.1.
- While $R_X(\tau) = R_X(-\tau)$, and thus we might think a unilateral transform would suffice in this case, since R_X is a particular case of R_{XY} when the two signals are the same, it is convenient to always use a bilateral transform.

Remark 38.2. Spectral densities are sometimes given as *Laplace transforms*, not Fourier transforms. The conversion is straightforward:

- From \mathcal{L} to \mathcal{F} :

$$s = j\omega \quad (38.25)$$

$$s^2 = -\omega^2 \quad (38.26)$$

$$s^3 = -j\omega^3 \quad (38.27)$$

$$s^4 = \omega^4 \quad (38.28)$$

$$s^5 = j\omega^5 \quad (38.29)$$

⋮

- From \mathcal{F} to \mathcal{L} :

$$\omega = \frac{s}{j} = \frac{js}{j^2} = -js \quad (38.30)$$

$$\omega^2 = -s^2 \quad (38.31)$$

$$\omega^3 = -\frac{s^3}{j} = js^3 \quad (38.32)$$

$$\omega^4 = s^4 \quad (38.33)$$

$$\omega^5 = \frac{s^5}{j} = -js^5 \quad (38.34)$$

⋮

When given as Fourier transforms, and thus functions of frequency ω , spectral densities are often written as $S_X(\omega)$ and $S_{XY}(\omega)$, omitting the imaginary unit j . \square

The PSD has several important properties.

Theorem 38.4.

$$\overline{X^2(t)} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_X(\omega) d\omega \quad (38.35)$$

Proof. Inverting definition (38.23),

$$\begin{aligned} R_X(\tau) &= \mathcal{F}^{-1}[S_X(\omega)] \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_X(\omega) e^{j\omega\tau} d\omega \end{aligned} \quad (38.36)$$

In particular,

$$R_X(0) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_X(\omega) \underbrace{e^{j\omega 0}}_1 d\omega \quad (38.37)$$

and we know from (37.38) that $R_X(0) = \overline{X^2(t)}$. \square

Theorem 38.5. A PSD is an even function of ω :

$$S_X(\omega) = S_X(-\omega)$$

$$S_X(\omega) = S_X(-\omega) \quad (38.38)$$

Proof. The PSD is a Fourier transform:

$$S_X(-\omega) = \int_{-\infty}^{+\infty} R_X(\tau) e^{-j(-\omega)\tau} d\tau \quad (38.39)$$

Now we apply variable change

$$t = -\tau \Leftrightarrow \tau = -t \quad (38.40)$$

$$\tau = -\infty \Leftrightarrow t = +\infty \quad (38.41)$$

$$\tau = +\infty \Leftrightarrow t = -\infty \quad (38.42)$$

$$d\tau = -dt \quad (38.43)$$

to get

$$S_X(-\omega) = \int_{+\infty}^{-\infty} R_X(-t)e^{j\omega(-t)}(-dt) \quad (38.44)$$

We know from (37.36) that $R_X(-t) = R_X(t)$, and so

$$S_X(-\omega) = \int_{-\infty}^{+\infty} R_X(t)e^{-j\omega t} dt = \mathcal{F}[R_X(t)] = S_X(\omega) \square \quad (38.45)$$

The PSD only has even powers of ω or s

The PSD is real valued

Remark 38.3. The PSD can be rational or irrational. Because it is even, if it is rational, it must be a function of even powers of ω when given as a Fourier transform, or of s when given as a Laplace transform: $\omega^0 = 1$, ω^2 , ω^4 , etc.. There can be no terms in s , s^3 , etc.. These missing odd powers of s are the ones that correspond to imaginary parts in the Fourier transform: $j\omega$, $(j\omega)^3 = -j\omega^3$, $(j\omega)^5 = j\omega^5$, etc.. Thus, the PSD, when given as a Fourier transform, has no imaginary part.

Terms in ω^2 , because of (38.31), correspond to Laplace transforms with pairs of symmetric poles or zeros: one in the left side of the complex plane and one in the right side. The same will happen for terms in ω^4 , ω^6 , ω^{10} , and so on.

We saw that the correlation, unlike the autocorrelation, is not in the general case an even function. Consequently, neither is the CSD, its Fourier transform, an even function in the general case; i.e. it can include odd powers of ω if given as a Fourier transform, or of s if given as a Laplace transform: s , s^3 , etc.. These odd powers of s mean that the CSD, when given as a Fourier transform, has an imaginary part. \square

The CSD is complex valued

Example 38.1. $S_X(\omega) = \frac{10}{\omega^2+1}$ corresponds to $S_X(s) = \frac{10}{-s^2+1} = -\frac{10}{s^2-1}$, with poles ± 1 , which are symmetric. \square

The usefulness of the PSD can be seen from the following results.

Lemma 38.2. The auto-correlation of $X(t) = \sin t$ is

$$R_X(\tau) = \frac{1}{2} \cos \tau \quad (38.46)$$

Proof. We will consider an arbitrarily large final time:

$$\begin{aligned} R_X(\tau) &= \lim_{t_{\text{final}} \rightarrow +\infty} \frac{1}{t_{\text{final}} - \tau} \int_0^{t_{\text{final}} - \tau} \sin(t) \sin(t + \tau) dt \\ &= \lim_{t_{\text{final}} \rightarrow +\infty} \frac{1}{t_{\text{final}} - \tau} \int_0^{t_{\text{final}} - \tau} \sin(t) \left(\sin(t) \cos(\tau) + \cos(t) \sin(\tau) \right) dt \\ &= \lim_{t_{\text{final}} \rightarrow +\infty} \frac{1}{t_{\text{final}} - \tau} \int_0^{t_{\text{final}} - \tau} \sin^2(t) \cos(\tau) dt + \\ &\quad \lim_{t_{\text{final}} \rightarrow +\infty} \frac{1}{t_{\text{final}} - \tau} \int_0^{t_{\text{final}} - \tau} \sin(t) \cos(t) \sin(\tau) dt \\ &= \cos(\tau) \lim_{t_{\text{final}} \rightarrow +\infty} \frac{1}{t_{\text{final}} - \tau} \int_0^{t_{\text{final}} - \tau} \sin^2(t) dt + \\ &\quad \sin(\tau) \lim_{t_{\text{final}} \rightarrow +\infty} \frac{1}{t_{\text{final}} - \tau} \int_0^{t_{\text{final}} - \tau} \frac{1}{2} \sin(2t) dt \end{aligned} \quad (38.47)$$

The last integral is limited; over each period, it has a zero mean: $\int_0^\pi \frac{1}{2} \sin(2t) dt = 0$. Thus, the second limit is zero. As to the first integral, since

$$\cos(2t) = \cos^2 t - \sin^2 t = 1 - \sin^2 t - \sin^2 t = 1 - 2\sin^2 t \quad (38.48)$$

we make

$$\begin{aligned} R_X(\tau) &= \cos(\tau) \lim_{t_{\text{final}} \rightarrow +\infty} \frac{1}{t_{\text{final}} - \tau} \int_0^{t_{\text{final}} - \tau} \left(\frac{1}{2} - \frac{1}{2} \cos(2t) \right) dt \\ &= \cos(\tau) \lim_{t_{\text{final}} \rightarrow +\infty} \frac{1}{t_{\text{final}} - \tau} \left[\frac{t}{2} \right]_0^{t_{\text{final}} - \tau} \\ &\quad - \frac{1}{2} \cos(\tau) \lim_{t_{\text{final}} \rightarrow +\infty} \frac{1}{t_{\text{final}} - \tau} \int_0^{t_{\text{final}} - \tau} \cos(2t) dt \end{aligned} \quad (38.49)$$

Again, the last integral is limited, and thus the limit is zero. We are left with

$$R_X(\tau) = \frac{1}{2} \cos(\tau) \lim_{t_{\text{final}} \rightarrow +\infty} \frac{t_{\text{final}} - \tau - 0}{t_{\text{final}} - \tau} \quad (38.50)$$

and the result follows immediately. \square

Lemma 38.3. The Fourier transform of the complex exponential is

$$\mathcal{F} [e^{jat}] = 2\pi\delta(\omega - a) \quad (38.51)$$

Proof. From (10.22) we know that $\mathcal{L} [\delta(t)] = 1$, and from (24.1) we know that $\mathcal{L} [f(t - \theta)] = F(s)e^{-\theta s}$. Thus $\mathcal{L} [\delta(t - \theta)] = e^{-\theta s}$; the corresponding Fourier transform is

$$\mathcal{F} [\delta(t - \theta)] = e^{-j\omega\theta} \quad (38.52)$$

Using the inverse Fourier transform (38.14),

$$\begin{aligned} \delta(t - \theta) &= \mathcal{F}^{-1} [e^{-j\omega\theta}] = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-j\omega\theta} e^{j\omega t} d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{j\omega(t-\theta)} d\omega \end{aligned} \quad (38.53)$$

We will use this integral to find the Fourier transform of the complex exponential:

$$\begin{aligned} \mathcal{F} [e^{jat}] &= \int_{-\infty}^{+\infty} e^{jat} e^{-j\omega t} dt \\ &= \int_{-\infty}^{+\infty} e^{jt(a-\omega)} dt \end{aligned} \quad (38.54)$$

The integral in (38.54) is the same as that in (38.53), with the following variable changes:

$$\omega \text{ becomes } t \quad (38.55)$$

$$t \text{ becomes } a \quad (38.56)$$

$$\theta \text{ becomes } \omega \quad (38.57)$$

Thus, $\mathcal{F} [e^{jat}] = 2\pi\delta(a - \omega)$, which is zero everywhere, save at $\omega = a$, just like $2\pi\delta(\omega - a)$. \square

Theorem 38.6. The PSD of $X(t) = \sin(at)$ is

PSD of a sinusoid is an impulse at its frequency

$$S_X(\omega) = \frac{\pi}{2} \delta(\omega - a) - \frac{\pi}{2} \delta(\omega + a) \quad (38.58)$$

Proof. (38.46) shows that $S_X(\omega) = \mathcal{F} [\frac{1}{2} \cos(at)]$, and since

$$\begin{cases} e^{jat} &= \cos at + j \sin at \\ e^{-jat} &= \cos at - j \sin at \end{cases} \Rightarrow e^{jat} + e^{-jat} = 2 \cos at \quad (38.59)$$

we can get

$$S_X(\omega) = \frac{1}{4} \mathcal{F} [e^{jat}] + \frac{1}{4} \mathcal{F} [e^{-jat}] \quad (38.60)$$

Applying (38.51), the result follows immediately. \square

When computing the PSD of a sinusoid in practice, it is not an impulse that is found, since that is of course impossible, but rather a significant peak at the frequency of the sinusoid. Likewise, when computing the PSD of a linear combination of sinusoids, peaks will be found at the corresponding frequencies, with heights proportional to the squares of the weights (because the signal is multiplied by itself when computing $R_X(\tau)$). In other words, the PSD of a signal gives us its *spectral content*.

A signal's PSD shows its spectral content

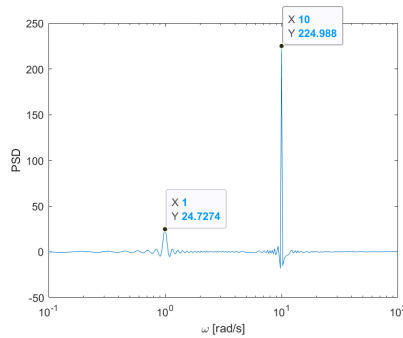


Figure 38.2: Power spectral density of $X = \sin(t) + 3\sin(10t)$, found with the code from Example 38.2.

Example 38.2. The PSD of $X(t) = \sin(t) + 3\sin(10t)$ can be found using a function to compute its autocorrelation

```
function [autocorrelation,tau] = R_X(t,X,m)
% function [autocorrelation,tau] = R_X(t,X,m)
% Finds the autocorrelation of X(t) at m points (default is length(t)/2).
n = length(t)-1;
if length(X)~=n+1, error('t and X must have the same length'), end
if nargin<3, m = floor(n/2); end
autocorrelation = zeros(1,m);
for k = 0 : m-1
    autocorrelation(k+1) = sum( X(1:n-k+1) .* X(k+1:n+1) )/( n-k+1 );
end
autocorrelation = [autocorrelation(end:-1:2) autocorrelation];
tau = [ -t(m:-1:2) 0 t(2:m) ];
```

and a function that computes the PSD using the one above

```
function [S,w]=S_X(t,X,number_w)
% function [S,w]=S_X(t,X,number_w)
% Finds the PSD S(w) of X(t) at number_w frequencies (default 100 per
% decade) in a reasonable range of frequencies.
[autocorrelation,tau] = R_X(t,X);
Ts = mean(diff(t)); % sample time
wUp = floor(log10(pi/Ts));
wLow = ceil(log10(pi/t(end))); if wLow == wUp, wLow = wUp-1; end % at least 1 decade
if nargin == 3
    wvector = logspace(wLow, wUp, number_w);
else
    wvector = logspace(wLow, wUp, (wUp-wLow)*100+1);
end
S = zeros(size(wvector));
for k = 1:length(wvector)
    w = wvector(k);
    S(k) = trapz( autocorrelation .* exp(-1i*w*tau) ) * Ts;
end
if norm(imag(S))/norm(real(S)) > 1e-6
    warning('Imaginary part of S not neglectable, something''s wrong')
end % sanity check
S = real(S);
w = wvector;
```

In this way, commands

```
>> Ts = 0.01; tfinal = 100;
>> t = 0 : Ts : tfinal;
>> X = sin(t) + 3*sin(10*t);
>> [S,w] = S_X(t,X); figure,semilogx(w,S),xlabel('\omega [rad/s]'),ylabel('PSD')
```

result in Figure 38.2.

Notice that:

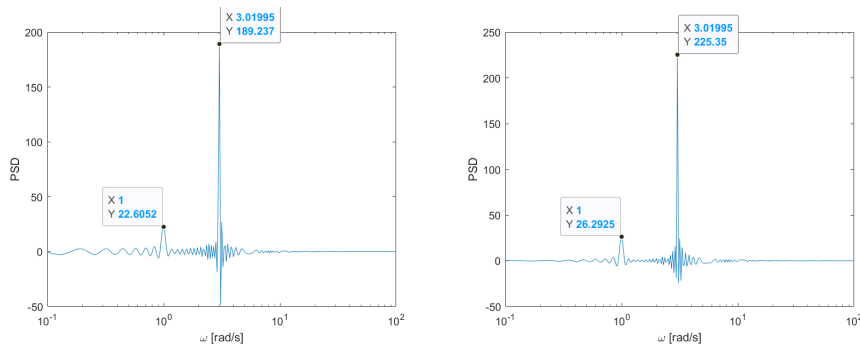


Figure 38.3: Power spectral density of $X = \sin(t) + 3 \sin(3t)$ (left) and $X = \sin(t) + 3 \sin(3.02t)$ (right), found with the code from Example 38.2.

- there are ripples in the PSD around the frequencies of the sinusoids;
- the amplitude of the peaks of $S_X(\omega)$ is very sensitive to the frequencies where it is found. Figure 38.2 shows the results obtained with the code above for signals $X = \sin(t) + 3 \sin(3t)$ and $X = \sin(t) + 3 \sin(3.02t)$. Even though the PSD is being computed for 100 frequencies in each decade, since in the first case there is no precise match for 3 rad/s, the amplitude is lower than in the second case, where the match is far better.

These numerical problems can be minimised, though never completely eliminated, using windows, addressed in the next section. \square

38.3 Numerical computation of the PSD and CSD

MATLAB's function `cpsd(x,y)` returns S_{YX} , since the definition employed switches MATLAB's *command cpsd* the roles of the two functions in relation to (37.81); it can be used to find both the CSD and the PSD. It may receive a third argument as `cpsd(x,y>window)`, *Windows* which requires an important explanation.

The PSD and the CSD are the Fourier transforms of the autocorrelation $R_X(\tau)$ and the cross correlation $R_{XY}(\tau)$, and these are in practice found for data recorded during a finite time interval t_{final} from (37.39) and (37.81). Furthermore, we argued $R_X(\tau)$ and $R_{XY}(\tau)$ cannot be found up to t_{final} or down to $-t_{\text{final}}$, but only for part of that interval; let it be $[-\tau_m, \tau_m]$. Thus, calculations are done as if $R_X(\tau)$ and $R_{XY}(\tau)$ were being multiplied by a sequence of two steps:

Rectangular window

$$S_X(j\omega) = \mathcal{F}[w_r(\tau)R_X(\tau)] \quad (38.61)$$

$$S_{XY}(j\omega) = \mathcal{F}[w_r(\tau)R_{XY}(\tau)] \quad (38.62)$$

$$w_r(\tau) = \begin{cases} 1, & \text{if } -\tau_m \leq \tau \leq \tau_m \\ 0, & \text{otherwise} \end{cases} \quad (38.63)$$

Here, $w_r(\tau)$ is called a **rectangular window**. Notice that this window is implicit in the calculation as it can be carried out in practice. The following result is now needed:

Theorem 38.7. If these Fourier transforms exist,

The Fourier transform of a product is the convolution of the Fourier transforms

$$\mathcal{F}[f(t)g(t)] = \frac{1}{2\pi} F(j\omega) * G(j\omega) \quad (38.64)$$

Proof. Use (38.14) in (38.1) to get

$$\begin{aligned}
 \mathcal{F}[f(t)g(t)] &= \int_{-\infty}^{+\infty} f(t)g(t)e^{-j\omega t} dt = \int_{-\infty}^{+\infty} \overbrace{\frac{1}{2\pi} \int_{-\infty}^{+\infty} F(j\Omega)e^{j\Omega t} d\Omega}^{f(t)} g(t)e^{-j\omega t} dt \\
 &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(j\Omega)g(t)e^{-j(\omega-\Omega)t} dt d\Omega \quad (38.65) \\
 &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(j\Omega) \underbrace{\int_{-\infty}^{+\infty} g(t)e^{-j(\omega-\Omega)t} dt}_{G(j(\omega-\Omega))} d\Omega = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(j\Omega)G(j(\omega-\Omega)) d\Omega
 \end{aligned}$$

This integral is the form that a convolution assumes when a bilateral transform is employed. \square

Thus, when computing a PSD or a CSD, if nothing is done about it, the result will be the convolution of what we wanted with the Fourier transform $W_r(j\omega)$ of a rectangular window $w_r(\tau)$:

$$S_X(j\omega) = \mathcal{F}[w_r(\tau)R_X(\tau)] = W_r(j\omega) * \overbrace{\mathcal{F}[R_X(\tau)]}^{\text{desired PSD}} \quad (38.66)$$

$$S_{XY}(j\omega) = \mathcal{F}[w_r(\tau)R_{XY}(\tau)] = W_r(j\omega) * \overbrace{\mathcal{F}[R_{XY}(\tau)]}^{\text{desired CSD}} \quad (38.67)$$

Leakage

It is this convolution that caused the ripples in Example 38.2, seen in Figures 38.2 and 38.3. Since the effect of each frequency shows up in other frequencies around the correct one, this effect is called **leakage**.

To minimise leakage, rather than trying to undo the effect of this convolution after the calculation is done, it is better to explicitly multiply $R_X(\tau)$ or $R_{XY}(\tau)$ by some window $w(\tau)$ with a more favourable Fourier transform. It turns out that there no ideal window that works in practice for all signals, but different signals require different windows.

Usual windows

Some of the most usual windows, represented in Figures 38.4 and 38.5, are the following. They are given as a function of t , in interval $[0, t_{\text{final}}]$, and are zero outside this range. MATLAB functions to create them are also shown; they receive as argument the number of samples in interval $[0, t_{\text{final}}]$, and are shown in the Figures using command `wvtool`, which shows the window's Fourier transform.

- **Rectangular** window, useful for signals with frequency content in a broadband (MATLAB function `rectwin`):

$$w_r(t) = 1 \quad (38.68)$$

Using no window is using a rectangular window

Remember that this window corresponds, in fact, to doing nothing to the signal. It is, thus, in a sense, a default window.

A Hann window often gives good results

- **Hann** (or Hanning) window, the most used one, useful for signals with frequency content in a narrowband (MATLAB function `hann`):

$$w_n(t) = \sin^2 \frac{\pi t}{t_{\text{final}}} \quad (38.69)$$

- **Hamming** window, useful for signals with closely spaced frequencies (MATLAB function `hamming`):

$$w_m(t) = 0.54 - 0.46 \cos \frac{2\pi t}{t_{\text{final}}} \quad (38.70)$$

MATLAB's function `cpsd` uses Hamming windows by default.

