INTRODUCTION TO KALMAN FILTERING

• What is a Kalman Filter?
  – Introduction to the Concept
  – A simple example
  – Which is the best estimate?
  – Basic Assumptions

• Discrete Kalman Filter
  – Problem Formulation
  – From the Assumptions to the Problem Solution
  – Towards the Solution
  – Filter dynamics
    • Prediction cycle
    • Filtering cycle
    • Summary
  – Properties of the Discrete KF
  – A simple example

• The meaning of the error covariance matrix

• The Extended Kalman Filter
WHAT IS A KALMAN FILTER?

• Optimal Recursive Data Processing Algorithm

• Typical Kalman filter application
WHAT IS A KALMAN FILTER?
Introduction to the Concept

- **Optimal** Recursive Data Processing Algorithm
  - Dependent upon the criteria chosen to evaluate performance
  - Under certain assumptions, KF is optimal with respect to virtually any criteria that makes sense.
  - KF incorporates all available information
    - knowledge of the system and measurement device dynamics
    - statistical description of the system noises, measurement errors, and uncertainty in the dynamics models
    - any available information about initial conditions of the variables of interest
A simple (static) example (from Maybeck)

- You are an inexperient sailor at sea and you do not know your location
- Take a (one-dimensional) star sighting to establish your position
- At time $t_1$ you establish your position to be $z_1$ (you measure)
- Because there is inherent measurement device inaccuracies you say that the precision is $\sigma_{z_1}$ (standard deviation)
- You can establish the conditional probability of $x(t_1)$ (your position at time $t_1$) conditioned on the observed value of measurement $z_1$.

Based on this conditional pdf the best estimate of your position is

$$\hat{x}(t_1) = z_1$$

and the variance of the error in the estimate is

$$\sigma^2_x(t_1) = \sigma^2_{z_1}$$
A simple (static) example (from Maybeck)

- You have a friend that is a trained sailor
- At time instant $t_2$ ($t_2 \approx t_1$) (the boat did not move) this trained sailor measures $z_2$ with a variance $\sigma_{z_2}^2$
- As this second sailor has larger skills, assume that the variance in his measurement is smaller than the measurement in yours.

Based on this conditional pdf the best estimate of the position given by the trained sailor $\hat{x}(t_2) = z_2$

and the variance of the error in the estimate is

$$\sigma_x^2(t_2) = \sigma_{z_2}^2$$
A simple (static) example (from Maybeck)

- At this point we have two measurements, with different uncertainty
- How to combine these data?
- What do we want to know?
  - Conditional position at time $t_2$ ($t_2 = t$) given both $z_1$ and $z_2$

\[
\mathbf{f}_{X(t_2)|Z(t_1),Z(t_2)}(X \mid Z_1, Z_2)
\]

is Gaussian

\[
\mu = \frac{\sigma_{z_2}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2} z_1 + \frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2} z_2
\]

\[
\frac{1}{\sigma^2} = \frac{1}{\sigma_{z_1}^2} + \frac{1}{\sigma_{z_2}^2}
\]

The uncertainty in the position estimate decreased by combining the two pieces of information.
A simple (static) example (from Maybeck)

• Which is the best estimate at time $t_2$ given $z_1$ and $z_2$?
  - The mean (also the maximum) of the conditional pdf

$$\hat{x}(t_2) = \frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2} z_1 + \frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2} z_2$$

$$\hat{x}(t_2) = z_1 + \frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2} (z_2 - z_1)$$

In the previous time instant
$$\hat{x}(t_1) = z_1$$

$$\hat{x}(t_2) = \hat{x}(t_1) + \frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2} (z_2 - \hat{x}(t_1))$$

This is the same as the Kalman filter implementation

The best prediction at time $t_2$, is given by the previous estimate plus an error multiplied by a gain

Interpret the error
A simple (static) example (from Maybeck)

• Which is the best estimate at time $t_2$ given $z_1$ and $z_2$?
  – The mean (also the maximum) of the conditional pdf

$$\hat{x}(t_2) = \hat{x}(t_1) + K(t_2)(z_2 - \hat{x}(t_1))$$

• Which is the uncertainty of the estimate at time $t_2$ given $z_1$ and $z_2$?
  – The variance of the conditional pdf

$$\frac{1}{\sigma^2} = \frac{1}{\sigma_{z_1}^2} + \frac{1}{\sigma_{z_2}^2}$$

$$K(t_2) = \frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2}$$

$$\sigma_x^2(t_2) = \sigma_x^2(t_1) - K(t_2)\sigma_x^2(t_1)$$
A simple (dynamic) example (from Maybeck)

- From now on observations are done at different time instants and the boat travels between the time instants where measurements are taken

\[
\frac{dx}{dt} = u + w
\]

Motion model
- \( u \) = velocity
- \( w \) = noise term representing uncertainty in the actual knowledge of velocity, assumed zero mean

- At time \( t_2 \) we already have \( \hat{x}(t_2) \) and \( \sigma_x^2(t_2) \)

- Now the vehicle is travelling

- **Before** doing a measurement at time instant \( t_3 \) (i.e., at \( t_3^- \)) which is the best prediction that we can do about the position and associated uncertainty?

\[
\hat{x}(t_3^-) = \hat{x}(t_2) + u(t_3 - t_2)
\]

\[
\sigma_x^2(t_3^-) = \sigma_x^2(t_2) + \sigma_w^2(t_3 - t_2)
\]
A simple (dynamic) example (from Maybeck)

• A measurement $z_3$ is done at time instant $t_3$ with an assumed variance $\sigma_{z_3}^2$

• Which is now the best estimate $\hat{x}(t_3)$?

