Instituto Superior Técnico

Calculus of variations and Optimal Control

Problem Series nº 2

P1. Snell's law of refraction and Fermat's Principle of Least Time

Consider the situation shown in figure P1-1 where a light ray travels from A to B through two optical media separated by a flat surface, and such that, from A to C the velocity in medium I is c_1, and from C to B the velocity in medium II is c_2.

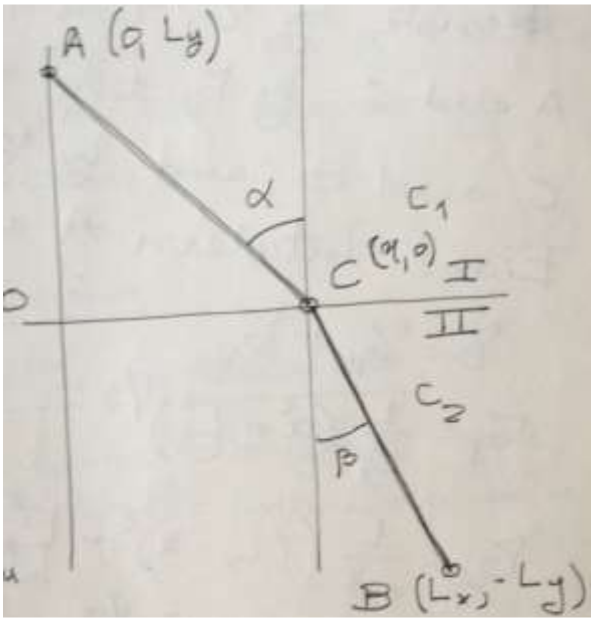

Figure P1-1. Deduction of Snell's law of refraction from Fermat's Principle of Least Time.

Write the first order necessary condition for x (the abscissa of point C) to be the minimum for the time of travel τ from from A to B, and show that this condition is equivalent to Snell's law:

$$\frac{\sin\alpha}{\sin\beta} = \frac{c_1}{c_2}$$

Hint: Denote by τ_1 the travel time between A and C, by τ_2 the travel time between C and B, and by τ the total travel time between A and B, that is a function of x. Compute τ_1 and τ_2 as a function of c_1 and c_2, and use the 1st order necessary conditions for minimum.

Remarks on exercise P1

The Principle of Least Time

The search for a general principle that allows to predict the path of a light ray in geometrical optics is very old and may be traced back to the Ancient Greece, where Heron of Alexandria stated that, for an homogeneous media, the light follows the shortest path ([FITAS2012], p. 14, Table 1.1). This principle allowed him to interpret reflection. In turn, the arab Ibn Haytham stated that a light ray follows the easiest way, in the sense that it founds the least resistance ([FITAS2012], p. 14, Table 1.1).

Of interest to modern thought in what concerns Variational Calculus, Fermat stated in 1162, in a letter to a correspondant, Marie Cureau de la Chambre, the Principle of Least Time ([GOLDSTINE1980], p. 1). Starting from the observation of Galileo that particles moving under the action of gravity follow paths that require the least time to travel, instead of the ones of least length, Fermat stated that *nature operates by means and ways that are easiest and fastest*. Combining this principle with his original method to find maxima and minima of functions, Fermat deduced Snell's law in a way that was very much similar to exercise P1. See [GOLDSTINE1980], p. 2-6, for a description of the original Fermat's deduction.

Fermat's principle of least time may be stated as follows: *From all the possible paths that light might follow to go from one point to another, the path that requires the shortest time is chosen*. This principle might be used to deduce all the results in geometrical optics ([FEYNMAN1975], ch. 26).

In addition to illustrating the use of necessary conditions for the minima of a $\mathbb{R}^n \rightarrow \mathbb{R}$ function, Fermat's principle has the interest of having been used by Johann Bernouilli in his solution of the Brachistochrone problem, a topic to be addressed latter, that amounts to find the shape of a wire such that a bead sliding on it by the sole action of the gravity force goes from a starting to an end point in the least time. Indeed, Bernouilli replaced the bead by a light ray that crosses a series of optically transparent media with different densities [SUSSMANN1997].