- **Flat top** window, useful for sinusoidal signals when amplitude accuracy is more important than frequency accuracy (MATLAB function `flattopwin`):

$$\begin{aligned}
 w_f(t) &= 0.21557895 - 0.41663158 \cos \frac{2\pi t}{t_{\text{final}}} + 0.277263158 \cos \frac{4\pi t}{t_{\text{final}}} \\
 &\quad - 0.083578947 \cos \frac{6\pi t}{t_{\text{final}}} + 0.006947368 \cos \frac{8\pi t}{t_{\text{final}}} \quad (38.71)
 \end{aligned}$$

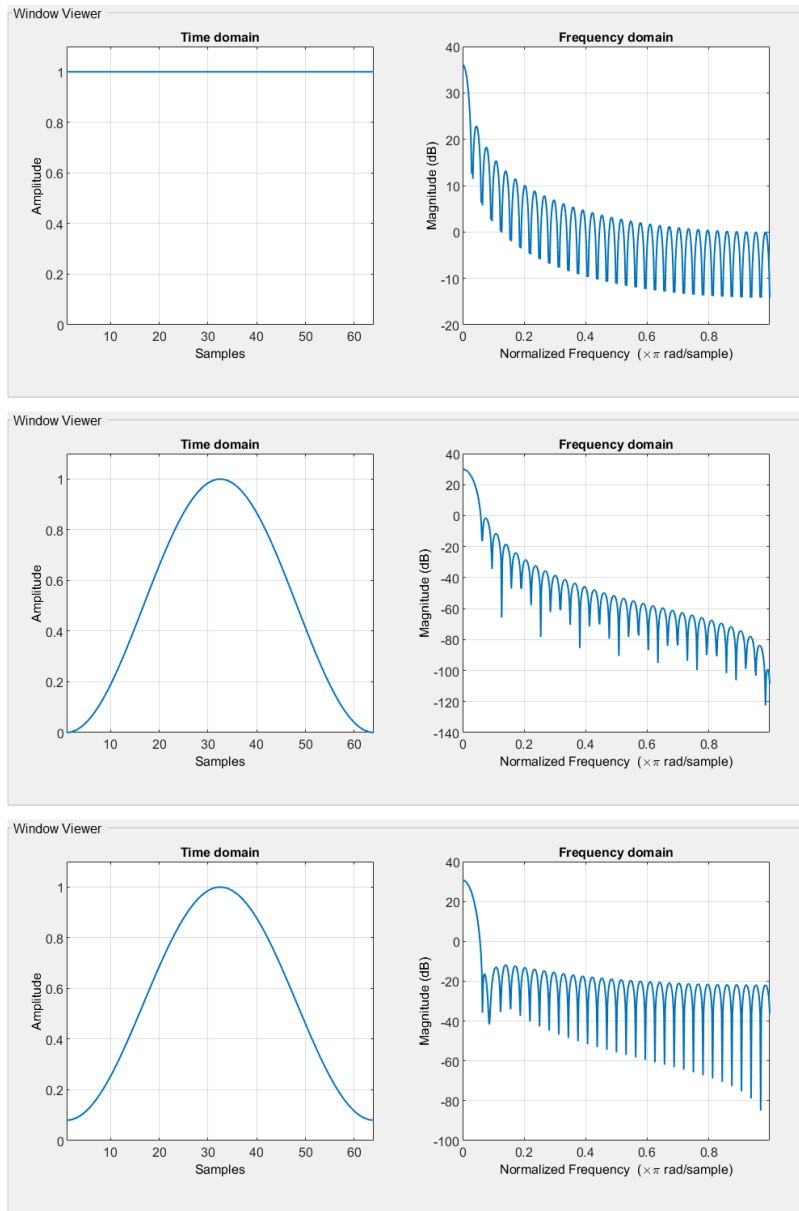


Figure 38.4: Windows, plot for 64 samples between 0 and t_{final} , together with their Fourier transforms. From top to bottom: rectangular window, Hanning window, Hamming window. Continues in Figure 38.5.

- **Bartlett** or triangular window (MATLAB function `bartlett`):

$$w_b(t) = 1 - \left| \frac{t - \frac{t_{\text{final}}}{2}}{\frac{t_{\text{final}}}{2}} \right| \quad (38.72)$$

- **Blackman** window (MATLAB function `blackman`):

$$w_k(t) = 0.42 - 0.5 \cos \frac{2\pi t}{t_{\text{final}}} + 0.08 \cos \frac{4\pi t}{t_{\text{final}}} \quad (38.73)$$

MATLAB's function `cpsd(x,y)` may receive a fourth argument as `cpsd(x,y>window)`'s command `cpds`. To explain it, we need the following result, often used to algorithms to calculate the PSD or the CSD.

Theorem 38.8. The PSD and CSD can be found as

$$\begin{aligned} S_X(j\omega) &= \lim_{t_{\text{final}} \rightarrow +\infty} \frac{1}{2t_{\text{final}}} E[X(j\omega)X(-j\omega)] \\ &= \lim_{t_{\text{final}} \rightarrow +\infty} \frac{1}{2t_{\text{final}}} E[|X(j\omega)|^2] \end{aligned} \quad (38.74)$$

$$S_{XY}(j\omega) = \lim_{t_{\text{final}} \rightarrow +\infty} \frac{1}{2t_{\text{final}}} E[X(j\omega)Y(-j\omega)] \quad (38.75)$$

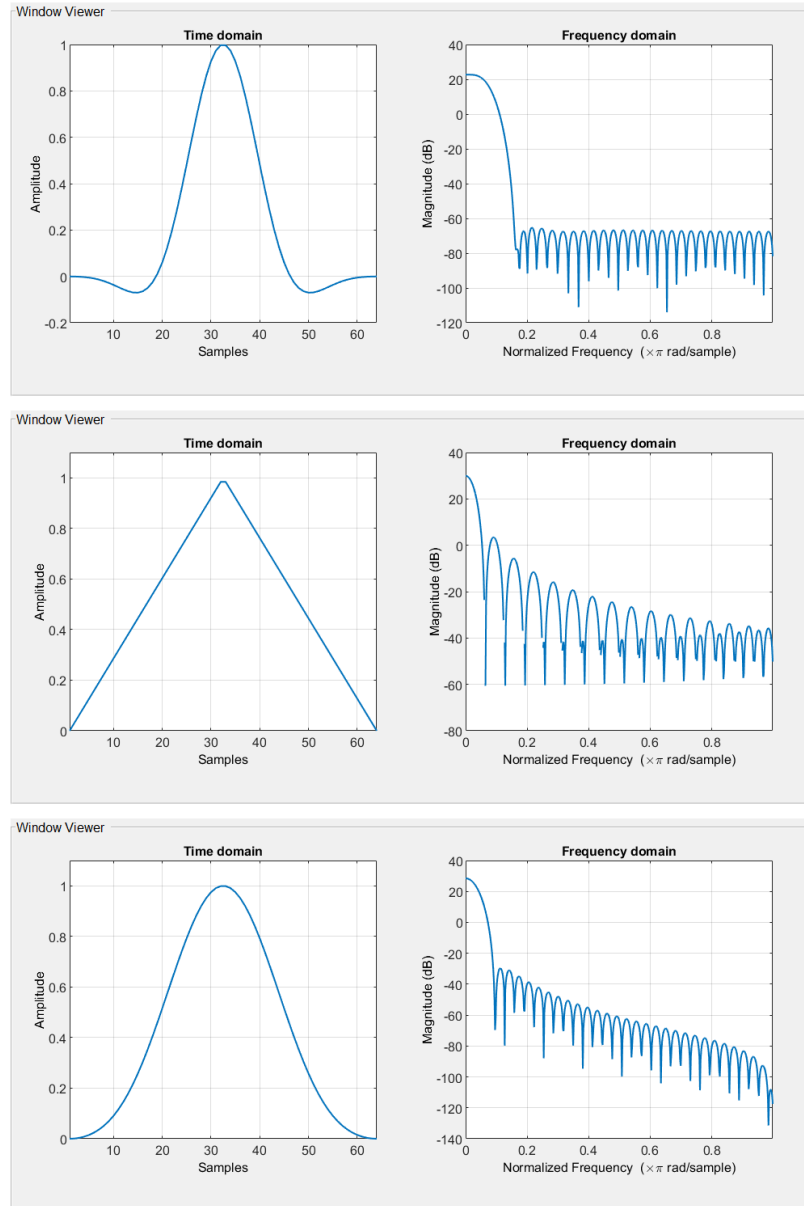


Figure 38.5: Windows, plot for 64 samples between 0 and t_{final} , together with their Fourier transforms. From top to bottom: flat top window, Bartlett window, Blackman window. Continued from Figure 38.4.

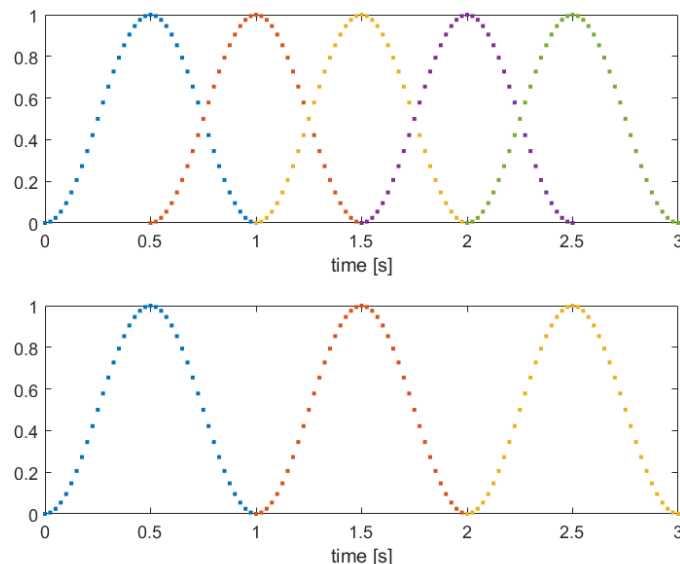


Figure 38.6: Top: five Hanning windows with a 50% overlap. Bottom: three Hanning windows with a 0% overlap. In both cases, the windows cover the time interval from 0 up to $t_{\text{final}} = 3$ s. Each window has 41 points and the sample time is $T_s = 25$ ms; thus, a window covers 1 s of data.

where t_{final} is the period of time during which the signal is measured.

Proof. We will not prove these results, but only sketch how the proof proceeds for the PSD, leaving the CSD as an Exercise. The PSD was defined in (38.23) as the Fourier transform of the autocorrelation (37.35), which, according to (37.39), is a convolution:

$$S_X(j\omega) = \mathcal{F}[R_X(\tau)] = \mathcal{F}[E[X(t)X(t+\tau)]] \quad (38.76)$$

The Fourier transform of a convolution is a product, just as the Laplace transform of a convolution is a product according to (2.78), and (38.74) has a product. The factor of 2 appears because (37.39) can be written as

$$R_X(\tau) = \frac{1}{2} \frac{1}{t_{\text{final}} - \tau} \int_{-t_{\text{final}} + \tau}^{t_{\text{final}} - \tau} x(t)x(t+\tau) dt \quad \square \quad (38.77)$$

The expected values in (38.74)–(38.75) can be estimated as the products $X(j\omega)X(-j\omega)$ and $X(j\omega)Y(-j\omega)$, as calculated from the available data measured from 0 to t_{final} . But it turns out to be better to:

- slice the data into different segments;
- estimate $X(j\omega)X(-j\omega)$ or $X(j\omega)Y(-j\omega)$, as the case may be, for each of the segments;
- and average the estimates.

When doing so, the segments of data may overlap, and the fourth argument of `cpsd(x,y>window,N)` determines how much overlap there is. By default, segments overlap by 50%, which means that each time instant (with the exception of those at the beginning and at the end) belongs to two segments at once (see Figure 38.6).

Overlapping the windows not only gives more estimates for the expected values in (38.74)–(38.75): it also prevents data from being unequally considered in calculations. For instance, with the 0% overlap Hanning windows of Figure 38.6, the signal at 0 s, 1 s, 2 s and 3 s is not used at all, and time instants close by have little influence in the result. Of course, when a rectangular window is used overlap makes no difference in what this is concerned.

38.4 White noise

Definition 38.4. **White noise** is a stochastic process $X(t)$ with autocorrelation $R_X(\tau) = R_X(0)\delta(\tau)$, i.e. the value assumed in each instant bears no relation whatsoever to past values, and has no bearing whatsoever on future values. \square

Example 38.3. In Figure 37.5, the signal in the lower plot, with normally distributed random numbers, is white noise, or, rather, such an approximation of white noise as is possible in practice.

The signal in the upper plot would also be (an approximation of) white noise if it had no mean value. \square

White noise has a constant PSD

Theorem 38.9. The PSD of white noise is constant over all frequencies.

Proof.

$$\mathcal{L}[\delta(\tau)] = 1 \Rightarrow \mathcal{F}[\delta(\tau)] = 1 \square \quad (38.78)$$

White noise does not exist

No analogical signal can have the autocorrelation of white noise. This is better seen taking the analogical signal as a digital signal with $T_s \rightarrow 0$. Since the autocorrelation is zero, the variation of the signal from one sample to the next can take any value. For vanishing sampling times, the velocity would become infinite, which is impossible in practice, since this would require an arbitrarily large energy. The same can be seen from the PSD: a constant value over all frequencies, including arbitrarily large ones, would mean arbitrarily fast oscillations, with an amplitude that never vanishes. (The argument bears a resemblance with the one that shows that transfer functions must be strictly proper, given in Section 11.4.)

White noise approximated in a frequency range

Digital approximations of white noise make more sense

Analogical signals can approximate white noise over a certain range of frequencies if they have a (fairly) constant PSD in that range. The PSD will have to decrease for larger frequencies. Digital signals have their frequency content limited by the sampling time; thus they can have a (fairly) constant PSD over all frequencies up to ω_s , and appear as better approximations of white noise. Of course, they do not have a constant frequency content for arbitrarily large frequencies, just as analogical approximations of white noise do not.

So white noise cannot be found in practice — just as an impulse cannot, and just as unit steps normally have a ramp, fast as it may be, when changing values. But the mathematical convenience of such signals, and their ability to approximate situations frequently found in practice, justify their widespread use.

Where white noise got its name from

Remark 38.4. White light is electromagnetic radiation white noise *in the visible spectrum*. White noise got its name due to its similarity with white light. Light, to be white, does not need to have frequency content outside the visible spectrum; it does not need to have any ultraviolet or infrared content; still less gamma rays or radio waves (see Figure 38.7). Likewise, all signals that in practice pass for white noise will have frequency content in a limited range of frequencies only. \square

Coloured noise

Definition 38.5. **Coloured noise** is noise which is not white. The **color of noise** is the particular evolution of the PSD with frequency. In particular:

- the PSD of pink noise has a slope of -10 dB/decade;
- the PSD of brown noise has a slope of -20 dB/decade;
- the PSD of blue noise has a slope of $+10$ dB/decade;
- the PSD of violet noise has a slope of $+20$ dB/decade.

There are other colours, but these and white are the most frequent. Notice that pink and blue noise can be obtained from white noise with fractional filter $s^{\pm\frac{1}{2}}$. \square

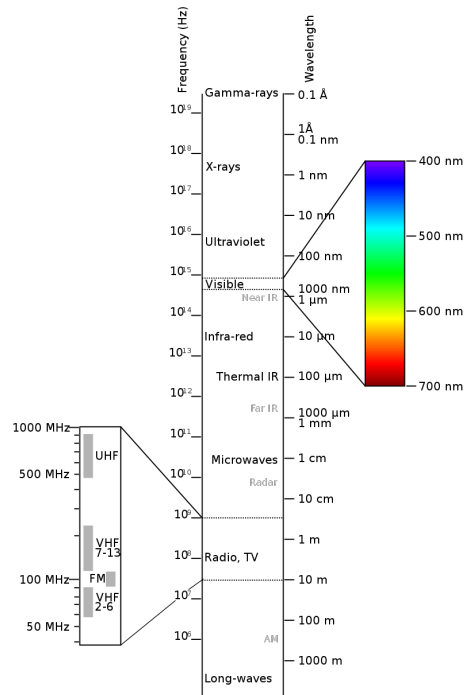


Figure 38.7: The electromagnetic spectrum (source: Wikimedia).

Glossary

“Cornelius, do you and your lawfully wedded spouse speak any language other than English?”

“What is English? I speak the language taught to me by my father and mother, who were taught by their fathers and mothers before them. It has been the language of our ancestors for nearly two thousand years. As to its origins, who can be sure?”

Paul DEHN (1912 — †1976), *Escape from the Planet of the Apes* (1971)

coloured (colored, US) noise ruído colorido

colour (colour, US) of noise cor do ruído

cross-spectral density densidade espectral cruzada

power spectral density densidade de potência espectral

white noise ruído branco

Exercises

1. Prove (38.75).

Chapter 39

Identification of continuous stochastic models

“Jenkins is a very good shot,” said Fisher. “A very good shot who can pretend to be a very bad shot. Shall I tell you the second hint I hit on, after yours, to make me think it was Jenkins? It was my cousin’s account of his bad shooting. He’d shot a cockade off a hat and a weathercock off a building. Now, in fact, a man must shoot very well indeed to shoot so badly as that. He must shoot very neatly to hit the cockade and not the head, or even the hat. If the shots had really gone at random, the chances are a thousand to one that they would not have hit such prominent and picturesque objects. They were chosen because they were prominent and picturesque objects. They make a story to go the round of society.”

Gilbert K. CHESTERTON (1874 — †1936), *The man who knew too much*, I
(*Harper’s Monthly Magazine*, April 1920)

This chapter concerns methods to identify a system when its input and output are stochastic. These methods provide either the impulse response or the frequency response of the system. From that point on, the procedures studied in Part VI are applied.

In all that follows, assume a linear, stable or marginally stable system $G(s)$, with impulse response $g(t)$ and static gain g_0 . Its input is $X(t)$ and its output is

$$Y(t) = X(t) * g(t) = \int_0^t X(t - \tau)g(\tau) d\tau \quad (39.1)$$

as shown in Figure 39.1.

39.1 Identification in time

Theorem 39.1. The expected value of the output \bar{Y} is the expected value of the input \bar{X} multiplied by the static gain g_0 . *Mean value of the output*
 Y

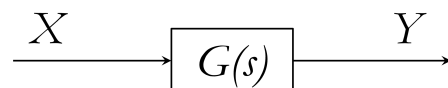


Figure 39.1: System studied in this chapter.

Proof.

$$\begin{aligned}
 \bar{Y} &= \lim_{t \rightarrow +\infty} E \left[\int_0^t X(t-\tau)g(\tau) d\tau \right] \\
 &= \lim_{t \rightarrow +\infty} \int_0^t E \left[\underbrace{X(t-\tau)}_{\text{stochastic}} \underbrace{g(\tau)}_{\text{deterministic}} \right] d\tau \\
 &= \lim_{t \rightarrow +\infty} \int_0^t \underbrace{E[X(t-\tau)]}_{\bar{X} \text{ (constant)}} g(\tau) d\tau \\
 &= \bar{X} \lim_{t \rightarrow +\infty} \int_0^t g(\tau) d\tau \\
 &= \bar{X} \lim_{t \rightarrow +\infty} \int_0^t g(\tau)H(t-\tau) d\tau \\
 &= \bar{X} \lim_{t \rightarrow +\infty} g(t) * H(t) \tag{39.2}
 \end{aligned}$$

This last limit is the steady-state value of the unit step response of $G(s)$, i.e. its static gain g_0 . As assumed throughout this chapter, the system cannot be unstable, otherwise the limit does not exist. \square

Example 39.1. A system with gain 5 receives an input with mean (approximately) equal to $\frac{1}{2}$:

```

>> t = 0 : 0.01 : 1000;
>> X = rand(1,length(t));
>> mean(X)
ans =
0.5004
>> s = tf('s'); G = 5/(s+1)
G =
    5
-----
s + 1
Continuous-time transfer function.
>> Y = lsim(G,X,t);
>> mean(Y)
ans =
2.4994

```

The mean of the output is (approximately) $5 \times \frac{1}{2} = 2.5$. \square

Mean square value of the output

Theorem 39.2. The expected value of the square of the output \bar{Y}^2 is

$$\bar{Y}^2 = \int_0^{+\infty} \int_0^{+\infty} R_X(\tau_1 - \tau_2)g(\tau_1)g(\tau_2) d\tau_1 d\tau_2 \tag{39.3}$$

Proof.

$$\begin{aligned}
 \bar{Y}^2 &= E \left[\overbrace{\int_0^t X(t-\tau_1)g(\tau_1) d\tau_1}^{Y(t)} \overbrace{\int_0^t X(t-\tau_2)g(\tau_2) d\tau_2}^{Y(t)} \right] \\
 &= E \left[\int_0^t \int_0^t \underbrace{X(t-\tau_1)X(t-\tau_2)}_{\text{stochastic}} \underbrace{g(\tau_1)g(\tau_2)}_{\text{deterministic}} d\tau_1 d\tau_2 \right] \\
 &= \int_0^t \int_0^t \underbrace{E[X(t-\tau_1)X(t-\tau_2)]}_{R_X(t-\tau_2-(t-\tau_1))=R_X(\tau_1-\tau_2)} g(\tau_1)g(\tau_2) d\tau_1 d\tau_2 \square \tag{39.4}
 \end{aligned}$$

Autocorrelation of the output

Theorem 39.3. The autocorrelation of the output $R_Y(\tau)$ is

$$R_Y(\tau) = \int_0^{+\infty} \int_0^{+\infty} R_X(\tau + \tau_1 - \tau_2)g(\tau_1)g(\tau_2) d\tau_1 d\tau_2 \tag{39.5}$$

Proof.