• Once again we have two Gaussian probability density functions
  • one associated with information up to $t_3^-$
  • one provided by the measurement itself

• Combining both

\[
\begin{align*}
\hat{x}(t_3) &= \hat{x}(t_3^-) + K(t_3)(z_3 - \hat{x}(t_3^-)) \\
\sigma_x^2(t_3) &= \sigma_x^2(t_3^-) - K(t_3)^2 \sigma_{z_3}^2 \\
K(t_3) &= \frac{\sigma_x^2(t_3^-)}{\sigma_x^2(t_3^-) + \sigma_{z_3}^2}
\end{align*}
\]

Kalman Gain
WHAT IS A KALMAN FILTER?  
Introduction to the concept

- **Optimal** Recursive Data Processing Algorithm

\[
x(k + 1) = f(x(k), u(k), w(k)) \\
z(k + 1) = h(x(k + 1), v(k + 1))
\]

- **Given**
  - \( f, h, \) noise characterization, initial conditions
  - \( z(0), z(1), z(2), \ldots, z(k) \)

- **Obtain**
  - the “best” estimate of \( x(k) \)

- \( x \) - state
- \( f \) - system dynamics
- \( h \) - measurement function
- \( u \) - controls
- \( w \) - system error sources
- \( v \) - measurement error sources
- \( z \) - observed measurements
WHAT IS A KALMAN FILTER?
Introduction to the concept

- Optimal **Recursive** Data Processing Algorithm
  - the KF does not require all previous data to be kept in storage and reprocessed every time a new measurement is taken.

\[
\begin{align*}
z(0) &\quad z(1) &\quad z(2) &\quad \ldots &\quad z(k) \\
\downarrow & & & & \\
\hat{x}(k) & & & &
\end{align*}
\]

\[
\begin{align*}
z(0) &\quad z(1) &\quad z(2) &\quad \ldots &\quad z(k) &\quad z(k+1) \\
\downarrow & & & & & \\
\hat{x}(k) &\quad \hat{x}(k+1) & & &
\end{align*}
\]

To evaluate $\hat{x}(k+1)$ the KF only requires $\hat{x}(k)$ and $z(k+1)$
WHAT IS A KALMAN FILTER?
Introduction to the concept

- Optimal Recursive **Data Processing Algorithm**
  - The KF is a data processing algorithm
  - The KF is a computer program running in a central processor
WHAT IS THE KALMAN FILTER?
Which is the best estimate?

- Any type of filter tries to obtain an **optimal** estimate of desired quantities from data provided by a noisy environment.

- **Best** = minimizing errors in some respect.

- Bayesian viewpoint - the filter propagates the **conditional probability density** of the desired quantities, conditioned on the knowledge of the actual data coming from measuring devices

- Why base the state estimation on the conditional probability density function?
WHAT IS A KALMAN FILTER?
Which is the best estimate?

Example

• x(i) one dimensional position of a vehicle at time instant i
• z(j) two dimensional vector describing the measurements of position at time j by two separate radars

• If z(1)=z₁, z(2)=z₂, …., z(j)=z_j

\[ p(x(i)|z(1),z(2),...,z(i)) \]

– represents all the information we have on x(i) based (conditioned) on the measurements acquired up to time i
– given the value of all measurements taken up time i, this conditional pdf indicates what the probability would be of x(i) assuming any particular value or range of values.
WHAT IS A KALMAN FILTER? Which is the best estimate?

• The shape of $p_{x(i)|z(1),z(2),...,z(i)}(x | z_1,z_2,....,z_i)$ conveys the amount of certainty we have in the knowledge of the value $x$.

• Based on this conditional pdf, the estimate can be:
  – the **mean** - the center of probability mass (MMSE)
  – the **mode** - the value of $x$ that has the highest probability (MAP)
  – the **median** - the value of $x$ such that half the probability weight lies to the left and half to the right of it.
WHAT IS THE KALMAN FILTER?

Basic Assumptions

• The Kalman Filter performs the conditional probability density propagation
  – for systems that can be described through a LINEAR model
  – in which system and measurement noises are WHITE and GAUSSIAN

• Under these assumptions,
  – the conditional pdf is Gaussian
  – mean=mode=median
  – there is a unique best estimate of the state
  – the KF is the best filter among all the possible filter types

• What happens if these assumptions are relaxed?
• Is the KF still an optimal filter? In which class of filters?
**DISCRETE KALMAN FILTER**

**Problem Formulation**

**MOTIVATION**

- Given a discrete-time, linear, time-varying plant
  - with random initial state
  - driven by white plant noise
- Given noisy measurements of linear combinations of the plant state variables
- Determine the best estimate of the system state variable