Fermat was much criticized by the disciples of Descartes on the basis that, if aa light ray would follow a least time path between two points, it would behave according to a moral principle [FITAS2012]. Although the argument of the Cartesians was sound, it was refuted because they committed a mistake. Indeed, the Cartesians were convinced that the speed of light was higher in more dense media, probably by assuming that the situation was similar to what happens with the speed of sound.

Instead, Fermat assumed the opposite, and together with his Principle, and his method to compute maxima and minima of functions, he could deduce in a rigorous way Snell's law, that had a striking experimental verification.

Can we thus conclude that light has a "moral sense"? In other terms, how does a light ray knows the correct path, for instance when passing fron media to another?

Actually, the answer is that it does not know, and was provided by Quantum Mechanics [FEYNMAN2005] and the solution of the Schroedinger equation. The solution of this equation, and its relation with the Principle of Least Time is explained in an elementary and beautiful way in [FEYNMAN1985]. It is therefore not a surprise that the Schroedinger equation may be deduced from Pontryagin's Maximum Principle [LEVI2010].

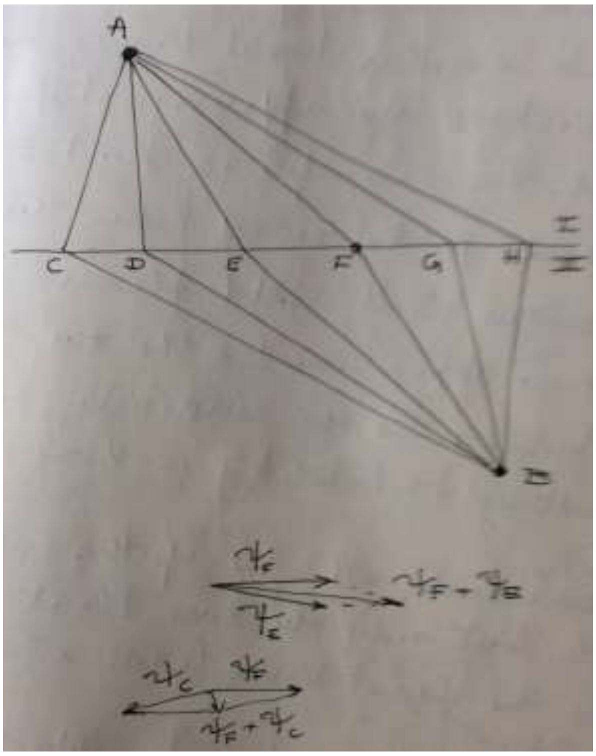

Figure P1-2.

In order to understand how light appears to follow the Principle of Least Time from a macroscopic point of view, consider (figure P1-2) a source of light at point A [FEYNMAN1985]. According to geometrical optics, and the Principle of

Least Time, a light ray that reaches point B will follow a path such that the travel time between A and $B$is minimum. Call F the point at the surface between medium I and medium II that corresponds to this path.

According to Quantum Mechanics, however, the description is not in terms of light rays but in probabilistic terms. As such, the probability that a particle of light – a photon – that leaves A reaches $B$is the probability that it passes through F compounded with the one corresponding to any other point , including not only C, D, E, etc., but all the points in the surface as well.

The probability associated to a single path i is the square o the modulus of a vector Ψ_i. The composition of the probabilities associated to different paths is made by adding the vectors, and then computing the square of the amplitude of the resulting vector.

Each individual vector is of unit length and has a phase that is proportional to the time of travel of the corresponding photon. Since F corresponds to a minimum, where the derivative vanishes (1^{st} order necessary condition), points close to it will have a similar travel time, and these vectors are approximately aligned. An example is point E. Instead, points away from the minimum will yield phases that can be in opposition to the one that corresponds to the minimum, and the resulting probability is very small (see the lower part of figure P1-2).