$$\begin{aligned}
 R_Y(\tau) &= E \left[\overbrace{\int_0^t X(t-\tau_1)g(\tau_1) d\tau_1}^{Y(t)} \overbrace{\int_0^t X(t+\tau-\tau_2)g(\tau_2) d\tau_2}^{Y(t+\tau)} \right] \\
 &= E \left[\int_0^t \int_0^t \underbrace{X(t-\tau_1)X(t+\tau-\tau_2)}_{\text{stochastic}} \underbrace{g(\tau_1)g(\tau_2)}_{\text{deterministic}} d\tau_1 d\tau_2 \right] \\
 &= \int_0^t \int_0^t \underbrace{E[X(t-\tau_1)X(t+\tau-\tau_2)]}_{R_X(t+\tau-\tau_2-(t-\tau_1))=R_X(\tau-\tau_2+\tau_1)} g(\tau_1)g(\tau_2) d\tau_1 d\tau_2 \quad (39.6)
 \end{aligned}$$

Theorem 39.4. If $G(s)$ is stable, the correlations of input and output $R_{XY}(\tau)$ and $R_{YX}(\tau)$ are *Correlation of input and output*

$$R_{XY}(\tau) = \int_0^{+\infty} R_X(\tau - \tau_1)g(\tau_1) d\tau_1 \quad (39.7)$$

$$R_{YX}(\tau) = \int_0^{+\infty} R_X(\tau + \tau_1)g(\tau_1) d\tau_1 \quad (39.8)$$

Proof.

$$\begin{aligned}
 R_{XY}(\tau) &= E \left[X(t) \overbrace{\int_0^t X(t+\tau-\tau_1)g(\tau_1) d\tau_1}^{Y(t+\tau)} \right] \\
 &= E \left[\int_0^t \underbrace{X(t)X(t+\tau-\tau_1)}_{\text{stochastic}} \underbrace{g(\tau_1)}_{\text{deterministic}} d\tau_1 \right] \\
 &= \int_0^t \underbrace{E[X(t)X(t+\tau-\tau_1)]}_{R_X(t+\tau-\tau_1-t)=R_X(\tau-\tau_1)} g(\tau_1) d\tau_1 \quad (39.9)
 \end{aligned}$$

$$\begin{aligned}
 R_{XY}(\tau) &= E \left[X(t+\tau) \overbrace{\int_0^t X(t-\tau_1)g(\tau_1) d\tau_1}^{Y(t)} \right] \\
 &= E \left[\int_0^t \underbrace{X(t+\tau)X(t-\tau_1)}_{\text{stochastic}} \underbrace{g(\tau_1)}_{\text{deterministic}} d\tau_1 \right] \\
 &= \int_0^t \underbrace{E[X(t+\tau)X(t-\tau_1)]}_{R_X(t-\tau_1-(t+\tau))=R_X(-\tau_1-\tau)} g(\tau_1) d\tau_1 \quad (39.10)
 \end{aligned}$$

Since $R_X(-\tau_1 - \tau) = R_X(\tau + \tau_1)$, (39.8) follows immediately from (39.10). \square

For discrete signals, the integral in (39.7) is replaced with a rectangular approximation. It is more expedient to make use of the fact that $R_X(\tau)$ is even, and consider instead *Correlation of input and output for discrete signals*

$$\begin{aligned}
 R_{XY}(\tau) &= \int_0^{+\infty} R_X(\tau - \tau_1)g(\tau_1) d\tau_1 \\
 &= \int_0^{+\infty} R_X(\tau_1 - \tau)g(\tau_1) d\tau_1 \quad (39.11)
 \end{aligned}$$

Suppose that there are $N + 1$ samples of the impulse response $g(kT_s)$, from $k = 0$ to $k = N$, separated by sampling time T_s ; the autocorrelation $R(kT_s)$, which is an even function, is consequently known from $k = -N$ to $k = N$. The integral is approximated by N rectangles, and correlation $R_{xXY}(nT_s)$ will be,

for successive values of $\tau = nT_s$, given by

$$R_{XY}(0) = T_s \sum_{k=0}^{N-1} R_X(kT_s) g(kT_s) \quad (39.12)$$

$$R_{XY}(T_s) = T_s \sum_{k=0}^{N-1} R_X(kT_s - T_s) g(kT_s) \quad (39.13)$$

$$R_{XY}(2T_s) = T_s \sum_{k=0}^{N-1} R_X(kT_s - 2T_s) g(kT_s) \quad (39.14)$$

⋮

$$R_{XY}(nT_s) = T_s \sum_{k=0}^{N-1} R_X(kT_s - nT_s) g(kT_s) \quad (39.15)$$

These results can be arranged in matrix form:

$$\underbrace{\begin{bmatrix} R_{XY}(0) \\ R_{XY}(T_s) \\ R_{XY}(2T_s) \\ \vdots \\ R_{XY}(nT_s) \end{bmatrix}}_{\mathbf{R}_{XY}} = T_s \underbrace{\begin{bmatrix} R_X(0) & R_X(T_s) & R_X(2T_s) & \cdots & R_X((N-1)T_s) \\ R_X(-T_s) & R_X(0) & R_X(T_s) & \cdots & R_X((N-2)T_s) \\ R_X(-2T_s) & R_X(-T_s) & R_X(0) & \cdots & R_X((N-3)T_s) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R_X(-nT_s) & R_X((1-n)T_s) & R_X((2-n)T_s) & \cdots & R_X((N-1-n)T_s) \end{bmatrix}}_{\mathbf{R}_X} \underbrace{\begin{bmatrix} g(0) \\ g(T_s) \\ g(2T_s) \\ \vdots \\ g(nT_s) \end{bmatrix}}_{\mathbf{g}} \quad (39.16)$$

Finding the impulse response

The matrix is the autocorrelation matrix found in (37.60), save that, if values of τ are never to approach t_{final} , there must be more columns than lines, i.e. $n < N$. If the autocorrelation of the input and the correlation of input and output are known, the impulse response of the plant can be found using the pseudo-inverse of \mathbf{R}_X :

$$\mathbf{h}(t) = \frac{1}{T_s} \mathbf{R}_X^+ \mathbf{R}_{XY} \quad (39.17)$$

Example 39.2. Suppose we have the response of a plant to a (normal) random input, obtained as follows:

```
s = tf('s'); G = 10/(s^2+0.25*s+1);
Ts = 0.01; tfinal = 60; t = 0 : Ts : tfinal;
X = randn(size(t));
Y = lsim(G,X,t);
```

The following function computes the correlation of two signals, and thus can also compute the autocorrelation of a signal:

```
function [correlation,tau] = R_XY(t,X,Y,m)
% function [correlation,tau] = R_X(t,X,Y,m)
% Finds the cross-correlation of X(t) and Y(t) at m points (default is length(t)/2).

n = length(X)-1;
if n+1 ~= length(Y) || n+1 ~= length(t)
    error('t, X and Y must have the same length.')
end
if nargin<4, m = floor(length(X)/2); elseif m>n, m=n; end % too many points

correlation1 = zeros(1,m); % correlation for positive values
for k = 0 : m-1
    correlation1(k+1) = sum( X(1:n-k+1).*Y(1+k:n+1) ) / (n-k+1);
end
correlation2 = zeros(1,m); % correlation for negative values
for k = 0 : m-1
    correlation2(k+1) = sum( Y(1:n-k+1).*X(1+k:n+1) ) / (n-k+1);
end
```

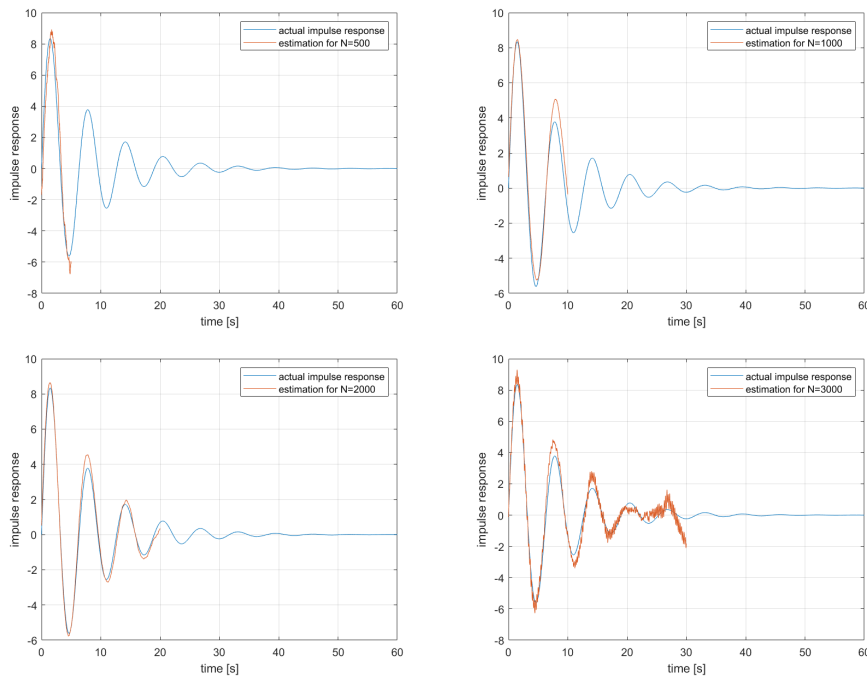


Figure 39.2: Identification of an impulse response from Example 39.2, for different time intervals (5 s, 10 s, 20 s, and 30 s).

```
% the correlation for 0 was obtained twice; we'll keep only one value
correlation = [correlation2(end:-1:2) correlation1];
tau = [ -t(m:-1:2) 0 t(2:m) ];
```

The impulse response of the plant can be recovered from the values of the input and the output as follows, for some desired number of samples, in this case 1000:

```
N = 1000;
[RXX,tauRXX] = R_XY(t,X,X,N);
[RXY,tauRXY] = R_XY(t,X,Y,N);
matrixRX = zeros(N,N);
for k = 1 : N
    matrixRX(k,:) = RXX(N-k+1:end-k+1);
end
estimated_g = 1/Ts * (matrixRX\RXY(N:end)');
```

Figure 39.2 shows the resulting impulse response obtained for different numbers of points (changing variable N in the code above), and compares them with the actual impulse response of the plant. Notice that the estimated impulse response is poor when N is small, because not enough data is being used for a good result, and also when N is high, which may be either because the estimates of $R_X(\tau)$ and $R_{XY}(\tau)$ become poor when τ is large, or because the least squares problem gets so big that numerical problems appear in its resolution. \square

39.2 Identification in frequency

The following theorems provide different ways of obtaining the frequency response of $G(s)$.

Theorem 39.5. The cross-spectral densities of input and output and the spectral density of the input are related by

$$S_{XY}(\omega) = G(j\omega)S_X(\omega) \quad (39.18)$$

$$S_{YX}(\omega) = G(-j\omega)S_X(\omega) \quad (39.19)$$

Proof. From (38.24) and (39.7),

$$\begin{aligned}
S_{XY}(\omega) &= \int_{-\infty}^{+\infty} R_{XY}(\tau) e^{-j\omega\tau} d\tau \\
&= \int_{-\infty}^{+\infty} \int_0^{+\infty} R_X(\tau - \tau_1) g(\tau_1) d\tau_1 e^{-j\omega\tau} d\tau \\
&= \int_0^{+\infty} g(\tau_1) \int_{-\infty}^{+\infty} R_X(\tau - \tau_1) e^{-j\omega\tau} d\tau d\tau_1
\end{aligned} \tag{39.20}$$

We now use the variable change

$$t = \tau - \tau_1 \Rightarrow \tau = t + \tau_1 \tag{39.21}$$

$$\tau = -\infty \Rightarrow t = -\infty \tag{39.22}$$

$$\tau = +\infty \Rightarrow t = +\infty \tag{39.23}$$

$$d\tau = dt \tag{39.24}$$

to get

$$\begin{aligned}
S_{XY}(\omega) &= \int_0^{+\infty} g(\tau_1) \int_{-\infty}^{+\infty} R_X(t) e^{-j\omega t} e^{-j\omega\tau_1} dt d\tau_1 \\
&= \underbrace{\int_0^{+\infty} g(\tau_1) e^{-j\omega\tau_1} d\tau_1}_{\mathcal{F}[g(t)] = G(j\omega)} \underbrace{\int_{-\infty}^{+\infty} R_X(t) e^{-j\omega t} dt}_{S_X(\omega)}
\end{aligned} \tag{39.25}$$

The proof of (39.19) is similar to that of (39.18) and is left as exercise. \square

Corollary 39.1. From (39.18) and (39.19),

$$\begin{aligned}
\frac{S_{XY}(\omega)}{S_{YX}(\omega)} &= \frac{G(j\omega) S_X(\omega)}{\underbrace{G(-j\omega)}_{\overline{G(j\omega)}} S_X(\omega)} \\
&= \frac{|G(j\omega)| e^{j\angle G(j\omega)}}{|G(j\omega)| e^{-j\angle G(j\omega)}} \\
&= e^{j2\angle G(j\omega)} \quad \square
\end{aligned} \tag{39.26}$$

Theorem 39.6. The spectral densities of input and output are related by

$$S_Y(\omega) = S_X(\omega) |G(j\omega)|^2 \tag{39.27}$$

Proof. From (38.23) and (39.5),

$$\begin{aligned}
S_Y(j\omega) &= \int_{-\infty}^{+\infty} R_Y(\tau) e^{-j\omega\tau} d\tau \\
&= \int_{-\infty}^{+\infty} \int_0^{+\infty} \int_0^{+\infty} R_X(\tau + \tau_1 - \tau_2) g(\tau_1) g(\tau_2) d\tau_1 d\tau_2 e^{-j\omega\tau} d\tau \\
&= \int_0^{+\infty} \int_0^{+\infty} g(\tau_1) g(\tau_2) \int_{-\infty}^{+\infty} R_X(\tau + \tau_1 - \tau_2) e^{-j\omega\tau} d\tau d\tau_1 d\tau_2
\end{aligned} \tag{39.28}$$

We now use the variable change

$$t = \tau + \tau_1 - \tau_2 \Rightarrow \tau = t - \tau_1 + \tau_2 \tag{39.29}$$

$$\tau = -\infty \Rightarrow t = -\infty \tag{39.30}$$

$$\tau = +\infty \Rightarrow t = +\infty \tag{39.31}$$

$$d\tau = dt \tag{39.32}$$

to get

$$\begin{aligned}
S_Y(j\omega) &= \int_0^{+\infty} \int_0^{+\infty} g(\tau_1) g(\tau_2) \int_{-\infty}^{+\infty} R_X(t) e^{-j\omega t} e^{j\omega\tau_1} e^{-j\omega\tau_2} dt d\tau_1 d\tau_2 \\
&= \underbrace{\int_0^{+\infty} g(\tau_2) e^{-j\omega\tau_2} d\tau_2}_{\mathcal{F}[g(t)] = G(j\omega)} \underbrace{\int_0^{+\infty} g(\tau_1) e^{j\omega\tau_1} d\tau_1}_{G(-j\omega)} \underbrace{\int_{-\infty}^{+\infty} R_X(t) e^{-j\omega t} dt}_{S_X(j\omega)}
\end{aligned} \tag{39.33}$$

Finally, $G(j\omega)G(-j\omega) = G(j\omega)\overline{G(j\omega)} = |G(j\omega)|^2$, which completes the proof. \square

Corollary 39.2. Replacing (39.27) in (38.35),

$$\begin{aligned}\overline{Y^2(t)} &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_Y(j\omega) d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_X(\omega) |G(j\omega)|^2 d\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_X(\omega) G(j\omega) G(-j\omega) d\omega \quad \square\end{aligned}\quad (39.34)$$

So this is how the frequency response of $G(s)$ can be found from spectral densities $S_X(j\omega)$ and $S_Y(j\omega)$ and from cross-spectral densities $S_{XY}(j\omega)$ and $S_{YX}(j\omega)$:

- The gain is either found from (39.27)

Finding the gain

$$|G(j\omega)| = \sqrt{\frac{S_Y(\omega)}{S_X(\omega)}} \quad (39.35)$$

or from (39.18)

$$|G(j\omega)| = \left| \frac{S_{XY}(\omega)}{S_X(\omega)} \right| \quad (39.36)$$

or from (39.19)

$$|G(j\omega)| = |G(-j\omega)| = \left| \frac{S_{YX}(\omega)}{S_X(\omega)} \right| \quad (39.37)$$

- The phase is either found from (39.26)

Finding the phase

$$\begin{aligned}\angle G(j\omega) &= \Im \left[\frac{1}{2} \log \frac{S_{XY}(\omega)}{S_{YX}(\omega)} \right] \\ &= -\frac{j}{2} \log \frac{S_{XY}(\omega)}{S_{YX}(\omega)}\end{aligned}\quad (39.38)$$

or from (39.18)

$$\angle G(j\omega) = \angle \frac{S_{XY}(\omega)}{S_X(\omega)} \quad (39.39)$$

or from (39.19)

$$\angle G(j\omega) = -\angle G(-j\omega) = -\angle \frac{S_{YX}(\omega)}{S_X(\omega)} \quad (39.40)$$

- The entire frequency response is either found from (39.18)

Finding the frequency response

$$G(j\omega) = \frac{S_{XY}(\omega)}{S_X(\omega)} \quad (39.41)$$

or from (39.19)

$$G(j\omega) = \overline{G(-j\omega)} = \overline{\left(\frac{S_{YX}(\omega)}{S_X(\omega)} \right)} \quad (39.42)$$

or conjugating (39.27) and (39.26)

$$G(j\omega) = |G(j\omega)| e^{j\angle G(j\omega)} = \sqrt{\frac{S_Y(\omega) S_{XY}(\omega)}{S_X(\omega) S_{YX}(\omega)}} \quad (39.43)$$

To verify if an identified frequency response is accurate, a function analogous to the correlation coefficient is used.

Definition 39.1. The **coherence function** γ_{XY}^2 is given by

Coherence function

$$\gamma_{XY}^2(\omega) = \frac{|S_{XY}(\omega)|^2}{|S_X(\omega)||S_Y(\omega)|} \quad \square \quad (39.44)$$

Theorem 39.7. In the conditions of Figure 39.1,

$\gamma_{XY}^2(\omega) = 1$ for a L all noise is accounted

$$\gamma_{XY}^2(\omega) \leq 1, \quad \forall \omega \quad (39.45)$$

Proof. From

$$\begin{aligned} \gamma_{XY}^2(\omega) &= \frac{|S_{XY}(\omega)|^2}{|S_X(\omega)||S_Y(\omega)|} \\ &= \frac{|G(j\omega)S_X(\omega)|^2}{|S_X(\omega)||S_X(\omega)|G(j\omega)|^2} \\ &= \frac{|G(j\omega)|^2|S_X(\omega)|^2}{|G(j\omega)|^2|S_X(\omega)|^2} = 1 \square \end{aligned} \quad (39.46)$$

$$0 \leq \gamma_{XY}^2(\omega) \leq 1$$

In practice, $\gamma_{XY}^2(\omega)$ is seldom 1. The limit case is that of an uncorrelated pair of input and output, and then

$$R_{XY}(\tau) = 0 \Rightarrow S_{XY}(\omega) = 0 \Rightarrow \gamma_{XY}^2(\omega) = 0 \quad (39.47)$$

Consequently:

- When $\gamma_{XY}^2(\omega) = 1$, a linear model $G(s)$ relating input X with output Y can be found from the data.
- When $\gamma_{XY}^2(\omega) < 1$, no model can predict Y solely from X . This may happen for several reasons: there is noise that could not be measured; there is another input; there are non-linearities.
- The magnitude of $\gamma_{XY}^2(\omega)$ shows how good the model will be, and at what frequencies it will perform better and worse.
- When $\gamma_{XY}^2(\omega) = 0$, no model can be found.

Example 39.3. Consider the following linear system:

```
Ts = 0.01; % sample time
t = 0 : Ts : 1000;
X = rand(1,length(t)); % input
s = tf('s'); G = 5/(s+1);
Y = lsim(G,X,t); % output
```

Figure 39.3 shows the values of input X and output Y of plant $G(s)$ during some seconds.

The gain and phase of the frequency response of $G(s)$ can be obtained from X and Y using (39.35)–(39.40) as follows. They will be compared with the result of command `freqresp`.

```
% spectral densities
[Syx,F] = cpsd(X,Y,[],[],[],1/Ts); w = 2*pi*F;
Sxy = cpsd(Y,X,[],[],[],1/Ts); % remember that the order of X and Y is like this
Sxx = cpsd(X,X,[],[],[],1/Ts);
Syy = cpsd(Y,Y,[],[],[],1/Ts);
% gain
Gfreqresp = squeeze(freqresp(G,w));
Ggain = 20*log10(abs(Gfreqresp)); % this is what we should obtain
Ggain1 = 20*log10(sqrt(Syy./Sxx));
Ggain2 = 20*log10(abs(Sxy./Sxx));
Ggain3 = 20*log10(abs(Syx./Sxx));
% phase
Gphase = rad2deg(angle(Gfreqresp));
Gphase1 = rad2deg(unwrap(imag(0.5*log(Sxy./Syx))));
Gphase2 = rad2deg(unwrap(angle(Sxy./Sxx)));
Gphase3 = -rad2deg(unwrap(angle(Syx./Sxx)));
```

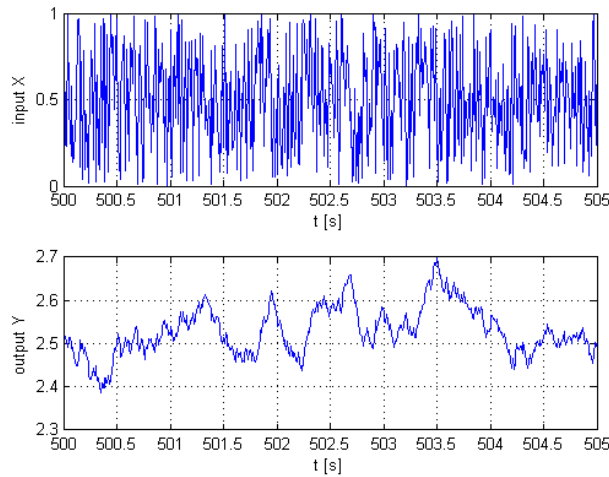


Figure 39.3: Sample of input and output of the plant of Example 39.3.