**STATE DYNAMICS AND MEASUREMENT EQUATION**

\[
x_{k+1} = A_k x_k + B_k u_k + G_k w_k, \quad k \geq 0
\]

\[
z_k = C_k x_k + v_k
\]
DISCRETE KALMAN FILTER

Problem Formulation

VARIABLE DEFINITIONS

\[ x_k \in \mathbb{R}^n \] state vector (stochastic non-white process)

\[ u_k \in \mathbb{R}^m \] deterministic input sequence

\[ w_k \in \mathbb{R}^n \] white Gaussian system noise
(assumed with zero mean)

\[ v_k \in \mathbb{R}^r \] white Gaussian measurement noise
(assumed with zero mean)

\[ z_k \in \mathbb{R}^r \] measurement vector (stochastic non-white sequence)
DISCRETE KALMAN FILTER
Problem Formulation

INITIAL CONDITIONS

• $x_0$ is a Gaussian random vector, with
  
  mean \quad E[x_0] = \bar{x}_0 

  covariance matrix \quad E[(x_0 - \bar{x}_0)(x_0 - \bar{x}_0)^T] = P_0 = P_0^T \geq 0

STATE AND MEASUREMENT NOISE

• zero mean \quad E[w_k]=E[v_k]=0

• \{w_k\}, \{v_k\} - white Gaussian sequences

$$E \begin{pmatrix} w_k \\ v_k \end{pmatrix} \begin{pmatrix} w_k^T \\ v_k^T \end{pmatrix} = \begin{bmatrix} Q_k & 0 \\ 0 & R_k \end{bmatrix}$$

$x(0)$, $w_k$ and $v_j$ are independent for all $k$ and $j$
DEFINITION OF FILTERING PROBLEM

• Let k denote present value of time

• Given the sequence of past inputs

\[ U_{0:k-1}^k = U_{0:k-1} = \{u_0, u_1, \ldots, u_{k-1}\} \]

• Given the sequence of past measurements

\[ Z_{1:k}^k = Z_{1:k} = \{z_1, z_2, \ldots, z_k\} \]

• Evaluate the best estimate of the state x(k)
DISCRETE KALMAN FILTER
Problem Formulation

- **Given** $x_0$
  - “Nature” apply $w_0$
  - We apply $u_0$
  - The system moves to state $x_1$
  - We make a measurement $z_1$

**Question:** which is the best estimate of $x_1$?

**Answer:** obtained from $p(x_1 \mid Z_1^1, U_0^0)$

- “Nature” apply $w_1$
- We apply $u_1$
- The system moves to state $x_2$
- We make a measurement $z_2$

**Question:** which is the best estimate of $x_2$?

**Answer:** obtained from $p(x_2 \mid Z_2^1, U_0^1)$
**DISCRETE KALMAN FILTER**

**Problem Formulation**

Question: which is the best estimate of $x_{k-1}$?

Answer: obtained from $p(x_{k-1} \mid Z_1^{k-1}, U_{0}^{k-2})$

- “Nature” apply $w_{k-1}$
- We apply $u_{k-1}$
- The system moves to state $x_k$
- We make a measurement $z_k$

Question: which is the best estimate of $x_k$?

Answer: obtained from $p(x_k \mid Z_1^k, U_{0}^{k-1})$
DISCRETE KALMAN FILTER
Towards the Solution

- The filter has to propagate the conditional probability density functions

\[ p(x_0) \]

\[ p(x_1 \mid Z_1^1, U_0^0) \rightarrow \hat{x}(1 \mid 1) \]

\[ p(x_2 \mid Z_1^2, U_0^1) \rightarrow \hat{x}(2 \mid 2) \]

\[ \vdots \]

\[ p(x_{k-1} \mid Z_1^{k-1}, U_0^{k-2}) \rightarrow \hat{x}(k-1 \mid k-1) \]

\[ \vdots \]

\[ p(x_k \mid Z_1^k, U_0^{k-1}) \rightarrow \hat{x}(k \mid k) \]

\[ \vdots \]
DISCRETE KALMAN FILTER
From the Assumptions to the Problem Solution

- **The LINEARITY** of
  - the system state equation
  - the system observation equation
- **The GAUSSIAN** nature of
  - the initial state, \( x_0 \)
  - the system **white** noise, \( w_k \)
  - the measurement **white** noise, \( v_k \)

Uniquely characterized by

- the conditional mean
  \[
  \hat{x}(k \mid k) = E[x_k \mid Z_1^k, U_0^{k-1}]
  \]
- the conditional covariance
  \[
P(k \mid k) = \text{cov}[x_k, x_k \mid Z_1^k, U_0^{k-1}]
  \]

\[
p(x_k \mid Z_1^k, U_0^{k-1}) \sim N(\hat{x}(k \mid k), P(k \mid k))
\]
DISCRETE KALMAN FILTER
Towards the Solution

- As the conditional probability density functions are Gaussian, the Kalman filter only propagates the first two moments

\[
p(x_0)
\]

\[
p(x_1 | Z_1^1, U_0^0)
\]

\[
p(x_2 | Z_1^2, U_0^1)
\]

\[\vdots\]

\[
p(x_{k-1} | Z_1^{k-1}, U_0^{k-2})
\]

\[
p(x_k | Z_1^k, U_0^{k-1})
\]

\[
p(x_{k+1} | Z_1^{k+1}, U_0^{k-1})
\]

\[\vdots\]

- \(E[x_1 | Z_1^1, U_0^0] = \hat{x}(1 | 1)\) \(P(1 | 1)\)

- \(E[x_2 | Z_1^2, U_0^1] = \hat{x}(2 | 2)\) \(P(2 | 2)\)

- \(E[x_{k-1} | Z_1^{k-1}, U_0^{k-2}] = \hat{x}(k-1 | k-1)P(k-1 | k-1)\)

- \(E[x_k | Z_1^k, U_0^{k-1}] = \hat{x}(k | k)\) \(P(k | k)\)
We stated that the state estimate equals the conditional mean

\[ \hat{x}(k \mid k) = \mathbb{E}[x_k \mid Z^k_1, U^{k-1}_0] \]

- Why?
- Why not the mode of \( p(x_k \mid Z^k_1, U^{k-1}_0) \)?
- Why not the median of \( p(x_k \mid Z^k_1, U^{k-1}_0) \)?

