Reading group: Calculus of Variations and Optimal Control Theory by Daniel Liberzon

Alexandre Vieira

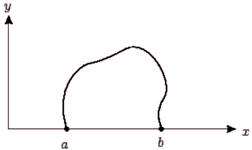
16th March 2017

Content

Content

Dido's isoperimetric problem

How to build the biggest city with the hide of an ox?



This can be formulated as a maximization problem :

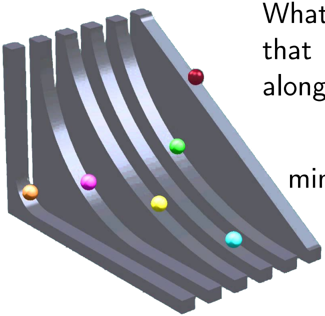
$$\text{maximize}_{y\in\mathscr{C}^1} \int_a^b y(x)dx$$
$$\text{such that } y(a) = y(b) = 0$$
$$\int_a^b \sqrt{1+(y'(x))^2}dx = C_0$$

Catenary problem

What equation describes a chain hanging between two fixed points ?

$$\begin{aligned}
\text{minimize}_{\, y \in \mathscr{C}^1} \quad & \int_a^b y(x)\sqrt{1+(y'(x))^2}\,dx \\
\text{such that} \quad & y(a) = y_a \\
& y(b) = y_b \\
& \int_a^b \sqrt{1+(y'(x))^2}\,dx = C_0
\end{aligned}$$

Brachistochrone problem



What path between two points is such that a particle sliding without friction along it takes the shortest time?

$$\text{minimize}_{y \in \mathscr{C}^1} \int_a^b \sqrt{\frac{1+(y'(x))^2}{y(x)}}dx$$

$$\text{such that } y(a) = y_a$$
$$y(b) = y_b$$

Content

Existence of minimum

We are usually interested not in local minima, but more on global minima. How could we be sure that they exist?

Theorem : Weierstrass Theorem

If f is a continuous functions and D is a compact set, then there exists a global minimum of f over D.

A compact set D is defined as a set for which every open cover has a finite subcover. Equivalently, D is compact if every sequence in D has a subsequence converging in D.
For a subset D of $\mathbb{R}^n$, compactness if easily defined: D must be closed and bounded.

First-order conditions

Consider $f : \mathbb{R}^n \to \mathbb{R}$ a $\mathscr{C}^1$-function, $D = dom\ f$ that we suppose open and non-empty.
Denote x^* a local minimum of f on D (which must be an interior point). For α small enough, $x^* + \alpha d \in D$.
Let us consider

$$g(\alpha) = f(x^* + \alpha d).$$

By construction, g is $\mathscr{C}^1$, so :

$$g(\alpha) = g(0) + \alpha g'(0) + o(\alpha)$$

First-order conditions

$$g(\alpha) = g(0) + \alpha g'(0) + o(\alpha)$$

We can easily prove that $g'(0) = 0$. For that, suppose that $g'(0) \neq 0$. Then, by the definition of $o(\alpha)$, there exists $\varepsilon > 0$ such that :

$$|\alpha| < \varepsilon,\ \alpha \neq 0 \implies |o(\alpha)| < |g'(0)\alpha|$$

And so :

$$g(\alpha) - g(0) < \alpha g'(0) + |\alpha g'(0)|$$

Restrict α to have the opposite sign to $g'(0)$, such that we have $g(\alpha) - g(0) < 0$ and so :

$$f(x^* + \alpha d) < f(x^*)$$

which contradicts the fact that x^* is a minimum. So $g'(0) = 0$.

First-order conditions

A simple application of the chain rule implies :

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

But since d is arbitrary, we conclude that :

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

Second-order conditions

Suppose now that f is $\mathscr{C}^2$. As before:

$$g(\alpha) = g(0) + \alpha g'(0) + \frac{1}{2}g''(0)\alpha^2 + o(\alpha^2)$$

We easily prove that $g''(0) \geq 0$. The same way we did for 1st order conditions, suppose $g''(0) < 0$. Then there exists $\varepsilon > 0$ such that :

$$|\alpha| < \varepsilon,\ \alpha \neq 0 \implies |o(\alpha^2)| < \frac{1}{2}|g''(0)\alpha^2|$$

And we prove exactely the same way that $g(\alpha) - g(0) < 0$, thus contradicting the fact that x^* is a local minimum.