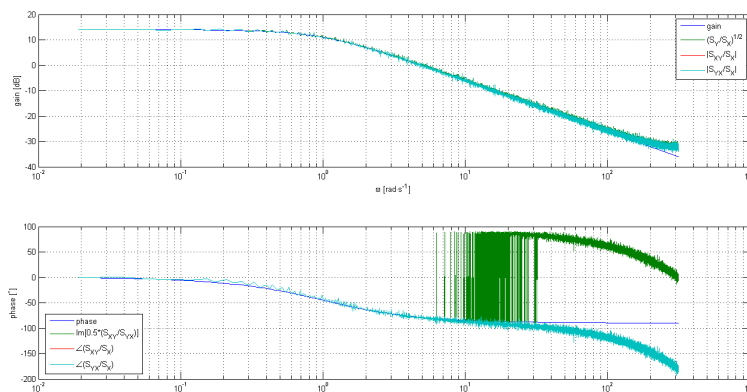


Figure 39.4: Bode diagrams of the plant of Example 39.3.

The corresponding Bode diagrams are shown in Figure 39.4, as plotted with the following commands:

```
figure, subplot(2,1,1), semilogx(w,Ggain, w,Ggain1, w,Ggain2, w,Ggain3)
grid on, xlabel('\omega [rad\cdots^{-1}]'), ylabel('gain [dB]')
legend({'gain', '(S_Y/S_X)^{1/2}', '|S_{XY}/S_X|', '|S_{YX}/S_X|'})
subplot(2,1,2), semilogx(w,Gphase, w,Gphase1, w,Gphase2, w,Gphase3)
grid on, ylabel('phase [^\circ]')
legend({'phase', 'Im[0.5*(S_{XY}/S_{YX})]', '\angle(S_{XY}/S_X)', '\angle(S_{YX}/S_X)'})
```

Two conclusions can at once be taken.

- There are numerical problems with the result of (39.38), namely oscillations of 180° which command `unwrap` does not solve, since it only deals with jumps of 360° . The way around this is to double the angle before applying `unwrap`, so that jumps of 180° become 360° wide, and then halve it:

```
Gphase1 = rad2deg(0.5*unwrap(2*imag(0.5*log(Sxy./Syx))));
```

- Results are poor in the last decade, something we already know to be expectable. This can be confirmed with γ_{XY}^2 , shown in Figure 39.5 and obtained as follows:

```
gamma2 = (abs(Sxy)).^2./(abs(Sxx).*abs(Syy));
figure, semilogx(w,gamma2, '.')
grid on, xlabel('\omega [rad\cdots^{-1}]'), ylabel('\gamma_{XY}^2')
```

Fixing the phase and neglecting the last decade, the Bode diagram in Figure 39.6 is obtained. \square

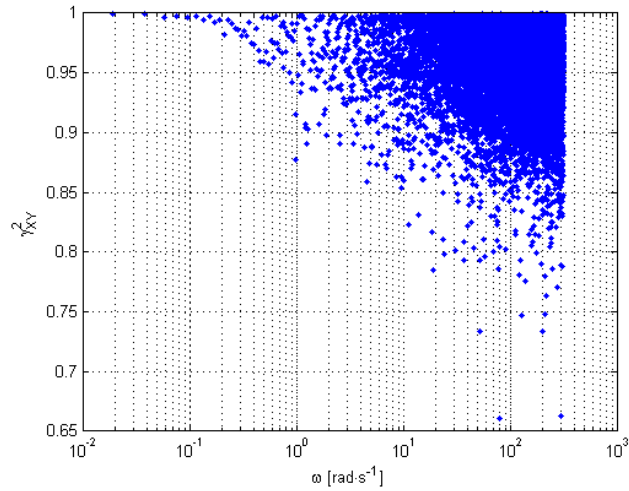


Figure 39.5: Coherence function of Example 39.3.

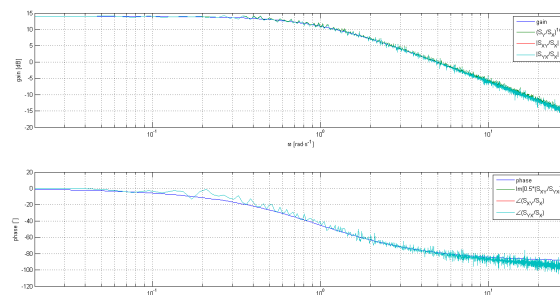


Figure 39.6: Improved Bode diagrams of the plant of Example 39.3.

Glossary

The political dialects to be found in pamphlets, leading articles, manifestos, White Papers and the speeches of Under-Secretaries do, of course, vary from party to party, but they are all alike in that one almost never finds in them a fresh, vivid, home-made turn of speech. When one watches some tired hack on the platform mechanically repeating the familiar phrases — *bestial atrocities*, *iron heel*, *blood-stained tyranny*, *free peoples of the world*, *stand shoulder to shoulder* — one often has a curious feeling that one is not watching a live human being but some kind of dummy: a feeling which suddenly becomes stronger at moments when the light catches the speaker's spectacles and turns them into blank discs which seem to have no eyes behind them. And this is not altogether fanciful. A speaker who uses that kind of phraseology has gone some distance toward turning himself into a machine. The appropriate noises are coming out of his larynx, but his brain is not involved as it would be if he were choosing his words for himself. If the speech he is making is one that he is accustomed to make over and over again, he may be almost unconscious of what he is saying, as one is when one utters the responses in church. And this reduced state of consciousness, if not indispensable, is at any rate favourable to political conformity.

George ORWELL (1903 — †1950), *Politics and the English Language* (1946)

coherence function função de coerência

Exercises

1. Prove (39.19).

Chapter 40

Filter design

Other noises were subdued in this city of rubber; the passenger-circles were a hundred yards away, and the subterranean traffic lay too deep for anything but a vibration to make itself felt. It was to remove this vibration, and silence the hum of the ordinary vehicles, that the Government experts had been working for the last twenty years.

Robert Hugh BENSON (1871 — †1914), *Lord of the world* (1907), Prologue

This chapter concerns the design of filters that receive a signal $x(t)$ corrupted by additive noise $n(t)$, with the objective of returning an output $y(t)$ as close as possible to $x(t)$, as seen in Figure 40.1. As might be expected, the ideal situation $y(t) = x(t)$ is seldom attainable, if ever; but useful approximations can be usually found.

40.1 Wiener filters

The output of the filter with transfer function $H(s)$ in Figure 40.1 will be, in the general case,

$$\begin{aligned} y(t) &= \mathcal{L}^{-1} [H(s) (X(s) + N(s))] \\ &= \underbrace{\mathcal{L}^{-1} [H(s) X(s)]}_{\text{distorted signal}} + \underbrace{\mathcal{L}^{-1} [H(s) N(s)]}_{\text{residual noise}} \end{aligned} \quad (40.1)$$

Consequently, there are two sources of error:

- signal distortion $e(t) = \mathcal{L}^{-1} [E(s)] = \mathcal{L}^{-1} [X(s) - \tilde{X}(s)] = \mathcal{L}^{-1} [X(s)(1 - H(s))]$;
- residual noise $m(t) = \mathcal{L}^{-1} [M(s)] = \mathcal{L}^{-1} [N(s)H(s)]$.

Definition 40.1. A **Wiener filter** minimises $\overline{e^2(t)} + \overline{m^2(t)}$. □

Theorem 40.1. A Wiener filter $H(s)$ for a signal $x(t)$ with spectral density $S_X(s)$ corrupted by noise $n(t)$ with spectral density $S_N(s)$ is given by

$$H(s) = \frac{S_X(s)}{S_X(s) + S_N(s)} \quad (40.2)$$

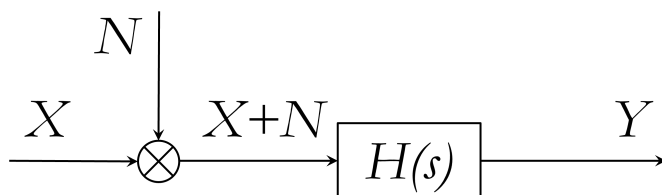


Figure 40.1: Block diagram for filters $H(s)$ addressed in this chapter.

Proof. According to (38.35), using Laplace rather than Fourier transforms thanks to variable change $j\omega = s \Leftrightarrow \omega = \frac{s}{j}$,

$$\begin{aligned}\overline{m^2} &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_N(\omega)H(j\omega)H(-j\omega) d\omega \\ &= \frac{1}{2\pi} \int_{-j\infty}^{+j\infty} S_N(s)H(s)H(-s) d\left(\frac{s}{j}\right)\end{aligned}\quad (40.3)$$

$$\begin{aligned}\overline{e^2} &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_X(\omega)(1-H(j\omega))(1-H(-j\omega)) d\omega \\ &= \frac{1}{2\pi} \int_{-j\infty}^{+j\infty} S_X(s)(1-H(s))(1-H(-s)) d\left(\frac{s}{j}\right) \\ &= \frac{1}{2\pi j} \int_{-j\infty}^{+j\infty} \left(S_X(s) - S_X(s)H(-s) - S_X(s)H(s) + S_X(s)H(s)H(-s) \right) ds\end{aligned}\quad (40.4)$$

and thus

$$\begin{aligned}\overline{e^2} + \overline{m^2} &= \frac{1}{2\pi j} \int_{-j\infty}^{+j\infty} \left(S_X(s) - S_X(s)H(-s) - S_X(s)H(s) \right. \\ &\quad \left. + S_X(s)H(s)H(-s) + S_N(s)H(s)H(-s) \right) ds\end{aligned}\quad (40.5)$$

It is now expedient to define

$$F_i(s)F_i(-s) = S_X(s) + S_N(s) \quad (40.6)$$

so that

- $F_i(s)$ will only have poles and zeros with negative real parts, or on the imaginary axis;
- the poles and zeros of $F_i(-s)$ will be the complex conjugates of those of $F_i(s)$.

We can now write

$$\overline{e^2} + \overline{m^2} = \frac{1}{2\pi j} \int_{-j\infty}^{+j\infty} \left(\left(F_i(s)H(s) - \frac{S_X(s)}{F_i(-s)} \right) \left(F_i(-s)H(-s) - \frac{S_X(s)}{F_i(-s)} \right) + \frac{S_X(s)S_N(s)}{F_i(s)F_i(-s)} \right) ds \quad (40.7)$$

as can be seen working out the products and sums:

$$\begin{aligned}& \frac{1}{2\pi j} \int_{-j\infty}^{+j\infty} \left(\left(F_i(s)H(s) - \frac{S_X(s)}{F_i(-s)} \right) \left(F_i(-s)H(-s) - \frac{S_X(s)}{F_i(-s)} \right) + \frac{S_X(s)S_N(s)}{F_i(s)F_i(-s)} \right) ds \\ &= \frac{1}{2\pi j} \int_{-j\infty}^{+j\infty} \left(\underbrace{F_i(s)F_i(-s)}_{S_X(s)+S_N(s)} H(s)H(-s) - \frac{F_i(s)H(s)S_X(s)}{F_i(s)} - \frac{F_i(-s)H(-s)S_X(s)}{F_i(-s)} \right. \\ &\quad \left. + \frac{S_X^2(s)}{F_i(s)F_i(-s)} + \frac{S_X(s)S_N(s)}{F_i(s)F_i(-s)} \right) ds \\ &= \frac{1}{2\pi j} \int_{-j\infty}^{+j\infty} \left(S_X(s)H(s)H(-s) + S_N(s)H(s)H(-s) \right. \\ &\quad \left. + \frac{F_i(s)F_i(-s)}{F_i(s)F_i(-s)} \left(S_X(s)(S_X(s) + S_N(s)) - F_i(s)F_i(-s)H(s)S_X(s) - F_i(s)F_i(-s)H(-s)S_X(s) \right) \right) ds \\ &= \frac{1}{2\pi j} \int_{-j\infty}^{+j\infty} \left(S_X(s)H(s)H(-s) + S_N(s)H(s)H(-s) + S_X(s) - H(s)S_X(s) - H(-s)S_X(s) \right) ds\end{aligned}\quad (40.8)$$

This is equal to (40.5).

We can now take a better look at the integrand of (40.7):

- $\frac{S_X(s)S_N(s)}{F_i(s)F_i(-s)} = \frac{S_X(s)S_N(s)}{S_X(s)+S_N(s)}$ does not depend on $H(s)$, so there is nothing to do to this term;
- $F_i(s)H(s) - \frac{S_X(s)}{F_i(-s)}$ and $F_i(-s)H(-s) - \frac{S_X(s)}{F_i(-s)}$ depend on $H(s)$. We know from (38.38) that the second term verifies

$$F_i(-s)H(-s) - \frac{S_X(s)}{F_i(-s)} = F_i(-s)H(-s) - \frac{S_X(-s)}{F_i(-s)} \quad (40.9)$$

Consequently, these two terms are complex conjugates, and their product is always positive or zero; it cannot be negative.

Thus, to minimise the integrand and thereby minimise $\overline{e^2} + \overline{m^2}$, we must have

$$F_i(s)H(s) - \frac{S_X(s)}{F_i(-s)} = 0 \quad (40.10)$$

from which the result is immediate. \square

Example 40.1. Let

$$S_X(s) = -\frac{24}{s^2 - 1} \quad (40.11)$$

$$S_N(s) = -\frac{1}{s^2} \quad (40.12)$$

Then

$$H(s) = \frac{-\frac{24}{s^2-1}}{-\frac{24}{s^2-1} - \frac{1}{s^2}} = \frac{-24s^2}{-25s^2 + 1} \quad (40.13)$$

which has poles at $25s^2 = 1 \Leftrightarrow s = \pm\frac{1}{5}$. \square

In the example above, the Wiener filter has a pole on the right-side of the complex plane. We saw in Remark 38.3 the reason why poles and zeros with a positive real part will often appear. Since this is undesirable, a **causal Wiener filter** is needed, which will no longer minimise $\overline{e^2(t)} + \overline{m^2(t)}$, but will provide the best possible proper approximation to a filter that does. Notice that the causal Wiener filter has this name even though the problem it solves is not necessarily that of causality. To distinguish it from the causal Wiener filter, the Wiener filter from Definition 40.1 is known as **infinite Wiener filter**.

Infinite Wiener filter

Definition 40.2. A causal Wiener filter is the result of removing zeros and poles with a positive real part from a Wiener filter. \square

When solving (40.10), $F_i(s)$ has by definition no poles and zeros with positive real part: these will appear (if they do) in $\frac{S_X(s)}{F_i(-s)}$. A partial fraction expansion can then be used to separate $\frac{S_X(s)}{F_i(-s)}$ into the sum of

- a rational function without poles and zeros with positive real part $\left[\frac{S_X(s)}{F_i(-s)} \right]_{\text{left}}$,
and
- a rational function without poles and zeros with negative real part $\left[\frac{S_X(s)}{F_i(-s)} \right]_{\text{right}}$,

so that (40.10) becomes

$$F_i(s)H(s) - \left[\frac{S_X(s)}{F_i(-s)} \right]_{\text{left}} - \left[\frac{S_X(s)}{F_i(-s)} \right]_{\text{right}} = 0 \quad (40.14)$$

Term $\left[\frac{S_X(s)}{F_i(-s)} \right]_{\text{right}}$ is neglected, and the causal Wiener filter is given by

$$H(s) = \frac{1}{F_i(s)} \left[\frac{S_X(s)}{F_i(-s)} \right]_{\text{left}} \quad (40.15)$$

While it might seem that it would suffice to neglect completely $F_i(-s)$, since by definition all its poles and zeros have non-negative real parts, it is better to proceed as above since such zeros and poles may cancel poles and zeros of $S_X(s)$.

Example 40.2. In Example 40.1, we make

$$\begin{aligned}
 F_i(s)F_i(-s) &= \underbrace{\frac{S_X(s)}{s^2-1}}_{-24} + \underbrace{\frac{S_N(s)}{s^2}}_{-1} \\
 &= \frac{-25s^2+1}{(s^2-1)s^2} = \frac{(5s-1)(5s+1)}{-s^2(s-1)(s+1)} \\
 &= \underbrace{\frac{5s+1}{s(s+1)}}_{F_i(s)} \underbrace{\frac{-5s+1}{-s(-s+1)}}_{F_i(-s)}
 \end{aligned} \tag{40.16}$$

Then

$$\begin{aligned}
 \frac{S_X(s)}{F_i(-s)} &= \frac{\underbrace{S_X(s)}_{-24} \underbrace{\frac{1}{F_i(-s)}}_{-s(-s+1)}}{s^2-1} \\
 &= \frac{24s(s-1)(-1)}{(s+1)(s-1)(5s-1)(-1)} = \frac{24s}{(s+1)(5s-1)}
 \end{aligned} \tag{40.17}$$

A straightforward partial fraction expansion gives

$$\frac{S_X(s)}{F_i(-s)} = \frac{24s}{(s+1)(5s-1)} = \underbrace{\frac{4}{s+1}}_{\left[\frac{S_X(s)}{F_i(-s)}\right]_{\text{left}}} + \underbrace{\frac{4}{5s-1}}_{\left[\frac{S_X(s)}{F_i(-s)}\right]_{\text{right}}} \tag{40.18}$$

Consequently

$$H(s) = \frac{s(s+1)}{\underbrace{5s+1}_{\frac{1}{F_i(s)}}} \underbrace{\frac{4}{s+1}}_{\left[\frac{S_X(s)}{F_i(-s)}\right]_{\text{left}}} = \frac{4s}{5s+1} \tag{40.19}$$

□

Wiener filters can also be used:

Extrapolation

- To estimate $x(t - \tau)$, i.e. the value that signal $x(t)$ had, $-\tau$ seconds into the past. This is called **extrapolation**.

Prediction

- To estimate $x(t + \tau)$, i.e. the value that signal $x(t)$ will have, τ seconds into the future. This is called **prediction**.

As a consequence of what we saw in Section 24.1, both estimations are obtained replacing the (causal) Wiener filter $H(s)$ with $H(s)e^{\pm\tau s}$.

40.2 Whitening filters

Filters may have different objectives than those of a Wiener filter.

Definition 40.3. A **whitening filter** is a filter that outputs white noise. □

Theorem 40.2. A causal whitening filter $H(s)$ for a signal $x(t)$ with spectral density $S_X(s)$ corrupted by noise $n(t)$ with spectral density $S_N(s)$ is given by

$$H(s) = \frac{1}{F_i(s)} \tag{40.20}$$

where $F_i(s)$ is defined as in (40.6).

Proof. According to (39.27), we shall have

$$\begin{aligned}
 S_Y(s) &= (S_X(s) + S_N(s)) |H(s)|^2 \\
 &= (S_X(s) + S_N(s)) \underbrace{\frac{1}{F_i(s)} \frac{1}{F_i(-s)}}_{\frac{1}{S_X(s) + S_N(s)}} = 1 \square
 \end{aligned} \tag{40.21}$$

As was the case with the Wiener filter, whitening filters may have poles and zeros with positive real parts, which in practice have to be neglected; or be non-causal and need to have additional poles added. Such approximations, of course, make the output of the filter deviate more from white noise.

Example 40.3. In Example 40.1, the whitening filter will be

$$H(s) = \frac{1}{F_i(s)} = \frac{s(s+1)}{5s+1} \quad (40.22)$$

which, being an improper transfer function, will have to have two additional poles added to become causal. \square

Glossary

Entam elle por lhe querer a cudir descuidara de si e o foguo fizeralhe algũ nojo por ptes de seu corpo, e direito do caualeiro topou com outro mateiro que pera ho mato hia que lhe perguntou vendo ho vir assi sem lenha que pera que fora ao mato. Respondendolhe o mateiro queimado falandolhe galego estas soos palauras, Bimarder, olhou o caualeiro pelo barbarismo das letras mudadas na pronunciaçam do, b, por, v, e pareceolhe misterio por que elle tambem na quelle se fora arder, e quis se chamar assi da hi auante (...)

Bernardim RIBEIRO (1482? — †1552?), *Hystoria de Menina e Moça* (1554, posth.)

causal Wiener filter filtro de Wiener causal

infinite Wiener filter filtro de Wiener infinito

Wiener filter filtro de Wiener

whitening filter filtro de branqueamento

Exercises

1. A signal with $S_X(s) = \frac{-1}{s^2-1}$ is corrupted by noise with $S_N(s) = \frac{-1}{s^2-4}$. Find:
 - (a) A Wiener filter.
 - (b) A causal Wiener filter.
 - (c) A causal Wiener filter to extrapolate the signal 5 seconds in the past.
2. A signal with $S_X(s) = \frac{-1}{s^2-4}$ is corrupted by noise with $S_N(s) = \frac{-1}{s^2-16}$. Find:
 - (a) A Wiener filter.
 - (b) A causal Wiener filter.
 - (c) A causal Wiener filter to predict the signal 2 seconds in the future.
3. A signal with $S_X(s) = \frac{-9}{s^2-9}$ is corrupted by noise with $S_N(s) = \frac{-1}{s^2-16}$. Find:
 - (a) A Wiener filter.
 - (b) A causal Wiener filter.
 - (c) A causal Wiener filter to extrapolate the signal 1 second in the past.
4. Find whitening filters for the inputs of the previous exercises.
5. Consider an input consisting exclusively of brown noise. Find:
 - (a) A whitening filter.
 - (b) A causal whitening filter.