As \( p(x_k \mid Z^k_1, U^{k-1}_0) \) is Gaussian
- mean = mode = median
DISCRETE KALMAN FILTER
Filter dynamics

- KF dynamics is recursive

\[ Z^k_1 = \{z_1, z_2, \ldots, z_k\} \]
\[ U^k_0 = \{u_0, u_1, \ldots, u_{k-1}\} \]

**Prediction cycle**

What can you say about \( x_{k+1} \) before we make the measurement \( z_{k+1} \)

\[ p(x_{k+1} | Z^k_1, U^k_0) \]

**Filtering cycle**

How can we improve our information on \( x_{k+1} \) after we make the measurement \( z_{k+1} \)

\[ z^k_{1+1} = \{z^k_1, z_{k+1}\} \]
\[ U^k_0 = \{U^k_{0-1}, u_k\} \]
**DISCRETE KALMAN FILTER**

Filter dynamics

\[ p(x_0) \]
\[ p(x_1 \mid U_{0}^{0}) \]
\[ p(x_2 \mid Z_{1}^{1}, U_{0}^{1}) \]
\[ p(x_{k+1} \mid Z_{1}^{k+1}, U_{0}^{k}) \]
\[ p(x_{k+2} \mid Z_{1}^{k+2}, U_{0}^{k+1}) \]
DISCRETE KALMAN FILTER

Filter dynamics - Prediction cycle

- Prediction cycle

\[ p(x_k \mid Z^k_1, U^{k-1}_0) \sim N(\hat{x}(k \mid k), P(k \mid k)) \]

\[ p(x_{k+1} \mid Z^k_1, U^k_0) \]

assumed known

- Is Gaussian

\[ \hat{x}(k+1 \mid k) = E(x_{k+1} \mid Z^k_1, U^k_0) \]

\[ P(k+1 \mid k) = \text{cov}[x_{k+1}; x_{k+1} \mid Z^k_1, U^k_0] \]

\[ x_{k+1} = A_k x_k + B_k u_k + G_k w_k \]

\[ E[x_{k+1} \mid Z^k_1, U^k_0] = A_k E[x_k \mid Z^k_1, U^k_0] + B_k E[u_k \mid Z^k_1, U^k_0] + G_k E[w_k \mid Z^k_1, U^k_0] \]

\[ \hat{x}(k+1 \mid k) = A_k \hat{x}(k \mid k) + B_k u_k \]
DISCRETE KALMAN FILTER
Filter dynamics - Prediction cycle

- Prediction cycle

\[ \tilde{x}(k+1 | k) = x_{k+1} - \hat{x}(k+1 | k) \]

\[ x(k+1) - \hat{x}(k+1 | k) = A_k x_k + B_k u_k + G_k w_k - (A_k \hat{x}(k | k) + B_k u_k) \]

\[ \tilde{x}(k+1 | k) = A_k \tilde{x}(k | k) + G_k w_k \]

\[ P(k+1 | k) = \text{cov}[x_{k+1}; x_{k+1} | Z_k^1, U_0^k] \]

\[ P(k+1 | k) = A_k P(k | k) A_k^T + G_k Q_k G_k^T \]

\[ \text{cov}[y; y] = E[(y - \bar{y})(y - \bar{y})^T] \]
DISCRETE KALMAN FILTER
Filter dynamics - Filtering cycle

- Filtering cycle

\[ p(x_{k+1} \mid Z_1^k, U_0^k) \]
\[ N(\hat{x}(k+1 \mid k), P(k+1 \mid k)) \]

\[ + \quad z_{k+1} \quad \rightarrow \quad p(x_{k+1} \mid Z_1^{k+1}, U_0^k) \]

**1st Step** | **Measurement prediction**

\[ p(z_{k+1} \mid Z_1^k, U_0^k) \]

\[ p(C_{k+1}x_{k+1} + v_{k+1} \mid Z_1^k, U_0^k) \]

\[ E[z_{k+1} \mid Z_1^k, U_0^k] = \hat{z}(k+1 \mid k) = C_{k+1}\hat{x}(k+1 \mid k) \]

\[ \text{cov}[z_{k+1}; z_{k+1} \mid Z_1^k, U_0^k] = P_z(k+1 \mid k) = C_{k+1}P(k+1 \mid k)C_{k+1}^T + R_{k+1} \]

What can you say about \( z_{k+1} \) before we make the measurement \( z_{k+1} \)?