To conclude, all the paths are possible, but only the ones that are close to the least time have a high probability of being followed. The above justification, presented in [FEYNMA1985], is somewhat *ad hoc*, but may actually be justified on the basis of Schroedinger equation.

References to the remark

[GOLDSTINE1980]

H. H. Goldstine (1980). *A history of the calculus of variations*. Springer-Verlag.

[FITAS2012]

A. Fitas (2012). *O Princípio da Menor Acção*. Caleidoscópio.

[FEYNMAN1975]

R. P. Feynman, R. B. Leighton, and M. Sands (1975). *The Feynman lectures on Physics*. 5th printing. Vol. I.

[FEYNMAN1985]
R. P. Feynman (1985). *QED – The strange theory of light and matter*. Penguin books.

[SUSSMANN1997]
H. J. Sussmann and J. C. Willems (1997). 300 years of Optimal Control: From the Brachistochrone to the Maximum Principle. *IEEE Control Systems*, June 1997, 32-44.

[FEYNMAN2005]
R. P.Feynman and A. R. Hibbs (2005). *Quantum mechanics and path integrals*. Dover, emended edition.

[LEVI2010]
M. Levi (2010). *Classical mechanics with calculus of variations and optimal control*. Americam Mathematical Society.

P2. Consider (figure P2-1) the rectangular box defined by the origin of the coordinates $\mathcal{O}$ $(0,0,0)$ and point $\wp$, with coordinates (x_1, x_2, x_3), where $x_i > 0$.

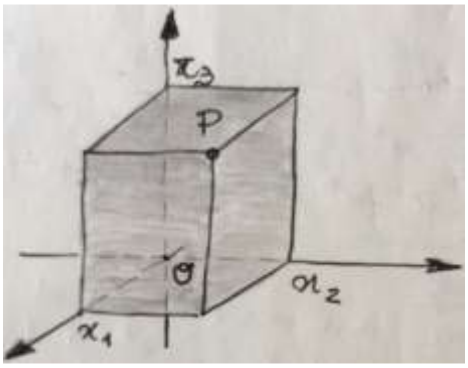

Figure P2-1. The rectangular box to be optimized in problem P1.

In this problem we want to find the values of the coordinates of the point $\wp$ such that the enclosed volume is a (a given generic value) and the sum S of the areas of all surfaces is minimum.

Solve this problem in two different ways:

a) Eliminate x_3 from $S(x_1, x_2, x_3)$ using the constraint on the volume to express x_3 as a function of x_1 and x_2 and obtain the stationary point(s) of the resulting function $S = \varphi(x_1, x_2)$. Use the 2nd order sufficient condition to prove that your solution is actually a minimum.

b) Use Lagrange multipliers to tackle the constraint on the volume.

P3. Consider the linear, time invariant, SISO system described by the discrete time state model

$$x(k+1) = \Phi x(k) + \Gamma u(k)$$
$$y(k) = Cx(k)$$

This is a 2nd order system and we know that

$$\Phi = \begin{bmatrix} 0 & 1 \\ 0{,}5 & 0{,}4 \end{bmatrix}, \quad \Gamma = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad C = [1 \quad 1],$$

The system starts from zero initial conditions:

$$x(0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Using the method of Lagrange multipliers, compute the sequence of values of the control variable $u(0), u(1), u(2)$ that drives the system output to the reference value $r = 5$, at time instant $k = 3$, that is to say, that forces $y(3) = r = 5$, and is such that the energy of the control sequence, given by

$$J(u) = u(0)^2 + u(1)^2 + u(2)^2,$$

Is minimum. Start by obtaining general expressions for $u(0), u(1), u(2)$ as functions of Φ, Γ, C e r, and only afterwards obtain numerical values.