6. A stochastic signal $X(t)$, corrupted with noise $N(t)$, is filtered by a filter with transfer function $H(t)$, that has an output $Y(t)$ consisting of distorted signal $X(t)$ and residual noise $M(t)$. The power spectral densities of the original signal and the noise are

$$S_X(\omega) = \frac{144}{\omega^2 + 1} \quad (40.23)$$

$$S_N(\omega) = \frac{25}{\omega^2} \quad (40.24)$$

- (a) Find $H(s)$, if the filter is a Wiener filter.
- (b) What does the filter you have just found minimise, and why is it that you cannot implement it?
- (c) Find the whitening filter $\frac{1}{F_i(s)}$.
- (d) Find $H(s)$, if the filter is a causal Wiener filter.

Chapter 41

Digital stochastic models

He held the last coin between his fingers, staring absently at it.

“Multivac is not the first computer, friends, nor the best-known, nor the one that can most efficiently lift the load of decision from the shoulders of the executive. A machine did win the war, John; at least a very simple computing device did; one that I used every time I had a particularly hard decision to make.”

With a faint smile of reminiscence, he flipped the coin he held. It glinted in the air as it spun and came down in Swift’s outstretched palm. His hand closed over it and brought it down on the back of his left hand. His right hand remained in place, hiding the coin.

“Heads or tails, gentlemen?” said Swift.

Isaac ASIMOV (1920 — †1992), *The machine that won the war*, The Magazine of Fantasy & Science Fiction, October 1961

We now turn our attention to digital stochastic models, often useful given that most signals are, as we saw in Chapter 3, discrete in time.

41.1 Types of digital stochastic models

Consider a digital system $G(z^{-1})$ with two inputs:

- u_k is a manipulated input, and is known as the **exogenous** input;
- e_k is stochastic, and thus, from the point of view of a control system, a disturbance. This is usually assumed to be white noise.

$G(z^{-1})$ is assumed as linear and proper, and thus its output y_k is a linear combination of the inputs and their past values. It will be a stochastic system if y_k depends only on e_k but not on u_k , or if it depends on both. Several paradigms of digital stochastic systems are usually found.

Definition 41.1. The following digital stochastic systems have particular names.

- Models that only depend on the stochastic input e_k :

– **Autoregressive model** of order p , AR(p): AR

$$\begin{aligned} y_k &= e_k + a_1 y_{k-1} + a_2 y_{k-2} + \dots + a_p y_{k-p} \\ \Leftrightarrow y_k - a_1 y_{k-1} - a_2 y_{k-2} - \dots - a_p y_{k-p} &= e_k \\ \Leftrightarrow y_k \underbrace{(1 - a_1 z^{-1} - a_2 z^{-2} - \dots - a_p z^{-p})}_{A(z^{-1})} &= e_k \end{aligned}$$

$$\Leftrightarrow \frac{y_k}{e_k} = \frac{1}{A(z^{-1})} \quad (41.1)$$

– **Moving average model** of order q , MA(q): MA

$$\begin{aligned}
y_k &= e_k - c_1 e_{k-1} - c_2 e_{k-2} - \dots - c_q e_{k-q} \\
\Leftrightarrow y_k &= e_k \underbrace{(1 - c_1 z^{-1} - c_2 z^{-2} - \dots - c_q z^{-q})}_{C(z^{-1})} \\
\Leftrightarrow \frac{y_k}{e_k} &= C(z^{-1}) \tag{41.2}
\end{aligned}$$

ARMA – **Autoregressive, moving average model** of orders p, q , ARMA(p, q):

$$\begin{aligned}
y_k &= e_k - c_1 e_{k-1} - c_2 e_{k-2} - \dots - c_q e_{k-q} + a_1 y_{k-1} + a_2 y_{k-2} + \dots + a_p y_{k-p} \\
\Leftrightarrow y_k \underbrace{(1 - a_1 z^{-1} - a_2 z^{-2} - \dots - a_p z^{-p})}_{A(z^{-1})} &= e_k \underbrace{(1 - c_1 z^{-1} - c_2 z^{-2} - \dots - c_q z^{-q})}_{C(z^{-1})} \\
\Leftrightarrow \frac{y_k}{e_k} &= \frac{C(z^{-1})}{A(z^{-1})} \tag{41.3}
\end{aligned}$$

ARIMA

– **Autoregressive, integral, moving average model** of orders p, q, d , ARIMA(p, q, d):

$$\frac{y_k}{e_k} = \frac{C(z^{-1})}{A(z^{-1})(1 - z^{-1})^d} \tag{41.4}$$

The poles in $z = 1$ correspond, as seen in Chapter 25, to integrations. In fact an ARIMA(p, q, d) model is only a particular case of an ARMA($p, q + d$), as the denominator of (41.4) is a polynomial of order $q + d$.

- Models that depend both on the stochastic input e_k and the exogenous input u_k :

ARX

– **Autoregressive, exogenous model** of orders p, m , ARX(p, m):

$$\begin{aligned}
y_k &= e_k + a_1 y_{k-1} + a_2 y_{k-2} + \dots + a_p y_{k-p} + b_0 u_k - b_1 u_{k-1} - b_2 u_{k-2} - \dots - b_m u_{k-m} \\
\Leftrightarrow y_k \underbrace{(1 - a_1 z^{-1} - a_2 z^{-2} - \dots - a_p z^{-p})}_{A(z^{-1})} &= e_k + u_k \underbrace{(b_0 - b_1 z^{-1} - b_2 z^{-2} - \dots - b_m z^{-m})}_{B(z^{-1})} \\
\Leftrightarrow y_k &= \frac{1}{A(z^{-1})} e_k + \frac{B(z^{-1})}{A(z^{-1})} u_k \tag{41.5}
\end{aligned}$$

An ARX model is sometimes given as ARX(p, m, d), allowing for a delay z^{-d} that affects the manipulated input (i.e. the exogenous input):

$$y_k = \frac{1}{A(z^{-1})} e_k + \frac{B(z^{-1})z^{-d}}{A(z^{-1})} u_k \tag{41.6}$$

This is the same as an ARX($p, m + d$) model, since the numerator of the second transfer function is of order $m + d$. It makes no sense to allow for a delay in the noise, precisely because it is noise, but a delay in a control action may exist, as seen in Chapter 24.

– **Autoregressive, moving average, exogenous model** of orders p, q, m , ARMAX(p, q, m):

$$\begin{aligned}
y_k &= e_k - c_1 e_{k-1} - c_2 e_{k-2} - \dots - c_q e_{k-q} + a_1 y_{k-1} + a_2 y_{k-2} + \dots + a_p y_{k-p} \\
&\quad + b_0 u_k - b_1 u_{k-1} - b_2 u_{k-2} - \dots - b_m u_{k-m} \\
\Leftrightarrow y_k \underbrace{(1 - a_1 z^{-1} - a_2 z^{-2} - \dots - a_p z^{-p})}_{A(z^{-1})} &= e_k \underbrace{(1 - c_1 z^{-1} - c_2 z^{-2} - \dots - c_q z^{-q})}_{C(z^{-1})} + u_k \underbrace{(b_0 - b_1 z^{-1} - b_2 z^{-2} - \dots - b_m z^{-m})}_{B(z^{-1})} \\
\Leftrightarrow y_k &= \frac{C(z^{-1})}{A(z^{-1})} e_k + \frac{B(z^{-1})}{A(z^{-1})} u_k \tag{41.7}
\end{aligned}$$

An ARMAX model is sometimes given as ARX(p, q, m, d), allowing for a delay z^{-d} that affects the manipulated input (i.e. the exogenous input):

$$y_k = \frac{C(z^{-1})}{A(z^{-1})} e_k + \frac{B(z^{-1})z^{-d}}{A(z^{-1})} u_k \tag{41.8}$$

This is the same as an ARMAX($p, q, m + d$) model, since the numerator of the second transfer function is of order $m + d$.

- **Autoregressive, integral, moving average, exogenous model** of orders p, q, d, m , ARIMAX(p, q, d, m): ARIMAX

$$y_k = \frac{C(z^{-1})}{A(z^{-1})(1 - z^{-1})^d} e_k + \frac{B(z^{-1})z^{-d}}{A(z^{-1})} u_k \quad (41.9)$$

- **Box-Jenkins model** of orders p, q, n, m , BJ(p, q, n, m): BJ

$$y_k = \frac{C(z^{-1})}{D(z^{-1})} e_k + \frac{B(z^{-1})}{F(z^{-1})} u_k \quad (41.10)$$

n and m are the orders of the denominators corresponding to p and q .

- **Output error model** of orders n, m , OE(n, m): OE

$$y_k = e_k + \frac{B(z^{-1})}{F(z^{-1})} u_k \quad (41.11)$$

This is in fact a BJ(0, 0, p, q) model.

- **General linear model (GLM)**: GLM

$$A(z^{-1})y_k = \frac{C(z^{-1})}{D(z^{-1})} e_k + \frac{B(z^{-1})z^{-d}}{F(z^{-1})} u_k \quad (41.12)$$

Multiplying both members by $A(z^{-1})$, this becomes a BJ model. □

Definition 41.2. A digital system can have:

- a **finite impulse response (FIR)**, if its impulse response is zero after some time, or FIR
- an **infinite impulse response (IIR)**, otherwise. FIR

More formally, a FIR system is a system with an impulse response g_k that verifies

$$\exists_n \forall_{m>k} g_m = 0, \quad k, n, m \in \mathbb{N}_0 \quad (41.13)$$

and an IIR system is one that does not verify (41.13). □

Obviously,

- all FIR systems are stable;
- some stable digital systems have an IIR;
- all unstable digital systems have an IIR;
- a MA system has a FIR;
- a AR system has an IIR.

Example 41.1. Consider a MA(2) given by

$$y_k = e_k + 5e_{k-1} - 7e_{k-2} \quad (41.14)$$

Letting the input e_k , which should be stochastic, be an impulse, the MA impulse response is found:

$$y_0 = \underbrace{e_0}_1 + 5 \underbrace{e_{-1}}_0 - 7 \underbrace{e_{-2}}_0 = 1 \quad (41.15)$$

$$y_1 = \underbrace{e_1}_0 + 5 \underbrace{e_0}_1 - 7 \underbrace{e_{-1}}_0 = 5 \quad (41.16)$$

$$y_2 = \underbrace{e_2}_0 + 5 \underbrace{e_1}_0 - 7 \underbrace{e_0}_1 = -7 \quad (41.17)$$

$$y_3 = \underbrace{e_3}_0 + 5 \underbrace{e_2}_0 - 7 \underbrace{e_1}_0 = 0 \quad (41.18)$$

Clearly, $y_k = 0$ for all subsequent time samples. □

Remark 41.1. It is evident that a MA(q) has an impulse response which is zero after q time samples. \square

Example 41.2. Consider an AR(2) given by

$$y_k = e_k + 0.1y_{k-1} - 0.2y_{k-2} \quad (41.19)$$

This system is stable, since its transfer function is

$$\frac{y_k}{e_k} = \frac{1}{1 - 0.1z^{-1} + 0.2z^{-2}} \quad (41.20)$$

and its poles are $0.05 \pm 0.444j$. Letting the input e_k , which should be stochastic, be an impulse, the MA impulse response is found:

$$y_0 = \underbrace{e_0}_1 + 0.1 \underbrace{y_{-1}}_0 - 0.2 \underbrace{y_{-2}}_0 = 1 \quad (41.21)$$

$$y_1 = \underbrace{e_1}_0 + 0.1 \underbrace{y_0}_1 - 0.2 \underbrace{y_{-1}}_0 = 0.1 \quad (41.22)$$

$$y_2 = \underbrace{e_2}_0 + 0.1 \underbrace{y_1}_{0.1} - 0.2 \underbrace{y_0}_1 = 0.01 - 0.2 = -0.19 \quad (41.23)$$

$$y_3 = \underbrace{e_3}_0 + 0.1 \underbrace{y_2}_{0.19} - 0.2 \underbrace{y_1}_{0.1} = -0.019 - 0.02 = -0.039 \quad (41.24)$$

$$y_4 = \underbrace{e_4}_0 + 0.1 \underbrace{y_3}_{-0.039} - 0.2 \underbrace{y_2}_{-0.19} = -0.0039 + 0.038 = 0.0341 \quad (41.25)$$

\vdots

The impulse response converges to zero, but asymptotically. \square

The identification of a digital stochastic system follows the usual steps: identifying the order of the model or models, identifying model parameters, and assessing performance (which includes, if several models were found, selecting one of them). Model parameter identification is done as seen in Section 31.3, and performance assessment as was covered in Section 30.2. To find reasonable orders for models, however, there are methods particularly suited to stochastic systems.

41.2 Autocorrelation of a MA

Theorem 41.1. The autocorrelation $R_Y(\tau)$ of the output y_k of a MA(q) model verifies

$$R_Y(\tau) \neq 0, \text{ if } |\tau| \leq q \quad (41.26)$$

$$R_Y(\tau) = 0, \text{ if } |\tau| > q \quad (41.27)$$

provided that e_k is white noise.

Proof.

$$\begin{aligned}
R_Y(\tau) &= E[(e_k - c_1 e_{k-1} - c_2 e_{k-2} - \dots - c_q e_{k-q})(e_{k+\tau} - c_1 e_{k+\tau-1} - c_2 e_{k+\tau-2} - \dots - c_q e_{k+\tau-q})] \\
&= E[e_k e_{k+\tau} - c_1 e_k e_{k+\tau-1} - c_2 e_k e_{k+\tau-2} - \dots - c_q e_k e_{k+\tau-q} \\
&\quad - c_1 e_{k-1} e_{k+\tau} + c_1^2 e_{k-1} e_{k+\tau-1} + c_1 c_2 e_{k-1} e_{k+\tau-2} + \dots + c_1 c_q e_{k-1} e_{k+\tau-q} \\
&\quad - c_2 e_{k-2} e_{k+\tau} + c_1 c_2 e_{k-2} e_{k+\tau-1} + c_2^2 e_{k-2} e_{k+\tau-2} + \dots + c_2 c_q e_{k-2} e_{k+\tau-q} \\
&\quad \dots \\
&\quad - c_q e_{k-q} e_{k+\tau} + c_1 c_q e_{k-q} e_{k+\tau-1} + c_2 c_q e_{k-q} e_{k+\tau-2} + \dots + c_q^2 e_{k-q} e_{k+\tau-q}] \\
&= E[e_k e_{k+\tau}] - c_1 E[e_k e_{k+\tau-1}] - c_2 E[e_k e_{k+\tau-2}] - \dots - c_q E[e_k e_{k+\tau-q}] \\
&\quad - c_1 E[e_{k-1} e_{k+\tau}] + c_1^2 E[e_{k-1} e_{k+\tau-1}] + c_1 c_2 E[e_{k-1} e_{k+\tau-2}] + \dots + c_1 c_q E[e_{k-1} e_{k+\tau-q}] \\
&\quad - c_2 E[e_{k-2} e_{k+\tau}] + c_1 c_2 E[e_{k-2} e_{k+\tau-1}] + c_2^2 E[e_{k-2} e_{k+\tau-2}] + \dots + c_2 c_q E[e_{k-2} e_{k+\tau-q}] \\
&\quad \dots \\
&\quad - c_q E[e_{k-q} e_{k+\tau}] + c_1 c_q E[e_{k-q} e_{k+\tau-1}] + c_2 c_q E[e_{k-q} e_{k+\tau-2}] + \dots + c_q^2 E[e_{k-q} e_{k+\tau-q}] \\
&= R_E(\tau) - c_1 R_E(\tau - 1) - c_2 R_E(\tau - 2) - \dots - c_q R_E(\tau - q) \\
&\quad - c_1 R_E(\tau + 1) + c_1^2 R_E(\tau) + c_1 c_2 R_E(\tau - 1) + \dots + c_1 c_q R_E(\tau - q + 1) \\
&\quad - c_2 R_E(\tau + 2) + c_1 c_2 R_E(\tau + 1) + c_2^2 R_E(\tau) + \dots + c_2 c_q R_E(\tau - q + 2) \\
&\quad \dots \\
&\quad - c_q R_E(\tau + q) + c_1 c_q R_E(\tau + q - 1) + c_2 c_q R_E(\tau + q - 2) + \dots + c_q^2 R_E(\tau)
\end{aligned} \tag{41.28}$$

If $\tau > q$, or if $\tau < -q$, the autocorrelation $R_E(0)$ never appears above. Since e_k is white noise, $R_E(0)$ is the only autocorrelation that is not zero. If $-q \leq \tau \leq q$, autocorrelation $R_E(0)$ appears at least once, with at least one coefficient different from zero, and thus the result is not zero. \square

In practice, this can be used to find the order of a MA model with an input which is white noise: calculate the autocorrelation of the output, and the number of delays for which it is not zero is the order of the model. However, there are some problems:

- No noise, as we know, is really white. — Nothing can be done about this.
- The input may be clearly different from white noise. — This may be solved if a whitening filter can be applied at the input of the plant.
- There is always a finite number of samples, and consequently even white noise itself would not have a zero autocorrelation for $\tau \neq 0$. — This can be improved increasing the number of samples, if possible. But, whatever the case, it is always necessary to establish a threshold below which the autocorrelation is assumed as zero. The usual threshold is the one in the following result, quoted without proof.

Theorem 41.2. If x_k is normally distributed, its autocorrelation coefficient $\rho_X(\tau)$ for $\tau \neq 0$, computed from N samples, is equal to zero, with a 5% significance level, if

$$-\frac{1.959964}{\sqrt{N}} < \rho_X(\tau) < \frac{1.959964}{\sqrt{N}} \quad \square \tag{41.29}$$

Corollary 41.1. If x_k is normally distributed, its autocorrelation $R_X(\tau)$ for $\tau \neq 0$, computed from N samples, is equal to zero, with a 5% significance level, if

$$-\frac{1.959964\sigma_X^2}{\sqrt{N}} < R_X(\tau) < \frac{1.959964\sigma_X^2}{\sqrt{N}} \quad \square \tag{41.30}$$

Proof. This is an obvious consequence of the definition of $\rho_X(\tau)$ (37.58). \square

In (41.29)–(41.30), the number in the numerators, usually approximated by 1.96 and sometimes even by 2, appears as shown in Figure 41.1.

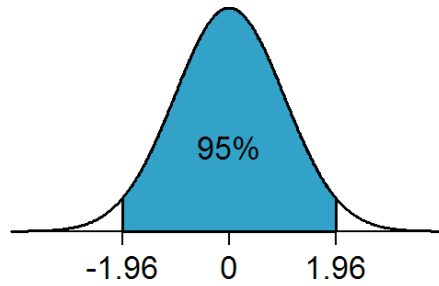


Figure 41.1: The normal distribution and the figure for a 5% significance level (source: Wikimedia). If X is a normally distributed random process with average 0 and variance 1, then $P(X > 1.96) = 2.5\%$, and $P(X < -1.96) = 2.5\%$. Consequently, $P(-1.96 < X < 1.96) = 95\%$.

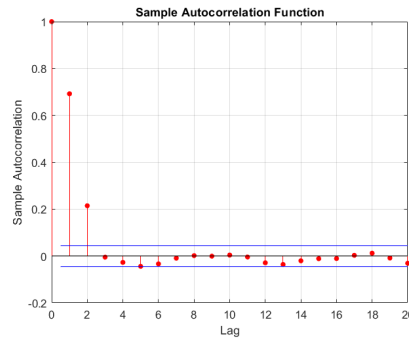


Figure 41.2: Autocorrelation coefficient of the output of plant (41.31) from Example 41.3.

Example 41.3. Consider a plant given by

$$\begin{aligned} \frac{y_k}{u_k} &= 3 + 4z^{-1} + 2z^{-2} \\ \Leftrightarrow y_k &= 3u_k + 4u_{k-1} + 2u_{k-2} \end{aligned} \quad (41.31)$$

The autocorrelation of this plant's output when fed 2001 samples of normally distributed white noise is shown in Figure 41.2 together with the threshold given by (41.29), which is $-0.044 < R_X(\tau) < 0.044$. The first three values, corresponding to three coefficients of the plant, are large enough to be outside the threshold. \square

The autocorrelation can be used as well to find the order of a model that only depends on an exogenous variable, as long as this variable can be approximated by white noise, and the MA part of the model can be neglected. In practice, orders adjacent to those found with this method should be considered for potential models as well, since numerical errors or noise can affect the results.

41.3 Partial autocorrelation of an AR

The role that the autocorrelation coefficient has in a MA model is taken by the partial autocorrelation coefficient when using an AR model instead. To introduce the partial autocorrelation, we must first establish some properties of an AR.