DISCRETE KALMAN FILTER
Filter dynamics - Filtering cycle

- Filtering cycle

2nd Step

\[ p(x_{k+1} \mid Z_1^{k+1}, U_0^k) \]

\[
E[x_{k+1} \mid Z_1^{k+1}, U_0^k] = E[x_{k+1} \mid Z_1^k, z_{k+1}, U_0^k]
\]

\[ Z_1^{k+1} \text{ and } \{Z_1^k, \tilde{z}(k+1 \mid k)\} \quad \text{Equivalent from the point of view of the contained information} \]

\[
E[x_{k+1} \mid Z_1^{k+1}, U_0^k] = E[x_{k+1} \mid Z_1^k, \tilde{z}(k+1 \mid k), U_0^k]
\]

Requires result on \( E[x \mid y, z] \)

when \( x, y \) and \( z \) are jointly Gaussian and \( y \) and \( z \) are statistically independent
**DISCRETE KALMAN FILTER**

Filter dynamics - Filtering cycle

\[
\hat{x}(k+1 | k) = \hat{x}(k+1 | k) + P(k+1 | k)C_{k+1}^T \left[ C_{k+1}P(k+1 | k)C_{k+1}^T + R_{k+1} \right]^{-1} (z_{k+1} - C_{k+1}\hat{x}(k+1 | k))
\]

\[
\hat{x}(k+1 | k+1) = \hat{x}(k+1 | k) + K(k+1)(z_{k+1} - C_{k+1}\hat{x}(k+1 | k))
\]

\[
P(k+1 | k+1) = P(k+1 | k) - P(k+1 | k)C_{k+1}^T \left[ C_{k+1}P(k+1 | k)C_{k+1}^T + R_{k+1} \right]^{-1} C_{k+1}P(k+1 | k)
\]

- **Filtering cycle**
DISCRETE KALMAN FILTER
Dynamics

- **Linear System**
  \[ x_{k+1} = A_k x_k + B_k u_k + G_k w_k, \quad k \geq 0 \]
  \[ z_k = C_k x_k + v_k \]

- **Discrete Kalman Filter**

  **prediction**
  \[ \dot{\hat{x}}(k + 1 \mid k) = A_k \hat{x}(k \mid k) + B_k u_k \]
  \[ P(k + 1 \mid k) = A_k P(k \mid k) A_k^T + G_k Q_k G_k^T \]

  **filtering**
  \[ \hat{x}(k + 1 \mid k + 1) = \hat{x}(k + 1 \mid k) + K(k + 1)(z_{k+1} - C_{k+1} \hat{x}(k + 1 \mid k)) \]
  \[ P(k + 1 \mid k + 1) = P(k + 1 \mid k) - K(k + 1)C_{k+1}P(k + 1 \mid k) \]
  \[ K(k + 1) = P(k + 1 \mid k)C_{k+1}^T \left[ C_{k+1}P(k + 1 \mid k)C_{k+1}^T + R_{k+1} \right]^{-1} \]

  **Initial conditions**
  \[ \hat{x}(0 \mid 0) = \bar{x}_0 \quad P(0 \mid 0) = P_0 \]
DISCRETE KALMAN FILTER
Properties

• The Discrete KF is a time-varying linear system

\[ \hat{x}_{k+1|k+1} = (I - K_{k+1}C_{k+1})A_k \hat{x}_{k|k} + K_{k+1}z_{k+1} + B_k u_k \]

– even when the system is time-invariant and has stationary noise

\[ \hat{x}_{k+1|k+1} = (I - K_{k+1}C)A \hat{x}_{k|k} + K_{k+1}z_{k+1} + B u_k \]

• the Kalman gain is not constant

• Does the Kalman gain matrix converges to a constant matrix? In which conditions?
DISCRETE KALMAN FILTER
Properties

- The state estimate is a linear function of the measurements

\[ \hat{x}_{k+1|k+1} = (I - K_{k+1}C_{k+1})A_k \hat{x}_{k|k} + K_{k+1}z_{k+1} + B_ku_k \]

\[ \Phi_k \]

Assuming null inputs for the sake of simplicity

\[ \hat{x}_{0|0} = \bar{x}_0 \]
\[ \hat{x}_{1|1} = \Phi_0 \hat{x}_{0|0} + K_1 z_1 \]
\[ \hat{x}_{2|2} = \Phi_1 \Phi_0 \hat{x}_{0|0} + \Phi_1 K_1 z_1 + K_2 z_2 \]
\[ \hat{x}_{3|3} = \Phi_2 \Phi_1 \Phi_0 \hat{x}_{0|0} + \Phi_2 \Phi_1 K_1 z_1 + \Phi_2 K_2 z_2 + K_3 z_3 \]
**DISCRETE KALMAN FILTER**

**Properties**

- **Innovation process**
  \[ r_{k+1} = z_{k+1} - C_{k+1} \hat{x}(k+1 \mid k) \]
  \[ \hat{x}(k+1 \mid k) = E(x_{k+1} \mid Z_1^k, U_0^k) \]
  - \( z(k+1) \) carries information on \( x(k+1) \) that was not available on \( Z_1^k \)
  - this new information is represented by \( r(k+1) \) - innovation process

- **Properties of the innovation process**
  - the innovations \( r(k) \) are orthogonal to \( z(i) \)
    \[ E[r(k)z^T(i)] = 0, \quad i = 1,2,\ldots,k-1 \]
  - the innovations are uncorrelated/white noise
    \[ E[r(k)r^T(i)] = 0, \quad i \neq k \]
  - this test can be used to access if the filter is operating correctly
DISCRETE KALMAN FILTER