Second-order conditions

Once again, using the chain rule:

$$g''(0) = d^\intercal \nabla^2 f(x^*) d \geq 0$$

and since d is arbitrary, $\nabla^2 f(x^*)$ must be positive semidefinite.

Unconstrained border conditions

We supposed so far that D is an open set. What if D is closed and $x^* \in \partial D$? There will be some d such that $x^* + \alpha d \notin D$ whatever α.

Definition: Feasible direction

We call a feasible direction at x^* a vector d such that $x^* + \alpha d \in D$ for small enough $\alpha > 0$.

Unconstrained border conditions

We can actually easily prove that we can replace the first and second-order conditions by: if x^* is a local minimum of f on D:

- if f is $\mathscr{C}^1$, then $\nabla f(x^*) \cdot d \geq 0$ for every feasible direction d
- if f is $\mathscr{C}^2$, then $d^\intercal \nabla^2 f(x^*) d \geq 0$ for all feasible directions satisfying $\nabla f(x^*) \cdot d = 0$.

Also, if D is convex, then every point of D is written $x^* + \alpha d$, so this kind of approach is clearly suitable for this kind of set.

Lagrange multipliers

We want now to minimize a function f on a set D defined by equalities:

$$h_1(x) = ... = h_m(x) = 0$$

Definition: regular point

x^* is called a regular point if $\nabla h_i(x^*)$, $i = 1, ..., m$ are linearly independant.

As we discussed before, the method used with feasible direction is not necessarily suitable for constrained cases. So here, we would rather use *feasible curves*.

Lagrange multipliers

We consider a curve $x(\alpha) \in D$, supposed $\mathscr{C}^1$, with $x(0) = x^*$. Given an aribtrary curve of this kind, we can consider

$$g(\alpha) = f(x(\alpha))$$

For $\alpha = 0$, we can prove:

$$g'(0) = \nabla f(x^*)x'(0) = 0$$

This vector $x'(0)$ is a tangent vector to D at x^*. It lives in the tangent space to D at x^*, which is denoted by $T_{x^*}D$.

From this, we can prove that if x^* is a minimum of f in D, then there exists $\lambda \in \mathbb{R}^m$ such that:

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) = 0$$

This is a really geometrical point of view. But we will prove that in a more algebraic way.

Lagrange multipliers

Suppose that $m = 1$. Given two vectors d_1, d_2 arbitrary, consider :

$$F : (\alpha_1, \alpha_2) \mapsto (f(x^* + \alpha_1 d_1 + \alpha_2 d_2), h(x^* + \alpha_1 d_1 + \alpha_2 d_2))$$

The Jacobian matrix of F at $(0, 0)$ is :

$$J_F(0,0) = \begin{pmatrix} \nabla f(x^*) d_1 & \nabla f(x^*) d_2 \\ \nabla h(x^*) d_1 & \nabla h(x^*) d_2 \end{pmatrix}$$

Theorem: Inverse Function Theorem

If the total derivative of a continuously differentiable function F defined from an open set of $\mathbb{R}^n$ into $\mathbb{R}^n$ is invertible at a point p (i.e., the Jacobian determinant of F at p is non-zero), then F is an invertible function near p. Moreover, the inverse function F^{-1} is also continuously differentiable.

Lagrange multipliers

If $J_F(0,0)$ is non singular, then by the Inverse Function Theorem, there are neighbourhoods of $(0,0)$ and $F(0,0) = (f(x^*), 0)$ on which F is a bijection.
That implies that there is a point $x = x^* + \alpha_1 d_1 + \alpha_2 d_2$, with (α_1, α_2) close to $(0,0)$, such that $h(x) = 0$ and $f(x) < f(x^*)$.
This cannot be true since x^* is a minimum.

So $J_F(0,0)$ must be singular.