The mean of an AR is 0

Theorem 41.3. If y_k is given by an AR with white noise input, its mean value \bar{y} is 0.

Proof. Since

$$y_k = e_k + a_1 y_{k-1} + a_2 y_{k-2} + \dots + a_p y_{k-p} \quad (41.32)$$

then

$$E[y_0] = \underbrace{E[e_0]}_0 \quad (41.33)$$

$$E[y_1] = \underbrace{E[e_1]}_0 + a_1 \underbrace{E[y_0]}_0 \quad (41.34)$$

$$E[y_2] = \underbrace{E[e_2]}_0 + a_1 \underbrace{E[y_1]}_0 + a_2 \underbrace{E[y_0]}_0 \quad (41.35)$$

and so on for all outputs. \square

Corollary 41.2. If y_k is given by an AR with white noise input:

- the autocovariance $C_y(\tau) = R_y(\tau) - \bar{y}^2$ is equal to the autocorrelation $R_y(\tau)$,
- the autocorrelation at 0 $R_y(0) = \bar{y}^2 + \sigma_y^2$ is the variance σ_y^2 ,
- the autocorrelation coefficient $\rho_y(\tau) = \frac{C_y(\tau)}{\sigma_y^2}$ is given by

$$\rho_y(\tau) = \frac{R_y(\tau)}{R_y(0)} \quad \square \quad (41.36)$$

Theorem 41.4. If y_k is given by an AR with white noise input, its parameters *Yule-Walker equations* a_1, a_2, \dots, a_p can be found from the Yule-Walker equations:

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_{p-1} \\ a_p \end{bmatrix} = \begin{bmatrix} 1 & \rho_y(1) & \rho_y(2) & \cdots & \rho_y(p-2) & \rho_y(p-1) \\ \rho_y(1) & 1 & \rho_y(1) & \cdots & \rho_y(p-3) & \rho_y(p-2) \\ \rho_y(2) & \rho_y(1) & 1 & \cdots & \rho_y(p-4) & \rho_y(p-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_y(p-2) & \rho_y(p-3) & \rho_y(p-4) & \cdots & 1 & \rho_y(1) \\ \rho_y(p-1) & \rho_y(p-2) & \rho_y(p-3) & \cdots & \rho_y(1) & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho_y(1) \\ \rho_y(2) \\ \rho_y(3) \\ \vdots \\ \rho_y(p-1) \\ \rho_y(p) \end{bmatrix} \quad (41.37)$$

Proof. From (41.32), we can write

$$\begin{aligned} y_{k+1} &= e_{k+1} + a_1 y_k + a_2 y_{k-1} + \dots + a_p y_{k-p+1} \\ &= e_{k+1} + \sum_{j=1}^p a_j y_{k-j+1} \end{aligned} \quad (41.38)$$

To find the autocorrelation for a delay of 1, we multiply (41.38) by y_k :

$$\begin{aligned} y_k y_{k+1} &= y_k e_{k+1} + \sum_{j=1}^p a_j y_k y_{k-j+1} \\ \Rightarrow \underbrace{E[y_k y_{k+1}]}_{R_y(1)} &= \underbrace{E[y_k e_{k+1}]}_0 + \sum_{j=1}^p a_j \underbrace{E[y_k y_{k-j+1}]}_{R_y(-j+1)=R_y(j-1)} \end{aligned} \quad (41.39)$$

Notice that $E[y_k e_{k+1}] = 0$ because y_k is an output previous to e_{k+1} , and thus not yet affected by it. Dividing both sides by variance σ_y^2 ,

$$\rho_y(1) = \sum_{j=1}^p a_j \rho_y(j-1) \quad (41.40)$$

For the autocorrelation with 2 delays, we multiply (41.38) by y_{k-1} :

$$\begin{aligned} y_{k-1} y_{k+1} &= y_{k-1} e_{k+1} + \sum_{j=1}^p a_j y_{k-1} y_{k-j+1} \\ \Rightarrow \underbrace{E[y_{k-1} y_{k+1}]}_{R_y(2)} &= \underbrace{E[y_{k-1} e_{k+1}]}_0 + \sum_{j=1}^p a_j \underbrace{E[y_{k-1} y_{k-j+1}]}_{R_y(-j+2)=R_y(j-2)} \\ \Rightarrow \rho_y(2) &= \sum_{j=1}^p a_j \rho_y(j-2) \end{aligned} \quad (41.41)$$

And, for a general case of i delays, we multiply (41.38) by $y_{k-(i-1)}$:

$$\begin{aligned}
 y_{k-i+1}y_{k+1} &= y_{k-i+1}e_{k+1} + \sum_{j=1}^p a_j y_{k-i+1}y_{k-j+1} \\
 \Rightarrow \underbrace{E[y_{k-i+1}y_{k+1}]}_{R_y(i)} &= \underbrace{E[y_{k-i+1}e_{k+1}]}_0 + \sum_{j=1}^p a_j \underbrace{E[y_{k-i+1}y_{k-j+1}]}_{R_y(-j+i)=R_y(j-i)} \\
 \Rightarrow \rho_y(i) &= \sum_{j=1}^p a_j \rho_y(j-i)
 \end{aligned} \tag{41.42}$$

We can thus collect p equations that give the autocorrelation coefficients from $\rho_y(1)$ to $\rho_y(p)$:

$$\begin{bmatrix} \rho_y(1) \\ \rho_y(1) \\ \rho_y(1) \\ \vdots \\ \rho_y(p-1) \\ \rho_y(p) \end{bmatrix} = \begin{bmatrix} \rho_y(0) & \rho_y(1) & \rho_y(2) & \cdots & \rho_y(p-2) & \rho_y(p-1) \\ \rho_y(-1) & \rho_y(0) & \rho_y(1) & \cdots & \rho_y(p-3) & \rho_y(p-2) \\ \rho_y(-2) & \rho_y(-1) & \rho_y(0) & \cdots & \rho_y(p-4) & \rho_y(p-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_y(-p+2) & \rho_y(-p+3) & \rho_y(-p+4) & \cdots & \rho_y(0) & \rho_y(1) \\ \rho_y(-p+1) & \rho_y(-p+2) & \rho_y(-p+3) & \cdots & \rho_y(-1) & \rho_y(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_{p-1} \\ a_p \end{bmatrix} \tag{41.43}$$

Since $\rho_y(0) = 1$, and $\rho_y(-i) = \rho_y(i)$, and since the matrix is invertible, the Yule-Walker equations result. \square

Finding an AR from the Yule-Walker equations

The Yule-Walker equations can be used to find the parameters of an AR model, if only its order p is known. What if it is not? In that case, we could try different values of p , and check if the last coefficient is still not zero. When the last coefficient is zero, we would have found the order of the model.

This is in fact not so, since we could have a model in which one coefficient is zero but is followed by another which is not — as, say, in $y_k = e_k + 0.5y_{k-1} - 0.4y_{k-2}$, a model for which $a_1 = 0.5$, then $a_2 = 0$, and finally $a_3 = -0.4$, followed at last by $a_4 = a_5 = a_6 = \dots = 0$. So, what we really need to do is to try successive values of p ; at some point, the last coefficient will *always* be zero, how much we keep increasing p . The order of the model will be given by the value of p for which the last coefficient was *not* zero.

This is how, for an AR, we arrive at a variable that is different from zero for as long as there are coefficients in the model, and becomes zero when coefficients are over — just as the autocorrelation for a MA. Its definition is as follows.

Definition 41.3. Given a discrete signal y_k , let

$$\mathbf{R}_\tau = \begin{bmatrix} 1 & \rho_y(1) & \rho_y(2) & \cdots & \rho_y(\tau-2) & \rho_y(\tau-1) \\ \rho_y(1) & 1 & \rho_y(1) & \cdots & \rho_y(\tau-3) & \rho_y(\tau-2) \\ \rho_y(2) & \rho_y(1) & 1 & \cdots & \rho_y(\tau-4) & \rho_y(\tau-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_y(\tau-2) & \rho_y(\tau-3) & \rho_y(\tau-4) & \cdots & 1 & \rho_y(1) \\ \rho_y(\tau-1) & \rho_y(\tau-2) & \rho_y(\tau-3) & \cdots & \rho_y(1) & 1 \end{bmatrix} \tag{41.44}$$

$$\mathbf{r}_\tau = \begin{bmatrix} \rho_y(1) \\ \rho_y(1) \\ \rho_y(1) \\ \vdots \\ \rho_y(\tau-1) \\ \rho_y(\tau) \end{bmatrix} \tag{41.45}$$

$$\begin{bmatrix} a_{1,\tau} \\ a_{2,\tau} \\ a_{3,\tau} \\ \vdots \\ a_{\tau-1,\tau} \\ a_{\tau,\tau} \end{bmatrix} = \mathbf{R}_\tau^{-1} \mathbf{r}_\tau \tag{41.46}$$

partial autocorrelation coefficient

The **partial autocorrelation coefficient** $\phi_y(\tau)$ of signal y_k is given by $\phi_y(\tau) = a_{\tau,tau}$. The **partial autocorrelation** $\Phi_y(\tau)$ of signal y_k is given by $\Phi_y(\tau) = \phi_y(\tau)\sigma_y^2$.

The above holds for $\tau \neq 0$; by definition, $\phi_y(0) = \rho_y(0) = 1$ and thus $\Phi_y(\tau) = R_y(0) = \sigma_y^2$. \square

Corollary 41.3. The partial autocorrelation coefficient for $\tau = 1$ is given by

$$\mathbf{R}_1 = [1] \tag{41.47}$$

$$\mathbf{r}_1 = [\rho_y(1)] \tag{41.48}$$

$$a_{1,1} = 1^{-1}\rho_y(1) \tag{41.49}$$

Thus $\phi_y(1) = \rho_y(1)$ and $\Phi_y(1) = R_y(1)$. \square

Remark 41.2. The partial autocorrelation is so called because it correlates the output with one of its previous values, removing the dependence from intermediate outputs; that is to say,

$$\Phi_y(\tau) = E[y_k y_{k+\tau} | y_{k+1}, y_{k+2}, \dots, y_{k+\tau-1}] \tag{41.50}$$

The conditional expected value above can be justified as follows. The last line of

$$\mathbf{R}_\tau [a_{1,\tau} \ a_{2,\tau} \ a_{3,\tau} \ \dots \ a_{\tau-1,\tau} \ a_{\tau,tau}]^T = \mathbf{r}_\tau \tag{41.51}$$

is

$$\begin{aligned} \rho_y(\tau-1)a_{1,\tau} + \rho_y(\tau-2)a_{2,\tau} + \rho_y(\tau-3)a_{3,\tau} + \dots + \rho_y(1)a_{\tau-1,\tau} + a_{\tau,tau} &= \rho_y(\tau) \\ \Leftrightarrow \phi_y(\tau) = a_{\tau,tau} = \rho_y(\tau) - \sum_{j=1}^{\tau-1} \rho_y(\tau-j)a_{j,\tau} \\ \Leftrightarrow \Phi_y(\tau) = R_y(\tau) - \sum_{j=1}^{\tau-1} R_y(\tau-j)a_{j,\tau} \\ &= E[y(k)y_{k+\tau}] - \sum_{j=1}^{\tau-1} a_{j,\tau} E[y(k)y_{k+\tau}] \end{aligned} \tag{41.52}$$

To the extent that the estimated model coefficients $a_{j,\tau}$ are correct, this is (41.50). And since, from (41.50),

$$\Phi_y(0) = E[y_k y_k] \tag{41.53}$$

$$\Phi_y(1) = E[y_k y_{k+1}] \tag{41.54}$$

without any conditional expectation (as there are no intermediate values of y), it makes sense that $\Phi_y(1) = R_y(1)$, as we saw above, and also that we should define $\Phi_y(0) = R_y(0)$. \square

Just as in (41.29)–(41.30) for the autocorrelation, the partial autocorrelation of a signal X_k computed from N samples is usually considered zero, with a 5% significance level, if

$$-\frac{1.959964}{\sqrt{N}} < \phi_X(\tau) < \frac{1.959964}{\sqrt{N}} \tag{41.55}$$

$$-\frac{1.959964\sigma_X^2}{\sqrt{N}} < \Phi_X(\tau) < \frac{1.959964\sigma_X^2}{\sqrt{N}} \tag{41.56}$$

In an AR model, additive noise at the output keeps influencing it in future samples. For this reason, the presence of noise easily leads to an overestimation of the order of the model.

Example 41.4. Consider a plant given by

$$\begin{aligned} \frac{y_k}{u_k} &= \frac{1}{3 + 4z^{-1} + 2z^{-2}} \\ \Leftrightarrow y_k &= \frac{1}{3}u_k - \frac{4}{3}y_{k-1} - \frac{2}{3}y_{k-2} \end{aligned} \tag{41.57}$$

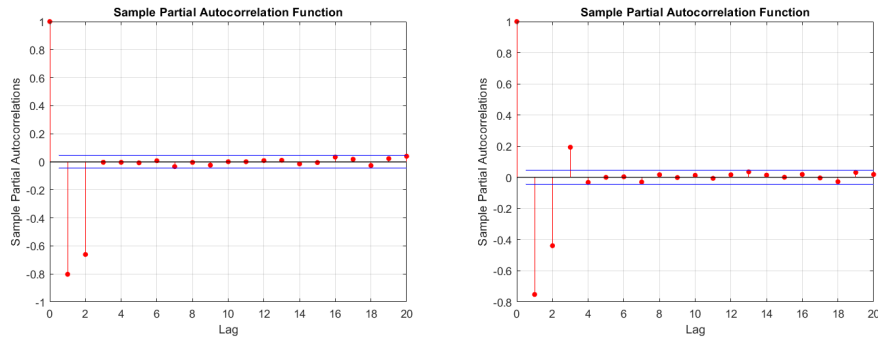


Figure 41.3: Left: partial autocorrelation of the output of plant (41.57) from Example 41.4. Right: the same, when there is noise as in (41.58).

The partial autocorrelation of this plant's output when fed 2001 samples of normally distributed white noise is shown in Figure 41.3 together with the threshold given by (41.29), which is $-0.044 < R_X(\tau) < 0.044$. The first three values, corresponding to three coefficients of the plant, are large enough to be outside the threshold.

The partial autocorrelation is also shown for the situation when there is additive noise at the output:

$$y_k = \frac{1}{3}u_k - \frac{4}{3}y_{k-1} - \frac{2}{3}y_{k-2} + 0.05e_k \quad (41.58)$$

Here e_k is normally distributed white noise, with zero-mean and variance 1. Notice how there are now four delays outside the threshold. \square

MATLAB's *commands*
autocorr and *parcorr*

MATLAB commands `autocorr` and `parcorr` plot the autocorrelation and the partial autocorrelation of a variable, and plot the threshold given by (41.29) approximating the numerator by 2.

41.4 Finding the orders of an ARMA

When we have the output of an ARMA model,

- we can find from the autocorrelation the order q of an MA(q) that might model it instead;
- we can find from the partial autocorrelation the order p of an AR(p) that might model it instead.

The number of coefficients of the ARMA model should not be larger than the smallest of these two orders; otherwise, there would be no interest in using an ARMA. Once this upper limit $\min\{p, q\}$ is established, all ARMA(p', q') models such that $p' + q' \leq \min\{p, q\}$ should be tried, and the performance of the resulting models compared with the AIC or the BIC, to determine the best option.

Example 41.5. Figure 41.4 shows the autocorrelation and partial autocorrelation coefficients of the 10000 samples long output of

$$G(z^{-1}) = \frac{4 + 5z^{-1} + 6z^{-2}}{1 + \frac{1}{2}z^{-1} + \frac{1}{3}z^{-2}} \quad (41.59)$$

From the figures, we can expect an AR model of order 13 or a MA model of order 6 to be able to model this output. So, an ARMA model should have, at most, 6 coefficients. Thus, we should find ARMA(p, q) models such that $p + q \leq 6$, and compare their AIC or BIC to choose one. \square

Another way to determine the orders of an ARMA model would be to verify which of them are needed using statistical tests such as those using t -values or p -values. We will not study this possibility further.

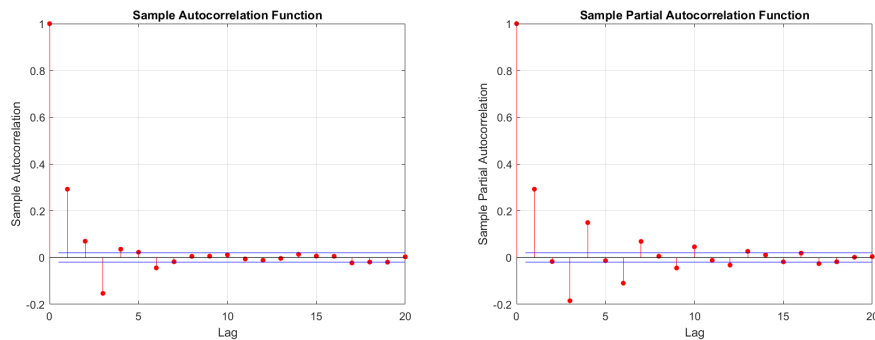


Figure 41.4: Autocorrelation and partial autocorrelation coefficients of the output of (41.59).

Glossary

I understood him in many Things, and let him know, I was very well pleas'd with him; in a little Time I began to speak to him, and teach him to speak to me; and first, I made him know his Name should be *Friday*, which was the Day I sav'd his Life; I call'd him so for the Memory of the Time; (...)

Daniel DEFOE (1660? — †1731), *The life and Strange Surprizing adventures of Robinson Crusoe, of York, Mariner* (1719)

autoregressive model modelo autorregressivo

autoregressive exogenous model modelo autorregressivo com variável exógena

autoregressive, integral, moving average, exogenous model modelo autorregressivo integral de média móvel com variável exógena

autoregressive, integral, moving average model modelo autorregressivo integral de média móvel

autoregressive, moving average, exogenous model modelo autorregressivo de média móvel com variável exógena

autoregressive, moving average model modelo autorregressivo de média móvel

Box-Jenkins model modelo de Box-Jenkins

finite impulse response resposta finita ao impulso, resposta impulsional finita

general linear model modelo linear genérico

infinite impulse response resposta infinita ao impulso, resposta impulsional infinita

moving average model modelo de média móvel

output error model modelo de média da saída

partial autocorrelation autocorrelação parcial

partial autocorrelation coefficient coeficiente de autocorrelação parcial

Exercises

1. Question.

Chapter 42

Control of stochastic systems

Prince Andrew listened attentively to Bagration's colloquies with the commanding officers and the orders he gave them and, to his surprise, found that no orders were really given, but that Prince Bagration tried to make it appear that everything done by necessity, by accident, or by the will of subordinate commanders was done, if not by his direct command, at least in accord with his intentions. Prince Andrew noticed, however, that though what happened was due to chance and was independent of the commander's will, owing to the tact Bagration showed, his presence was very valuable. Officers who approached him with disturbed countenances became calm; soldiers and officers greeted him gaily, grew more cheerful in his presence, and were evidently anxious to display their courage before him.

Leo TOLSTOY (1828 — †1910), *War and Peace* (1869), II 17 (transl. Louise Maude and Aylmer Maude, 1922–1923)

In this chapter, three different ways of designing closed-loop controllers for stochastic processes are presented. The first two can only be applied to follow constant references, i.e. in regulation problems.

42.1 Minimum variance control

This technique can be used to design a regulator. The effect of the noise in the output can be minimised, but not eliminated.