Properties

- Covariance matrix of the innovation process

\[ S(k + 1) = C_{k+1} P(k + 1 | K) C_{k+1}^T + R_{k+1} \]

\[ K(k + 1) = P(k + 1 | k) C_{k+1}^T \left[ C_{k+1} P(k + 1 | k) C_{k+1}^T + R_{k+1} \right]^{-1} \]

\[ K(k + 1) = P(k + 1 | k) C_{k+1}^T S_{k+1}^{-1} \]
The Discrete KF provides an unbiased estimate of the state

- $\hat{x}_{k+1|k+1}$ is an unbiased estimate of the state $x(k+1)$, providing that the initial conditions are $\hat{x}(0 \mid 0) = \bar{x}_0$ and $P(0 \mid 0) = P_0$

- Is this still true if the filter initial conditions are not the specified?
DISCRETE KALMAN FILTER
Steady state Kalman Filter

- Time invariant system; stationary white system and observation noise
  \[ x_{k+1} = Ax_k + Gw_k, \quad k \geq 0 \]
  \[ z_k = Cx_k + v_k \]
  \[ E[w_k w_k^T] = Q \]
  \[ E[v_k v_k^T] = R \]

- Filter dynamics
  \[ \hat{x}(k + 1 \mid k + 1) = A \hat{x}(k + 1 \mid k) + K(k + 1)(z_{k+1} - C \hat{x}(k + 1 \mid k)) \]

  Discrete Riccati Equation

  \[ P(k + 1 \mid k) = AP(k \mid k - 1)A^T - AP(k \mid k - 1)C^T [CP(k \mid k - 1)C^T + R]^{-1} CP(k \mid k - 1)A^T + GQG^T \]
If $Q$ is positive definite, $(A, G\sqrt{Q})$ is controllable, and $(A, C)$ is observable, then

- the steady state Kalman filter exists
- the limit exists $\lim_{k \to \infty} P(k+1 | k) = P_\infty^-$
- $P_\infty^-$ is the unique, finite positive-semidefinite solution to the algebraic equation
  \[
P_\infty^- = AP_\infty^- A^T - AP_\infty^- C^T [CP_\infty^- C^T + R]^{-1} CP_\infty^- A^T + GQG^T
  \]
- $P_\infty^-$ is independent of $P_0$ provided that $P_0 \geq 0$
- the steady-state Kalman filter is asymptotically unbiased

$K_\infty = P_\infty^- C^T [CP_\infty^- C^T + R]^{-1}$
MEANING OF THE COVARIANCE MATRIX

Generals on Gaussian pdf

- Let $z$ be a Gaussian random vector of dimension $n$

  $$ E[z] = m, \quad E[(z - m)(z - m)^T] = P $$

- $P$ - covariance matrix - symmetric, positive definite

- Probability density function

  $$ p(z) = \frac{1}{\sqrt{(2\pi)^n \det P}} \exp\left[ -\frac{1}{2} (z - m)^T P^{-1} (z - m) \right] $$

$n=1$

$n=2$
MEANING OF THE COVARIANCE MATRIX

Generals on Gaussian pdf

- Locus of points where the pdf is greater or equal than a given threshold
  \[(z - m)^T P^{-1} (z - m) \leq K\]

  - \(n=1\) line segment
  - \(n=2\) ellipse and inner points
  - \(n=3\) 3D ellipsoid and inner points
  - \(n>3\) hiperellipsoid and inner points

- If \(P = \text{diag} \ (\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)\)
  - the ellipsoid axis are aligned with the axis of the referencial where the vector \(z\) is defined
    \[(z - m)^T P^{-1} (z - m) \leq K \iff \sum_{i=1}^{n} \frac{(z_i - m_i)^2}{\sigma_i^2} \leq 1\]
  - length of the ellipse semi-axis = \(\sigma_i \sqrt{K}\)
MEANING OF THE COVARIANCE MATRIX
Generals on Gaussian pdf - Error elipsoid

Example
n=2

\[ P = \begin{bmatrix}
\sigma_1^2 & 0 \\
0 & \sigma_2^2 \\
\end{bmatrix} \]

\[ P = \begin{bmatrix}
\sigma_1^2 & \sigma_{12} \\
\sigma_{12} & \sigma_2^2 \\
\end{bmatrix} \]
MEANING OF THE COVARIANCE MATRIX
Generals on Gaussian pdf - Error ellipsoid and axis orientation

- Error ellipsoid \((z - m_z)\mathbf{P}^{-1}(z - m_z) \leq K\)
- \(\mathbf{P} = \mathbf{P}^T\) - to distinct eigenvalues correspond orthogonal eigenvectors
- Assuming that \(\mathbf{P}\) is diagonalizable
  
  \[
  \mathbf{P} = \mathbf{T}\mathbf{D}\mathbf{T}^{-1} \quad \text{with} \quad \mathbf{D} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)
  \]
  \[
  \mathbf{T}\mathbf{T}^T = \mathbf{I}
  \]
- Error ellipsoid (after coordinate transformation)
  
  \[
  \mathbf{w} = \mathbf{T}^T\mathbf{z} \quad \mathbf{(z - m_z)}^T\mathbf{D}^{-1}\mathbf{T}^T(z - m_z) \leq K
  \]
  \[
  \mathbf{(w - m_w)}^T\mathbf{D}^{-1}(\mathbf{w} - \mathbf{m}_w) \leq K
  \]
- At the new coordinate system, the ellipsoid axis are aligned with the axis of the new referencial
MEANING OF THE COVARIANCE MATRIX