Lagrange multipliers

$$J_F(0,0) = \begin{pmatrix} \nabla f(x^*)d_1 & \nabla f(x^*)d_2 \\ \nabla h(x^*)d_1 & \nabla h(x^*)d_2 \end{pmatrix}$$

We suppose that x^* is a regular point. In this case, it means that $\nabla h(x^*) \neq 0$. Choose d_1 such that $\nabla h(x^*)d_1 \neq 0$. Let

$$\lambda = -\frac{\nabla f(x^*)d_1}{\nabla h(x^*)d_1}$$

Since $J_F(0,0)$ is singular, its first row must be a constant multiple of its second row. Thus :

$$\nabla f(x^*)d_2 + \lambda \nabla h(x)d_2 = 0, \ \forall d_2 \in \mathbb{R}^n$$

Since d_2 is arbitrary, we have the well know equation:

$$\nabla f(x^*) + \lambda \nabla h(x) = 0$$

Content

Can we use this in an infinite-dimensional framework?

Consider now

$$J : V \to \mathbb{R}$$

V : infinite-dimensional vector space (usually, a function space).
J is called a functional.
Main difference with finite-dimensional case: no "generic" space.

There can be several V where the problem may be well defined, and choosing this V may be part of the problem.

Can we use this in an infinite-dimensional framework?

In finite-dimensional cases, all norms are equivalent. It is not the case for function spaces.

- On $\mathscr{C}^0([a,b],\mathbb{R}^n)$, we can use:

$$\|y\|_0 = \max_{a \leq x \leq b} |y(x)|$$

- On $\mathscr{C}^1([a,b],\mathbb{R}^n)$, we can use:

$$\|y\|_1 = \max_{a \leq x \leq b} |y(x)| + \max_{a \leq x \leq b} |y'(x)|$$

- There may be other norms (L^p,...)

Once this norm is chosen, we can clearly define a local optimum for the functional J over V.

First variation

Also, what do we call a derivative of a functional?

Definition: First Variation

For a function $y \in V$, we call the first variation of J at y, the linear functional $\delta J|_y : V \to \mathbb{R}$ satisfying, for all η and all α:

$$J(y + \alpha\eta) = J(y) + \delta J|_y(\eta)\alpha + o(\alpha)$$

First-order condition

Can we the same way, define a first-order condition? Suppose we want to find a local minimum of a functional J over a subset A of V.

Definition: Admissible perturbation

We call a perturbation $\eta \in V$ *admissible* if $y^* + \alpha\eta \in A$ for all α close enough to 0.

Reasonning the same way as before, we prove that if y^* is a local minimum, then for all admissible perturbations η, we must have

$$\delta J_{|y}(\eta) = 0$$

Problem : how do we compute $\delta J_{|y}$?

Second variation

Definition: Bilinear functional and quadratic form

$B : V \times V \to \mathbb{R}$ is bilinear if it is linear in each argument.
$Q(y) = B(y, y)$ is called a quadratic form on V.

Definition: Second variation

For a function $y \in V$, we call the second variation of J at y, the quadratic form $\delta^2 J|_y : V \to \mathbb{R}$ satisfying, for all η and all α:

$$J(y + \alpha\eta) = J(y) + \delta J|_y(\eta)\alpha + \delta^2 J|_y(\eta)\alpha^2 + o(\alpha^2)$$

Second-order condition

We still have that, if y^* is a local minimum of J over A, then for all admissible perturbations η, we have

$$\delta^2 J|_y(\eta) \geq 0$$

But the condition here is not sufficient! In the finite dimensional-case, sufficiency is proved by the fact that the unit ball is compact (and then using Weierstrass Theorem). This is not the case in infinite-dimensional case!

Content

Exercices

- page 6-7: prove the necessary (and sufficient) conditions in the case where $x^* \in \partial D$ (suppose D convex).
- page 15: Prove the first order condition in the constrained case with $m > 1$.
- Exercise 1.2 : have a look at the following constrained problem :

$$\begin{aligned} &\min x_1 + x_2 \\ \text{s.t. } &(x_1 + 2)^2 + x_2^2 = 2 \\ &(x_1 - 2)^2 + x_2^2 = 2 \end{aligned}$$

  Find the minimum solution (you can prove before that it exists!). Is it a regular point? Have a look also on the KKT conditions (Lagrange multipliers).
- Exercise 1.5 and 1.6 (first and second variations)