Theorem 42.1. Consider a plant with an output given by

$$y(t) = \frac{B(z^{-1})}{A(z^{-1})} z^{-d} u(t) + \frac{C(z^{-1})}{A(z^{-1})} e(t) \quad (42.1)$$

where $u(t)$ is a manipulated input, and $e(t)$ is white noise. (If there is no pure delay of d sample times from $u(t)$ to the output, then $d = 0$.) This plant is controlled in closed loop, as seen in Figure 42.1, and, without loss of generality, reference $r(t)$ is taken as 0 (otherwise a variable change is used). In this situation, the controller that minimises the variance of error $\varepsilon(t) = -y(t)$

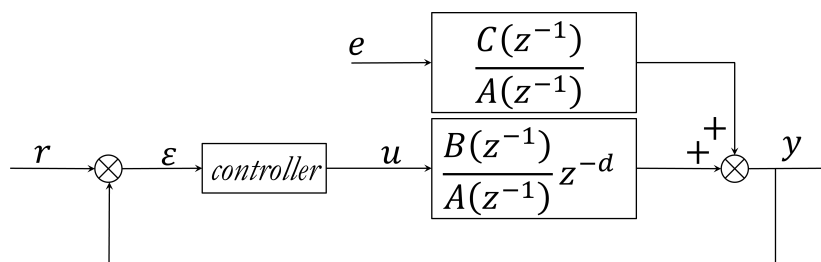


Figure 42.1: Closed loop control of a stochastic digital system.

is given by

$$u(t) = -\frac{G(z^{-1})}{B(z^{-1})F(z^{-1})}y(t) \quad (42.2)$$

where $F(z^{-1})$ and $G(z^{-1})$ are found solving

$$C(z^{-1}) = F(z^{-1})A(z^{-1}) + G(z^{-1})z^{-d} \quad (42.3)$$

The output when control law (42.2) is used is given by the MA

$$y(t) = F(z^{-1})e(t) \quad (42.4)$$

Proof. Dropping the dependence on z^{-1} to alleviate the notation, (42.1) can be rewritten as

$$y(t+d) = \frac{B}{A}u(t) + \frac{C}{A}e(t+d) \quad (42.5)$$

(42.3) is the same as separating the dynamics of the noise $\frac{C(z^{-1})}{A(z^{-1})}$ into two parts by means of polynomial division:

$$\begin{aligned} \frac{C}{A} &= F + \frac{G}{A}z^{-d} \\ \Leftrightarrow F &= \frac{C - Gz^{-d}}{A} \end{aligned} \quad (42.6)$$

Replacing (42.6) in (42.5),

$$y(t+d) = \frac{B}{A}u(t) + Fe(t+d) + \frac{G}{A}e(t) \quad (42.7)$$

Now notice that (42.1) can also be rewritten as

$$e(t) = \frac{A}{C}y(t) - \frac{B}{C}z^{-d}u(t) \quad (42.8)$$

Replacing this in (42.7), and using (42.6),

$$\begin{aligned} y(t+d) &= \frac{B}{A}u(t) + Fe(t+d) + \frac{G}{A}\left(\frac{A}{C}y(t) - \frac{B}{C}z^{-d}u(t)\right) \\ &= Fe(t+d) + \frac{G}{C}y(t) - \frac{BG}{CA}z^{-d}u(t) + \frac{B}{A}u(t) \\ &= Fe(t+d) + \frac{G}{C}y(t) + \underbrace{\frac{B(C - Gz^{-d})}{CA}}_{\frac{BF}{C}}u(t) \\ &= \underbrace{Fe(t+d)}_{\substack{\text{this can only be} \\ \text{known at time } t+d}} + \underbrace{\left(\frac{G}{C}y(t) + \frac{BF}{C}u(t)\right)}_{\substack{\text{this is already known at time } t}} \end{aligned} \quad (42.9)$$

Consequently, to minimise $E[(\varepsilon(t))^2]$, we minimise

$$\begin{aligned} E[(y(t+d))^2] &= E\left[\left(Fe(t+d) + \frac{G}{C}y(t) + \frac{BF}{C}u(t)\right)^2\right] \\ &= E[(Fe(t+d))^2] + E\left[\left(\frac{G}{C}y(t) + \frac{BF}{C}u(t)\right)^2\right] \end{aligned} \quad (42.10)$$

where there are no cross-terms because white noise $e(t+d)$ is independent from $y(t)$ and $u(t)$. There is nothing to do about the expected value that depends on the white noise, but the second term can be minimised:

$$\frac{G}{C}y(t) + \frac{BF}{C}u(t) = 0 \quad (42.11)$$

From here (42.2) is immediate.

As to the performance of the controller, replacing (42.2) in (42.9) gives

$$y(t+d) = Fe(t+d) + \frac{G}{C}y(t) + \underbrace{\frac{BF}{C} \left(-\frac{G}{BF}y(t) \right)}_{-\frac{G}{C}y(t)} \quad (42.12)$$

which is (42.4). \square

Remark 42.1. Let the orders of polynomials $A(z^{-1})$, $B(z^{-1})$, etc. be n_A , n_B , etc.. Orders n_F and n_G must be large enough for (42.3) to be possible. They can be found by inspection in each case; these orders turn out to be, in every situation, the following:

$$n_F = d - 1 \quad (42.13)$$

$$n_G = \max\{n_A - 1, n_C - d\} \quad \square \quad (42.14)$$

Example 42.1. Consider a plant with an output given by

$$\begin{aligned} y(t) &= 0.5y(t - T_s) + 0.2y(t - 2T_s) + 0.7u(t - T_s) + 0.3u(t - 2T_s) \\ &\quad + e(t) + 0.4e(t - T_s) + 0.1e(t - 2T_s) \\ \Leftrightarrow y(t)(1 - 0.5z^{-1} - 0.2z^{-2}) &= u(t) \underbrace{(0.7z^{-1} + 0.3z^{-2})}_{z^{-1}(0.7+0.3z^{-1})} + e(t)(1 + 0.4z^{-1} + 0.1z^{-2}) \end{aligned} \quad (42.15)$$

That is to say,

$$A(z^{-1}) = 1 - 0.5z^{-1} - 0.2z^{-2} \quad (42.16)$$

$$B(z^{-1}) = 0.7 + 0.3z^{-1} \quad (42.17)$$

$$C(z^{-1}) = 1 + 0.4z^{-1} + 0.1z^{-2} \quad (42.18)$$

$$d = 1 \quad (42.19)$$

We must solve (42.3):

$$1 + 0.4z^{-1} + 0.1z^{-2} = (1 - 0.5z^{-1} - 0.2z^{-2})F(z^{-1}) + z^{-1}G(z^{-1}) \quad (42.20)$$

The left side of the equation is of order 2. The right side will be of order 2 if $n_F = 0$ and $n_G = 1$, i.e. $F(z^{-1}) = f_0$ and $G(z^{-1}) = g_0 + g_1z^{-1}$. Notice that the only independent term on the right side will be f_0 ; thus we must have $f_0 = 1$. This still leaves two variables, g_0 and g_1 , which suffice to make the equation possible.

Instead of reasoning like this we could apply (42.13)–(42.14) and write

$$\begin{aligned} 1 + 0.4z^{-1} + 0.1z^{-2} &= (1 - 0.5z^{-1} - 0.2z^{-2})(f_0) + z^{-1}(g_0 + g_1z^{-1}) \\ &= f_0 - 0.5f_0z^{-1} - 0.2f_0z^{-2} + g_0z^{-1} + g_1z^{-2} \end{aligned} \quad (42.21)$$

Equalling the coefficients of the same order on both sides,

$$\begin{cases} 1 = f_0 \\ 0.4 = -0.5f_0 + g_0 \\ 0.1 = -0.2f_0 + g_1 \end{cases} \Leftrightarrow \begin{cases} f_0 = 1 \\ g_0 = 0.9 \\ g_1 = 0.3 \end{cases} \Rightarrow \begin{cases} F(z^{-1}) = 1 \\ G(z^{-1}) = 0.9 + 0.3z^{-1} \end{cases} \quad (42.22)$$

Thus

$$u(t) = -y(t) \frac{0.9 + 0.3z^{-1}}{0.7 + 0.3z^{-1}} \quad (42.23)$$

and this control law will achieve $y(t) = e(t)$. This is not surprising, since the noise itself cannot be eliminated; it is impossible to do any better.

Figure 42.2 shows the output, equal to the error, and the control action, for a simulation with 100 time steps. \square

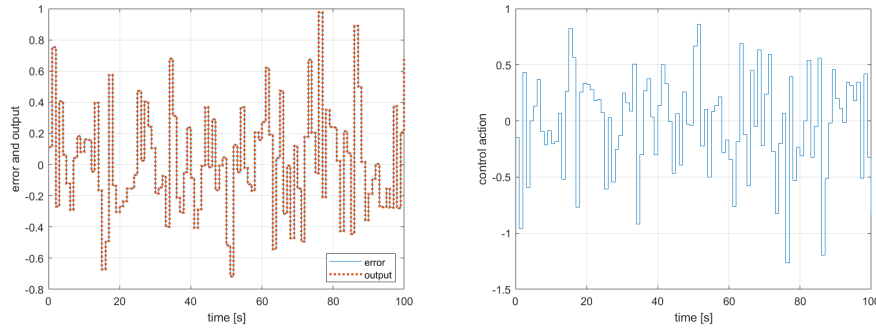


Figure 42.2: Simulation of plant (42.15) controlled by (42.23), from Example 42.1.

Example 42.2. Consider a plant similar to that of Example 42.1, save that now the output is given by

$$\begin{aligned}
 y(t) &= 0.5y(t - T_s) + 0.2y(t - 2T_s) + 0.7u(t - 2T_s) + 0.3u(t - 3T_s) \\
 &\quad + e(t) + 0.4e(t - T_s) + 0.1e(t - 2T_s) \\
 \Leftrightarrow y(t) \underbrace{(1 - 0.5z^{-1} - 0.2z^{-2})}_{A(z^{-1})} &= u(t) \underbrace{(0.7z^{-2} + 0.3z^{-3})}_{\substack{z^{-2} \\ z^{-d}} \underbrace{(0.7 + 0.3z^{-1})}_{B(z^{-1})}} + e(t) \underbrace{(1 + 0.4z^{-1} + 0.1z^{-2})}_{C(z^{-1})}
 \end{aligned} \tag{42.24}$$

This is in fact the same plant of Example 42.1 but with $d = 2$. We must solve (42.3):

$$1 + 0.4z^{-1} + 0.1z^{-2} = (1 - 0.5z^{-1} - 0.2z^{-2}) F(z^{-1}) + z^{-2} G(z^{-1}) \tag{42.25}$$

The left side of the equation is still of order 2. The right side will be of order 2 if $n_F = 0$ and $n_G = 0$, i.e. $F(z^{-1}) = f_0$ and $G(z^{-1}) = g_0$. Once more, the only independent term on the right side will be f_0 ; thus we must have $f_0 = 1$. This leaves only one variable, g_0 , and the equation is impossible. Consequently, we need $n_F = 1$ and $n_G = 1$, i.e. $F(z^{-1}) = f_0 + f_1z^{-1}$ and $G(z^{-1}) = g_0 + g_1z^{-1}$; the right member will be of order 3, and the terms of order 3 will have to cancel, since there is none on the right member.

Instead of reasoning like this we could apply (42.13)–(42.14) and write

$$\begin{aligned}
 1 + 0.4z^{-1} + 0.1z^{-2} &= (1 - 0.5z^{-1} - 0.2z^{-2}) (f_0 + f_1z^{-1}) + z^{-2} (g_0 + g_1z^{-1}) \\
 &= f_0 - 0.5f_0z^{-1} - 0.2f_0z^{-2} + f_1z^{-1} - 0.5f_1z^{-2} - 0.2f_1z^{-3} + g_0z^{-2} + g_1z^{-3}
 \end{aligned} \tag{42.26}$$

Equalling the coefficients of the same order on both sides,

$$\begin{cases} 1 = f_0 \\ 0.4 = -0.5f_0 + f_1 \\ 0.1 = -0.2f_0 - 0.5f_1 + g_0 \\ 0 = -0.2f_1 + g_1 \end{cases} \Leftrightarrow \begin{cases} f_0 = 1 \\ f_1 = 0.9 \\ g_0 = 0.3 + 0.5 \times 0.9 = 0.75 \\ g_1 = 0.2 \times 0.9 = 0.18 \end{cases} \Rightarrow \begin{cases} F(z^{-1}) = 1 + 0.9z^{-1} \\ G(z^{-1}) = 0.75 + 0.18z^{-1} \end{cases} \tag{42.27}$$

Thus

$$u(t) = -y(t) \frac{0.75 + 0.18z^{-1}}{(0.7 + 0.3z^{-1})(1 + 0.9z^{-1})} \tag{42.28}$$

and this control law will achieve $y(t) = e(t) (1 + 0.9z^{-1})$. Notice how one additional delay from the control action to the output leads to more complicated calculations, to a controller of higher order (there are two poles now rather than only one), and to an output with more noise.

Figure 42.3 shows the output, no longer equal to the error, and the control action, for a simulation with 100 time steps. \square

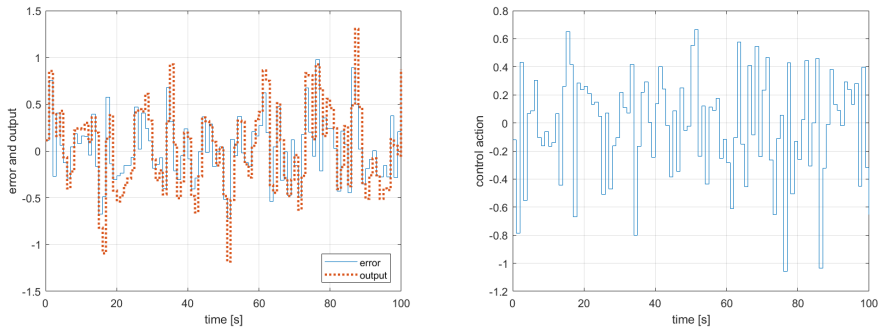


Figure 42.3: Simulation of plant (42.24) controlled by (42.28), from Example 42.2.

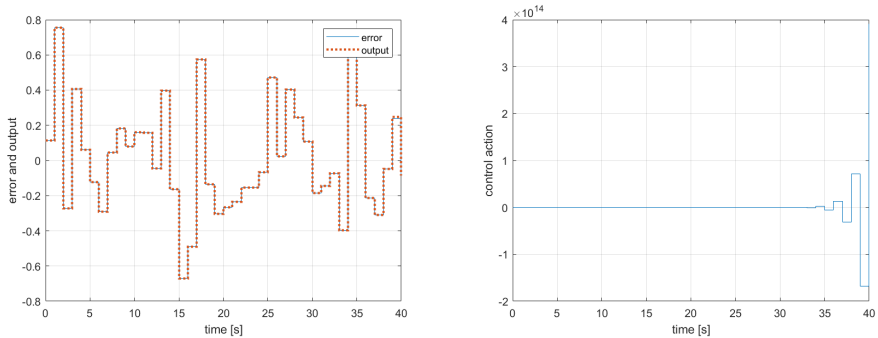


Figure 42.4: Simulation of plant (42.29) controlled by (42.33), from Example 42.3.

of a non-minimum phase zero

Example 42.3. Consider yet another plant similar to that of Example 42.1, save that now the output is given by

$$\begin{aligned}
 y(t) &= 0.5y(t - T_s) + 0.2y(t - 2T_s) + 0.3u(t - T_s) + 0.7u(t - 2T_s) \\
 &\quad + e(t) + 0.4e(t - T_s) + 0.1e(t - 2T_s) \\
 \Leftrightarrow y(t) \underbrace{(1 - 0.5z^{-1} - 0.2z^{-2})}_{A(z^{-1})} &= u(t) \underbrace{\left(\underbrace{z^{-1}}_{z^{-d}} \underbrace{(0.3 + 0.7z^{-1})}_{B(z^{-1})} \right)}_{B(z^{-1})} + e(t) \underbrace{(1 + 0.4z^{-1} + 0.1z^{-2})}_{C(z^{-1})}
 \end{aligned}
 \tag{42.29}$$

This is the same plant of Example 42.1 but this time with a non-minimum phase zero, since

$$B(z^{-1}) = 0 \Leftrightarrow 0.7z^{-1} = -0.3 \Leftrightarrow z^{-1} = -\frac{3}{7} \Leftrightarrow z = -\frac{7}{3}
 \tag{42.30}$$

which is outside the unit radius circle. Solving (42.3) is done as in Example 42.1; the results are the same:

$$F(z^{-1}) = 1
 \tag{42.31}$$

$$G(z^{-1}) = 0.9 + 0.3z^{-1}
 \tag{42.32}$$

Consequently, the control action will be

$$u(t) = -y(t) \frac{0.9 + 0.3z^{-1}}{0.3 + 0.7z^{-1}}
 \tag{42.33}$$

and this control law will achieve $y(t) = e(t)$.

Figure 42.4 shows the output, which equal to the error, and the control action that achieves this result. Notice how the amplitude of the control action grows exponentially. This is because control law is unstable: (42.2) cancels the zeros of the plant, and so in this case the controller tries to cancel the non-minimum phase zero. Such a controller is of course impossible in practice: as

soon as actuators saturate, the output will be far from what is desired. (In fact, even in simulation, sooner or later an overflow prevents good results; that is the reason why Figure 42.4 only shows 40 time steps). \square

What to do with non-minimum phase zeros

This last example shows that, if there are non-minimum phase zeros in $B(z^{-1})$, we cannot have (42.2), but must use only the minimum-phase zeros in the denominator. That is to say, we separate $B(z^{-1})$ as product

$$B(z^{-1}) = B_{\min}(z^{-1})B_{\text{non}}(z^{-1}) \quad (42.34)$$

so that $B_{\text{non}}(z^{-1})$ will have all the non-minimum phase zeros, and $B_{\min}(z^{-1})$ all the minimum phase zeros; and then make

$$u(t) = -\frac{G(z^{-1})}{B_{\min}(z^{-1})F(z^{-1})}y(t) \quad (42.35)$$

Price to pay for a stable control action

The effect of the error in the output will be larger, but the control system (and in particular the control action) will be stable.

Theorem 42.2. Consider a plant with an output given by

$$y(t) = \frac{\overbrace{B_{\min}(z^{-1})B_{\text{non}}(z^{-1})}^{B(z^{-1})}}{A(z^{-1})}z^{-d}u(t) + \frac{C(z^{-1})}{A(z^{-1})}e(t) \quad (42.36)$$

where $u(t)$ is a manipulated input, and $e(t)$ is white noise. (If there is no pure delay of d sample times from $u(t)$ to the output, then $d = 0$.) All the roots of $B_{\text{non}}(z^{-1})$ have positive real parts; none of the roots of $B_{\min}(z^{-1})$ do. This plant is controlled in closed loop, as seen in Figure 42.1, and, without loss of generality, reference $r(t)$ is taken as 0 (otherwise a variable change is used). In this situation, the controller that minimises the variance of error $\varepsilon(t) = -y(t)$ and ensures a stable control action is given by

$$u(t) = -\frac{G(z^{-1})}{B_{\min}(z^{-1})F(z^{-1})}y(t) \quad (42.37)$$

where $F(z^{-1})$ and $G(z^{-1})$ are found solving

$$C(z^{-1})B_{\min}(z^{-1}) = F(z^{-1})A(z^{-1}) + B_{\text{non}}(z^{-1})G(z^{-1})z^{-d} \quad (42.38)$$

The output when control law (42.37) is used is given by

$$y(t) = \frac{F(z^{-1})}{B_{\min}(z^{-1})}e(t) \quad (42.39)$$

Proof. Replacing the desired control action (42.37) in (42.36), and dropping the dependence on t and z^{-1} to alleviate the notation,

$$\begin{aligned} y &= \frac{B_{\min}B_{\text{non}}}{A}z^{-d} \left(-\frac{G}{B_{\min}F}y \right) + \frac{C}{A}e \\ \Leftrightarrow y \underbrace{\left(1 + \frac{B_{\text{non}}Gz^{-d}}{AF} \right)}_{\frac{AF+B_{\text{non}}Gz^{-d}}{AF}} &= \frac{C}{A}e \\ \Leftrightarrow y &= \frac{C}{A} \frac{AF}{AF+B_{\text{non}}Gz^{-d}}e = \frac{CF}{AF+B_{\text{non}}Gz^{-d}}e \end{aligned} \quad (42.40)$$

If (42.38) holds, the denominator is $C(z^{-1})B_{\min}(z^{-1})$ and (42.39) is obtained. \square

Remark 42.2. As before, orders n_F and n_G must be enough for (42.57) to be possible, and can be found by inspection in each case; these orders now turn out to be:

$$n_F = n_{B_{\text{non}}} + d - 1 \quad (42.41)$$

$$n_G = \max\{n_A - 1, n_C - d\} \quad \square \quad (42.42)$$

Example 42.4. We are now in position of finding a much better controller for the non-minimum phase plant (42.29) of Example 42.3. We have

$$A(z^{-1}) = 1 - 0.5z^{-1} - 0.2z^{-2} \quad (42.43)$$

$$B_{\text{non}}(z^{-1}) = 0.3 + 0.7z^{-1} \quad (42.44)$$

$$B_{\text{min}}(z^{-1}) = 1 \quad (42.45)$$

$$C(z^{-1}) = 1 + 0.4z^{-1} + 0.1z^{-2} \quad (42.46)$$

$$z^{-d} = z^{-1} \quad (42.47)$$

Instead of solving (42.3), we solve (42.38):

$$(1 + 0.4z^{-1} + 0.1z^{-2}) \times 1 = (1 - 0.5z^{-1} - 0.2z^{-2}) F(z^{-1}) + z^{-1} (0.3 + 0.7z^{-1}) G(z^{-1}) \quad (42.48)$$

The reasoning of Example 42.2 applies here and shows that $n_F = 1$ and $n_G = 1$, i.e. $F(z^{-1}) = f_0 + f_1z^{-1}$ and $G(z^{-1}) = g_0 + g_1z^{-1}$. Or we could apply (42.41)–(42.42) and write

$$\begin{aligned} (1 + 0.4z^{-1} + 0.1z^{-2}) \times 1 &= (1 - 0.5z^{-1} - 0.2z^{-2}) (f_0 + f_1z^{-1}) + z^{-1} (0.3 + 0.7z^{-1}) (g_0 + g_1z^{-1}) \\ &= f_0 - 0.5f_0z^{-1} - 0.2f_0z^{-2} + f_1z^{-1} - 0.5f_1z^{-2} - 0.2f_1z^{-3} \\ &\quad + 0.3g_0z^{-1} + 0.3g_1z^{-2} + 0.7g_0z^{-2} + 0.7g_1z^{-3} \end{aligned} \quad (42.49)$$

Equating the coefficients of the same order on both sides,

$$\begin{cases} 1 = f_0 \\ 0.4 = -0.5f_0 + f_1 + 0.3g_0 \\ 0.1 = -0.2f_0 - 0.5f_1 + 0.3g_1 + 0.7g_0 \\ 0 = -0.2f_1 + 0.7g_1 \end{cases} \Leftrightarrow \begin{cases} f_0 = 1 \\ f_1 + 0.3g_0 = 0.9 \\ -0.5f_1 + 0.7g_0 + 0.3g_1 = 0.3 \\ -0.2f_1 + 0.7g_1 = 0 \end{cases} \quad (42.50)$$

The last three equations are better solved with MATLAB:

```
>> [1 0.3 0; -0.5 0.7 0.3; -0.2 0 0.7] \ [0.9 0.3 0]'
```

```
ans =
```

```
0.6551
0.8163
0.1872
```

Thus

$$F(z^{-1}) = 1 + 0.6551z^{-1} \quad (42.51)$$

$$G(z^{-1}) = 0.8163 + 0.1872z^{-1} \quad (42.52)$$

and so, according to (42.37),

$$u(t) = -y(t) \frac{0.8163 + 0.1872z^{-1}}{1 \times (1 + 0.6551z^{-1})} \quad (42.53)$$

As shown in Figure 42.5, this control law will achieve $y(t) = e(t) (1 + 0.6551z^{-1})$, an output larger than that of Example 42.3: the controller (42.37) is not in fact a minimum variance controller; it is known as a **sub-optimal minimum variance controller**. However, this output is in practice feasible (since the control action is now bounded), unlike the original controller, which would be optimal — if only it could ever work. \square

Sub-optimal minimum variance control

Remark 42.3. Minimum variance control minimises the variance of the error $E[(\varepsilon(t))^2]$, however large the control action has to be. It can be generalised so as to minimise a cost function J instead, given by

$$J = P E[(\varepsilon(t))^2] + Q E[(u(t))^2] \quad (42.54)$$

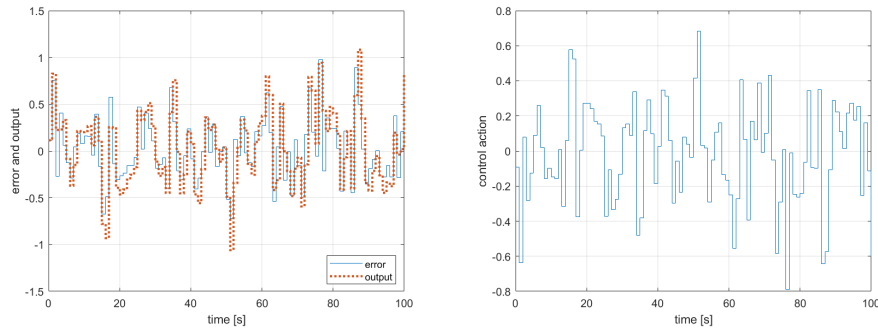


Figure 42.5: Simulation of plant (42.29) controlled by (42.53), from Example 42.4.

where P and Q are weights. J defines how large we are willing to have the control action to lower the variance of the output. Of course, when $Q = 0$ we get minimum variance control back.