Generals on Gaussian pdf - Error ellipse and referencial axis

\[ z = \begin{bmatrix} x \\ y \end{bmatrix} \quad \begin{bmatrix} x - m_x \\ y - m_y \end{bmatrix} \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix}^{-1} \begin{bmatrix} x - m_x \\ y - m_y \end{bmatrix} \leq K \]

\[ \frac{(x - m_x)^2}{K\sigma_x^2} + \frac{(y - m_y)^2}{K\sigma_y^2} \leq 1 \]
MEANING OF THE COVARIANCE MATRIX

Generals on Gaussian pdf - Error ellipse and referencial axis

• \( n=2 \)

\[
\begin{bmatrix}
    x \\
    y
\end{bmatrix}
\]

\[
P = \begin{bmatrix}
    \sigma_x^2 & \rho \sigma_x \sigma_y \\
    \rho \sigma_x \sigma_y & \sigma_y^2
\end{bmatrix}
\]

\[
\begin{bmatrix}
    x \\
    y
\end{bmatrix}
\]

\[
\begin{bmatrix}
    \sigma_x^2 & \rho \sigma_x \sigma_y \\
    \rho \sigma_x \sigma_y & \sigma_y^2
\end{bmatrix}
\]

\[= K \]

\[
\lambda_1 = \frac{1}{2} \left[ \sigma_x^2 + \sigma_y^2 + \sqrt{(\sigma_x^2 - \sigma_y^2)^2 + 4 \rho^2 \sigma_x^2 \sigma_y^2} \right]
\]

\[
\lambda_2 = \frac{1}{2} \left[ \sigma_x^2 + \sigma_y^2 - \sqrt{(\sigma_x^2 - \sigma_y^2)^2 + 4 \rho^2 \sigma_x^2 \sigma_y^2} \right]
\]

\[
\frac{w_1^2}{K \lambda_1} + \frac{w_2^2}{K \lambda_2} \leq 1
\]

\[
\alpha = \frac{1}{2} \tan^{-1} \left( \frac{2 \rho \sigma_x \sigma_y}{\sigma_x^2 - \sigma_y^2} \right),
\]

\[-\frac{\pi}{4} \leq \alpha \leq \frac{\pi}{4}, \quad \sigma_x^2 \neq \sigma_y^2\]
DISCRETE KALMAN FILTER

Probabilistic interpretation of the error ellipsoid

\[ p(x_k \mid Z_0^k, U_0^{k-1}) \sim N(\hat{x}(k \mid k), P(k \mid k)) \]

• Given \( \hat{x}(k \mid k) \) and \( P(k \mid k) \) it is possible to define the locus where, with a given probability, the values of the random vector \( x(k) \) lie.

Hiperellipsoid with center in \( \hat{x}(k \mid k) \) and with semi-axis proportional to the eigenvalues of \( P(k \mid k) \)
DISCRETE KALMAN FILTER
Probabilistic interpretation of the error ellipsoid

\[ p(x_k \mid Z^k_0, U^{k-1}_0) \sim N(\hat{x}(k \mid k), P(k \mid k)) \]

- Example for \( n=2 \)

\[ M = \{ x_k : (x_k - \hat{x}(k \mid k))^T P(k \mid k)^{-1} [x_k - \hat{x}(k \mid k)] \leq K \} \]

\[ \text{Pr}\{x_k \in M\} \]

- is a function of \( K \)
- a pre-specified values of this probability can be obtained by an appropriate choice of \( K \)
DISCRETE KALMAN FILTER

Probabilistic interpretation of the error ellipsoid

\[ p(x_k \mid Z_0^k, U_0^{k-1}) \sim N(\hat{x}(k \mid k), P(k \mid k)) \]

\[ [x_k - \hat{x}(k \mid k)]^T P(k \mid k)^{-1} [x_k - \hat{x}(k \mid k)] \leq K \]

(Scalar) random variable with a \( \chi^2 \) distribution with \( n \) degrees of reedom

- How to chose \( K \) for a desired probability?
  - Just consult a Chi square distribution table

<table>
<thead>
<tr>
<th>Probability</th>
<th>( n=1 )</th>
<th>( n=2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>( K=2.71 )</td>
<td>( K=4.61 )</td>
</tr>
<tr>
<td>95%</td>
<td>( K=3.84 )</td>
<td>( K=5.99 )</td>
</tr>
</tbody>
</table>
DISCRETE KALMAN FILTER
The error ellipsoid and the filter dynamics

- Prediction cycle

\[ x_{k+1} = A_k x_k + B_k u_k + G_k w_k \]

\[ \hat{x}(k+1 | k) = A_k \hat{x}(k | k) + B_k u_k \]

\[ P(k+1 | k) = A_k P(k | k) A_k^T + G_k Q_k G_k^T \]
DISCRETE KALMAN FILTER
The error ellipsoid and the filter dynamics

\[ \hat{x}(k + 1 | k + 1) = \hat{x}(k + 1 | k) + K(k + 1)r(k + 1) \]

\[ P(k + 1 | k + 1) = P(k + 1 | k) - K(k + 1)C_{k+1}P(k + 1 | k) \]