Hint: By iterating the state equation, start by obtaining expressions for $x(k)$, $k = 1, 2, 3$.

P4. In general, the 1st and 2nd order necessary conditions for minimum are not sufficient. This exercise provides an example.

On the (x_1, x_2) plane, consider the function $x_1(1 + x_1) + x_2(1 + x_2)$. Let D be the union of the closed first quadrant $\{(x_1, x_2) : x_1 \geq 0, \ x_2 \geq 0\}$ and some curve (*e.g.* a circular arc) directed from the origin into the first quadrant (figure P4-1).

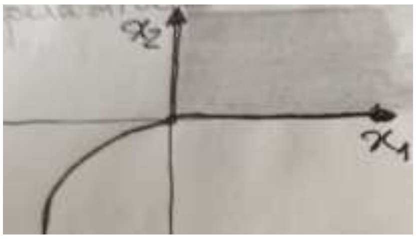

Figure P4-1. The domain of the function considered in problem P4-1.

a) Show that the origin is **not** a local minimum;
b) State what are the feasible directions at the origin;
c) Show that, although the necessary conditions for a local minimum hold, the origin is not a local minimum. Take into consideration that you are in the presence of an inequality constraint.

P5. Prove that if x^* is a local minimum of f (not necessarily in the interior of D), then

$$d^T \nabla^2 f(x^*) \cdot d \geq 0$$

for every feasible direction d that satisfy

$$\nabla f(x^*) \cdot d = 0$$

Hint: Modify the argument used in the 2st order necessary condition for minimum at an interior point. Consider the function $g(\alpha) \coloneqq f(x^* + \alpha d)$, where $\alpha \in \mathbb{R}$ and show that $g'(0) = 0$. Use a 2nd order Taylor expansion of g to conclude that $g''(0) \geq 0$. Finally, prove that $g''(0) = d^T \nabla^2 f(x^*) d$.

P6. Consider $x_1, \dots, x_k \in C \subset \mathbb{R}^n$ a convex set, and $\theta_1, \dots, \theta_k \in \mathbb{R}$ that satisfy $\theta_i \geq 0\ \forall_{i=1,\dots,k}$ and $\theta_1 + \cdots + \theta_k = 1$. Show that $\theta_1 x_1 + \cdots + \theta_k x_k \in C$.

Hint: Use mathematical induction on k.

Proof by mathematical induction

Consider the problem of proving that some proposition $\wp_k$ that depends on a mathematical entity that varies with the $k \in \mathbb{N}$ holds for any integer k. The method of Mathematical induction consists of the following steps:

1. Prove that it holds for $k = 1$
2. Assume that the proposition is valid for a generic k. Then show that if this assumption is valid, it must hold for $k + 1$ as well.

As the method of proof by contradiction, proof by mathematical induction is not accepted by the Intuitionists.

Example

Prove that

$$1 + 2 + 3 + \cdots + k = \frac{k(k+1)}{2}$$

Step 1: For $k = 1$ the equality is true because

$$1 = \frac{1 \times (1+1)}{2}$$

Step 2: Assume that the equality is true for some $k \geq 1$, that is, assume that it is true that

$$1 + 2 + 3 + \cdots + k = \frac{k(k+1)}{2}$$

In this case, one has

$$1 + 2 + 3 + \cdots + k + (k+1) = \frac{k(k+1)}{2} + (k+1) = (k+1)\left(\frac{k}{2} + 1\right) = \frac{(k+1)(k+2}{2},$$

Which means that, if the equality holds for k, then it also holds for $k + 1$.

P7. Consider the problem:

Minimize $f(x_1, x_2, x_3) = -3x_1^2 + x_2^2 + 2x_3^2 + 2(x_1 + x_2 + x_3)$

Subject to $x_1^2 + x_2^2 + x_3^2 = 1$

a) Find the stationary points of $L(x, \lambda)$ using the KKT conditions.

b) Compute f at each stationary point.

Use MATLAB when appropriate.