We will not explore this possibility, called **generalised minimum variance control**, and which is a particular case of **optimal control**, addressed below in Section 43.5. \square

42.2 Pole-assignment control

Digital stochastic systems can be controlled by pole-assignment, as done in Chapters 21 and 22 for deterministic plants in continuous time and in Chapter 27 for digital deterministic plants. We will assume a regulation problem, as we did for minimum variance control.

Theorem 42.3. Consider a plant with an output given by

$$y(t) = \frac{B(z^{-1})}{A(z^{-1})} z^{-d} u(t) + \frac{C(z^{-1})}{A(z^{-1})} e(t) \quad (42.55)$$

where $u(t)$ is a manipulated input, and $e(t)$ is white noise. (If there is no pure delay of d sample times from $u(t)$ to the output, then $d = 0$.) This plant is controlled in closed loop, as seen in Figure 42.1, and, without loss of generality, reference $r(t)$ is taken as 0 (otherwise a variable change is used). In this situation, the controller that places the poles of the closed loop in the locations defined by $T(z^{-1})$ is given by

$$u(t) = -\frac{G(z^{-1})}{F(z^{-1})} y(t) \quad (42.56)$$

where $F(z^{-1})$ and $G(z^{-1})$ are found solving

$$T(z^{-1})C(z^{-1}) = F(z^{-1})A(z^{-1}) + B(z^{-1})G(z^{-1})z^{-d} \quad (42.57)$$

The output when control law (42.2) is used is given by

$$y(t) = \frac{F(z^{-1})}{T(z^{-1})} e(t) \quad (42.58)$$

Proof. Replacing (42.56) in (42.55), and dropping the dependence on t and z^{-1} to alleviate the notation,

$$\begin{aligned} y &= \frac{Bz^{-d}}{A} \left(-\frac{G}{F} y \right) + \frac{C}{A} e \\ \Leftrightarrow y \left(\underbrace{1 + \frac{BGz^{-d}}{AF}}_{\frac{AF+BGz^{-d}}{AF}} \right) &= \frac{C}{A} e \\ \Leftrightarrow y &= \frac{C}{A} \frac{AF}{AF+BGz^{-d}} e = \frac{CF}{AF+BGz^{-d}} e \end{aligned} \quad (42.59)$$

We want the denominator to be equal to TC , i.e. we want (42.57) to hold, so that

$$y = \underbrace{\frac{CF}{AF + BGz^{-d}}}_{TC} e = \frac{F}{T} e \quad (42.60)$$

which is (42.58). \square

Remark 42.4. Orders n_F and n_G must again be enough for (42.57) to be possible, and can be found by inspection in each case; these orders turn out to be:

$$n_F = n_B + d - 1 \quad (42.61)$$

$$n_G = n_A - 1 \quad (42.62)$$

The maximum number of poles that can be placed is given by

$$n_T \leq n_A + n_B + d - 1 - n_C \quad \square \quad (42.63)$$

Example 42.5. Consider a plant with an output given by

$$\begin{aligned} y(t) &= 2y(t - T_s) + u(t - T_s) + 0.5u(t - 2T_s) + e(t) + 0.3e(t - T_s) \\ \Leftrightarrow y(t) \underbrace{(1 - 2z^{-1})}_{A(z^{-1})} &= u(t) \underbrace{z^{-1}}_{z^{-d}} \underbrace{(1 + 0.5z^{-1})}_{B(z^{-1})} + e(t) \underbrace{(1 + 0.3z^{-1})}_{C(z^{-1})} \end{aligned} \quad (42.64)$$

for which we want a controller placing a pole at $z = 0.5$, i.e.

$$T(z^{-1}) = 1 - 0.5z^{-1} \quad (42.65)$$

We must solve (42.57):

$$(1 - 0.5z^{-1})(1 + 0.3z^{-1}) = (1 - 2z^{-1})F(z^{-1}) + z^{-1}(1 + 0.5z^{-1})G(z^{-1}) \quad (42.66)$$

The left side of the equation is of order 2. The right side will be of order 2 if $n_F = 1$ and $n_G = 0$, i.e. $F(z^{-1}) = f_0 + f_1z^{-1}$ and $G(z^{-1}) = g_0$. Notice that the only independent term on the right side will be f_0 ; thus we must have $f_0 = 1$. This still leaves two variables, g_0 and f_1 , which suffice to make the equation possible.

Instead of reasoning like this we could apply (42.61)–(42.62) and write

$$\begin{aligned} 1 - 0.2z^{-1} - 0.15z^{-2} &= (1 - 2z^{-1})(f_0 + f_1z^{-1}) + (z^{-1} + 0.5z^{-2})g_0 \\ &= f_0 + f_1z^{-1} - 2f_0z^{-1} - 2f_1z^{-2} + g_0z^{-1} + 0.5g_0z^{-2} \end{aligned} \quad (42.67)$$

Equalling the coefficients of the same order on both sides,

$$\begin{cases} 1 = f_0 \\ -0.2 = f_1 - 2f_0 + g_0 \\ -0.15 = -2f_1 + 0.5g_0 \end{cases} \Leftrightarrow \begin{cases} f_0 = 1 \\ f_1 + g_0 = 1.8 \\ -4f_1 + g_0 = -0.3 \end{cases} \Rightarrow \begin{cases} F(z^{-1}) = 1 + 0.42z^{-1} \\ G(z^{-1}) = 1.38 \end{cases} \quad (42.68)$$

Thus

$$u(t) = -y(t) \frac{1.38}{1 + 0.42z^{-1}} \quad (42.69)$$

and this control law will achieve

$$y(t) = -e(t) \frac{1 + 0.42z^{-1}}{1 - 0.5z^{-1}} \quad (42.70)$$

with the desired pole. Notice that (42.69) is stable. \square

Remark 42.5. Unlike (42.2), which has poles that cancel the zeros of the plant, and can therefore be known in advance, the controller given by (42.56) has poles wherever the roots of $F(z^{-1})$ turn out to be. Consequently, its stability should always be verified; an unstable controller would result in a situation similar to that of Example 42.3.

If the controller is unstable, the only solution is changing $T(z^{-1})$. \square

Always check if a pole placement controller is stable

42.3 Control design for time-varying references

When the reference varies with time, controllers are more difficult to design, and let through more noise to the output.

Tracking controller

Theorem 42.4. Consider a plant with an output given by

$$y(t) = \frac{B(z^{-1})}{A(z^{-1})} z^{-d} u(t) + \frac{C(z^{-1})}{A(z^{-1})} e(t) \quad (42.71)$$

where $u(t)$ is a manipulated input, and $e(t)$ is white noise. (If there is no pure delay of d sample times from $u(t)$ to the output, then $d = 0$.) This plant is controlled in closed loop, as seen in Figure 42.1, and no restriction on reference $r(t)$ is assumed. In this situation, the controller given by

$$u(t) = \frac{G(z^{-1})}{B(z^{-1})F(z^{-1})} \underbrace{\frac{1}{1-z^{-1}}}_{\text{integration}} (r(t) - y(t)) \quad (42.72)$$

where $F(z^{-1})$ and $G(z^{-1})$ are found solving

$$C(z^{-1}) = A(z^{-1})F(z^{-1})(1-z^{-1}) + z^{-d}G(z^{-1}) \quad (42.73)$$

achieves an output given by

$$y(t) = F(z^{-1})(1-z^{-1})e(t) + \frac{G(z^{-1})}{C(z^{-1})} z^{-d} r(t) \quad (42.74)$$

Proof. Replacing (42.72) in (42.71), and dropping the dependence on t and z^{-1} to alleviate the notation,

$$\begin{aligned} y &= \frac{B}{A} z^{-d} \frac{G}{BF} \frac{1}{1-z^{-1}} (r - y) + \frac{C}{A} e \\ \Leftrightarrow y \left(\underbrace{1 + \frac{G}{AF} \frac{z^{-d}}{1-z^{-1}}}_{\frac{AF(1-z^{-1}) + Gz^{-d}}{AF(1-z^{-1})}} \right) &= \frac{C}{A} e + \frac{G}{AF} \frac{z^{-d}}{1-z^{-1}} r \\ \Leftrightarrow y &= \frac{CF(1-z^{-1})}{AF(1-z^{-1}) + Gz^{-d}} e + \frac{Gz^{-d}}{AF(1-z^{-1}) + Gz^{-d}} r \end{aligned} \quad (42.75)$$

Replacing (42.73) in the denominator, (42.74) is obtained. \square

Remark 42.6. Once more, orders n_F and n_G must be enough for (42.73) to be possible, can be found by inspection in each case, and turn out to be:

$$n_F = d - 1 \quad (42.76)$$

$$n_G = \max\{n_A, n_C - d\} \quad \square \quad (42.77)$$

Example 42.6. Consider plant (42.64) from Example 42.5:

$$A(z^{-1}) = 1 - 2z^{-1} \quad (42.78)$$

$$B(z^{-1}) = 1 + 0.5z^{-1} \quad (42.79)$$

$$C(z^{-1}) = 1 + 0.3z^{-1} \quad (42.80)$$

$$d = 1 \quad (42.81)$$

To find a tracking controller, we must solve (42.73):

$$1 + 0.3z^{-1} = (1 - 2z^{-1})(1 - z^{-1})F(z^{-1}) + z^{-1}G(z^{-1}) \quad (42.82)$$

The left side of the equation is of order 1. The right side is at least of order 2; this is the case when $n_F = 1$ and $n_G = 2$, i.e. $F(z^{-1}) = f_0$ and $G(z^{-1}) = g_0 + g_1 z^{-1}$. Notice that the only independent term on the right side will be f_0 ; thus we must have $f_0 = 1$. This still leaves two variables, g_0 and g_1 , which suffice to make the equation possible.

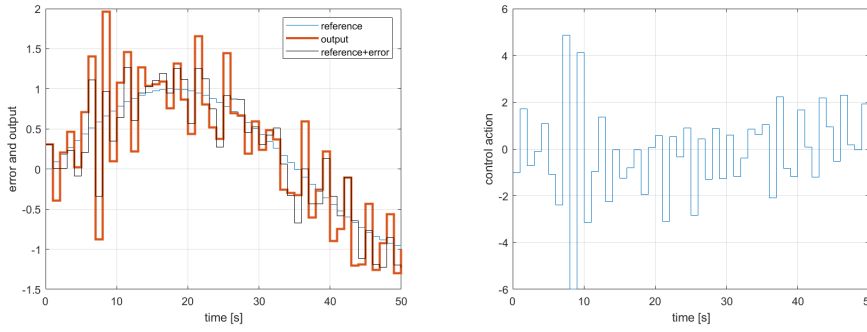


Figure 42.6: Simulation of plant (42.64) controlled by (42.85), from Example 42.6.

Instead of reasoning like this we could apply (42.76)–(42.77) and write

$$\begin{aligned} 1 + 0.3z^{-1} &= (1 - 2z^{-1})(1 - z^{-1})f_0 + z^{-1}(g_0 + g_1z^{-1}) \\ &= f_0 - 3f_0z^{-1} + 2f_0z^{-2} + g_0z^{-1} + g_1z^{-2} \end{aligned} \quad (42.83)$$

Equalling the coefficients of the same order on both sides,

$$\begin{cases} 1 = f_0 \\ 0.3 = -3f_0 + g_0 \\ 0 = 2f_0 + g_1 \end{cases} \Leftrightarrow \begin{cases} f_0 = 1 \\ g_0 = 3.3 \\ g_1 = -2 \end{cases} \Rightarrow \begin{cases} F(z^{-1}) = 1 \\ G(z^{-1}) = 3.3 - 2z^{-1} \end{cases} \quad (42.84)$$

Thus

$$u(t) = -y(t) \frac{3.3 - 2z^{-1}}{(1 + 0.5z^{-1})(1 - z^{-1})} = -\frac{3.3 - 2z^{-1}}{1 - 0.5z^{-1} - 0.5z^{-2}}y(t) \quad (42.85)$$

and this control law will achieve

$$y(t) = (1 - z^{-1})e(t) + \frac{3.3 - 2z^{-1}}{1 + 0.3z^{-1}}z^{-1}r(t) \quad (42.86)$$

Figure 42.6 shows the results of a simulation with 50 time steps and a sinusoidal reference. Notice that the output is not much far from the reference corrupted by the noise (which as usual cannot be eliminated). \square

42.4 Adaptive control

In the adaptive control of a discrete time plant, in each time instant:

- its model is updated using a recursive identification algorithm, as we saw in Section 31.5;
- the updated model is then used to update its controller, using one of the methods from this Chapter;
- the controller provides a control action for the next time instant.

This corresponds to what is represented in Figure 42.7.

Adaptive control may have stability problems caused by the identification algorithm, by the controller design algorithm, or even by any abrupt change of the control action resulting from an abrupt change in the model or in the controller. We will not further address these problems or their possible solutions.

Glossary

I shook my head: I could not see how poor people had the means of being kind; and then to learn to speak like them, to adopt their manners, to be uneducated, to grow up like one of the poor women I saw sometimes nursing their children or washing their clothes at the cottage doors of the village of Gateshead: no, I was not heroic enough to purchase liberty at the price of caste.

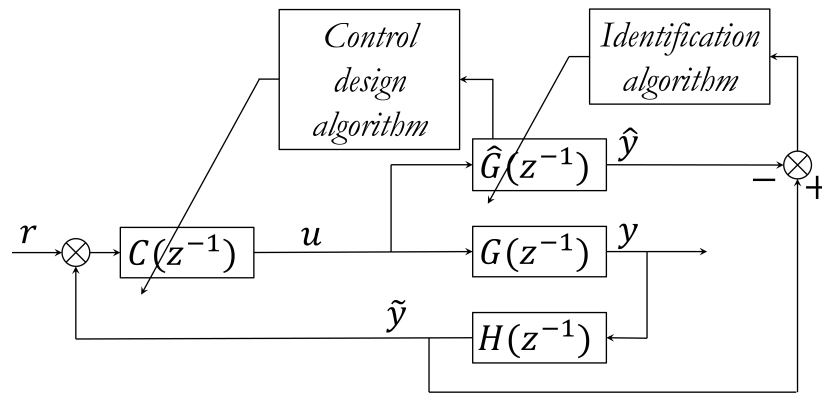


Figure 42.7: Adaptive control. $G(z^{-1})$ is the plant, $\hat{G}(z^{-1})$ is the model of the plant, $H(z^{-1})$ is the sensor, and $C(z^{-1})$ is the controller. $r(t)$ is the reference, $u(t)$ is the control action, $y(t)$ is the output, $\tilde{y}(t)$ is the measured output, and $\hat{y}(t)$ is the estimated output.

Charlotte BRONTË (1816 — †1855), *Jane Eyre: An Autobiography* (1847), III

adaptive control controle adaptativo

minimum variance control controle de variância mínima

generalised (generalized, US) minimum variance control controle de variância mínima generalizado

sub-optimal minimum variance control controle sub-ótimo de variância mínima

pole assignment control controle por colocação de polos

Exercises

1. Consider a plant with sampling time T_s and an output given by

$$y(t) = 0.6y(t-T_s) + 0.1y(t-2T_s) + u(t-T_s) + 2u(t-2T_s) + e(t) - 0.5e(t-T_s) \quad (42.87)$$

where $e(t)$ is white noise and $u(t)$ is a manipulated input. Find the transfer function $R(z^{-1})$ of a minimum variance regulator for this plant. *Hint:* notice that there is a non-minimum phase zero.

2. Find a controller for plant (42.15) of Example 42.1 that places the poles of the closed loop at $z = \frac{1}{10}$ and $z = \frac{1}{20}$. *Hint:* use MATLAB for the calculations.

Epilogue

Chapter 43

What next?

Beloved pupil! Tamed by thee,
Addish-, Subtrac-, Multiplica-tion,
Division, Fractions, Rule of Three,
Attest thy deft manipulation!

Caro discípulo, que bem dominas
somas, produtos e subtrações,
regras de três, razões e quocientes,
fazendo dexas manipulações:

Then onward! Let the voice of Fame
From Age to Age repeat thy story,
Till thou hast won thyself a name
Exceeding even Euclid's glory!

avança, pois, para que a voz da Fama
de era em era ecoe a tua história,
até que alcances para ti um nome
que exceda até de Euclides mesmo a glória!

Lewis CARROLL (1832 — †1898), *A tangled tale* (1885), To my pupil (transl. Duarte Valério, 2002)

This chapter not only brings together some odd ends left behind in previous chapters, as it tries to show that what you have learned is related to many subjects outside the scope of these Lecture Notes, which you may with this background easily learn in dedicated courses.

This chapter is still being written. In the picture: National Pantheon, or Church of Saint Engratia, Lisbon (source: <http://www.panteaonacional.gov.pt/171-2/historia-2/>).

43.1 Discrete events and automation

This section concerns a particular type of control systems, called automation systems, involving binary variables.

43.2 State-space representations of systems continuous in time

Linear systems, both continuous in time and discrete, can be put in a matrix form that is convenient for several techniques of modelling, control, and identification. Let us first see how this can be done for continuous time.

43.3 State-space representations of systems discrete in time

Systems which are discrete in time can also be put in a state-space representation.

43.4 MIMO systems and MIMO control

While we have only been concerned with SISO systems, MIMO systems are in fact ubiquitous. They can be represented in matrix form, either using transfer functions or state-space representations.

43.5 Optimal control

In this section we have a very brief introduction to techniques of controller design called optimal control.

43.6 Fractional order control

Fractional order derivatives and systems, which were dealt with in Part VII, can be used for controller design in a variety of ways. In short, the main ones are:

- Fractional PID control
- First-generation CRONE control
- Second-generation CRONE control
- Third-generation CRONE control
- Non-linear fractional control
- Variable order fractional control

43.7 Other areas

In this course we have studied many tools of modelling, control, and identification that can be applied to systems and signals that have little or nothing to do with mechatronics. They can be used in many areas as different as biology, economy, or military sciences, once you grasp how.



This chapter is still being written. In the picture: National Pantheon, or Church of Saint Engratia, Lisbon (source: <http://www.panteaonacional.gov.pt/171-2/historia-2/>).

Glossary

Had We revealed it as a non-Arabic Quran, they would have certainly argued, “If only its verses were made clear [in our language]. What! A non-Arabic revelation for an Arab audience!”

MUHAMMAD ibn Abdullah (570? — †632), *Quran* (610–632), xli 44, Mustafa Khattab version (2015)

discrete event evento discreto
demand procura (demanda, bras.)
optimal control controlo ótimo
predator-prey predador-presa
supply oferta