- Filtering cycle

\[ z_{k+1} = C_{k+1}x_{k+1} + v_{k+1} \]

\[ r_{k+1} = z_{k+1} - C_{k+1}\hat{x}(k + 1 | k) \]

\[ S(k + 1) = C_{k+1}P(k + 1 | k)C_{k+1}^T + R_{k+1} \]
Extended Kalman Filter

• **Non linear** dynamics

• **White Gaussian** system and observation noise

\[
x_{k+1} = f_k(x_k, u_k) + w_k
\]
\[
z_k = h_k(x_k) + v_k
\]

\[
x_0 \sim N(\bar{x}_0, P_0)
\]
\[
E[w_k w_j^T] = Q_k \delta_{kj}
\]
\[
E[v_k v_j^T] = R_k \delta_{kj}
\]

• **QUESTION:** Which is the MMSE (minimum mean-square error) estimate of \(x(k+1)\)?

  - Conditional mean \(\hat{x}(k + 1 | k) = E(x_{k+1} | Z_1^k, U_0^k)\)?
  - Due to the non-linearity of the system,

\[
p(x_k | Z_1^k, U_0^{k-1}) \quad p(x_{k+1} | Z_1^k, U_0^k) \quad p(x_{k+1} | Z_1^{k+1}, U_0^k)
\]

are non Gaussian
Extended Kalman Filter

- **(Optimal)** ANSWER: The MMSE estimate is given by a non-linear filter, that propagates the conditonal pdf.

- The **EKF** gives an approximation of the optimal estimate
  - The non-linearities are approximated by a linearized version of the non-linear model around the last state estimate.
  - For this approximation to be valid, this linearization should be a good approximation of the non-linear model in all the uncertainty domain associated with the state estimate.
Extended Kalman Filter

\[ p(x_k | Z_1^k, U_0^{k-1}) \quad \xrightarrow{\text{Apply KF to the linear dynamics}} \quad \hat{x}(k | k) \]

\[ p(x_{k+1} | Z_1^k, U_0^k) \quad \xrightarrow{\text{Apply KF to the linear dynamics}} \quad \hat{x}(k+1 | k) \]

\[ x_{k+1} = f_k(x_k, u_k) + w_k \]

linearize around \( \hat{x}(k | k) \)

\[ z_{k+1} = h_{k+1}(x_{k+1}) + v_{k+1} \]

linearize around \( \hat{x}(k+1 | k) \)
Extended Kalman Filter

linearize around $\hat{x}(k \mid k)$

\[
f_k(x_k, u_k) \approx f_k(\hat{x}_{k\mid k}, u_k) + \nabla f_k \bigg|_{x_k = \hat{x}_{k\mid k}} (x_k - \hat{x}_{k\mid k}) + \ldots
\]

\[
x_{k+1} \equiv \nabla f_k \bigg|_{x_k = \hat{x}_{k\mid k}} x_k + w_k + (f_k(\hat{x}_{k\mid k}, u_k) - \nabla f_k \hat{x}_{k\mid k})
\]

Prediction cycle of KF

\[
\hat{x}_{k+1\mid k} = \nabla f_k \hat{x}_{k\mid k} + (f_k(\hat{x}_{k\mid k}, u_k) - \nabla f_k \hat{x}_{k\mid k})
\]

\[
P(k+1 \mid k) = \nabla f_k P(k \mid k) \nabla f_k^T + Q_k
\]
Extended Kalman Filter

linearize $z_{k+1} = h_{k+1}(x_{k+1}) + v_{k+1}$ around $\hat{x}(k+1 \mid k)$

$$h_{k+1}(x_{k+1}) \approx h_{k+1}(\hat{x}_{k+1 \mid k}) + \nabla h_{k+1} \bigg|_{x_{k+1} = \hat{x}_{k+1 \mid k}}(x_{k+1} - \hat{x}_{k+1 \mid k}) + \ldots$$

$$z_{k+1} \approx \nabla h_{k+1} \bigg|_{x_{k+1} = \hat{x}_{k+1 \mid k}} x_{k+1} + v_k + (h_{k+1}(\hat{x}_{k+1 \mid k}) - \nabla h_{k+1} \bigg|_{x_{k+1} = \hat{x}_{k+1 \mid k}} \hat{x}_{k+1 \mid k})$$

Update cycle of KF

$$\hat{x}_{k+1 \mid k+1} = \hat{x}_{k+1 \mid k} + P(k+1 \mid k)\nabla h_{k+1}^T(\nabla h_{k+1} P(k+1 \mid k) \nabla h_{k+1}^T + R_{k+1})^{-1}[z_{k+1} - h_{k+1}(\hat{x}_{k+1 \mid k})]$$

$$P(k+1 \mid k+1) = P(k+1 \mid k) - P(k+1 \mid k) \nabla h_{k+1}^T [\nabla h_{k+1} P(k+1 \mid k) \nabla h_{k+1}^T + R_{k+1}]^{-1} \nabla h_{k+1} P(k+1 \mid k)$$
References

- Peter S. Maybeck, “The Kalman Filter: an Introduction to Concepts”