HUMBERTO STEIN SHIROMOTO

OPTIMISATION

The public is more familiar with bad design than good design. It is, in effect, conditioned to prefer bad design, because that is what it lives with. The new becomes threatening, the old reassuring.

Paul Rand, *Design, Form, and Chaos*

La perfection est atteinte, non pas lorsquíl n'y a plus rien à ajouter, mais lorsquíl n'y a plus rien à retirer.

Antoine de Saint-Exupéry

PUBLISHED BY

HUMBERTO.STEIN-SHIROMOTO.INFO

First printing, August 2021

Contents

Note use [**Liberzon2012**] as main reference

Part I

Convex Optimisation

1 | Linear Programming

2 | Convex programming

Part II

Infinite Dimensional

3 | Variational Analysis

This is all from Liberzon's book

To formulate the variational problem, first consider the $\mathcal{L}_p$ norm defined as

$$\|y\|_{\mathcal{L}_p} = \left(\int_a^b |y(x)|^p \, dx \right)^{1/p} , \tag{3.1}$$

where $p \in \mathbb{R}_{\geqslant 0}$. Now, the local minima of function can be defined

Definition 3.1. Let $\mathbf{V}$ be a vector space of functions equipped with a norm $\|\cdot\|$, let $\mathbf{A}$ be a subset of $\mathbf{V}$, and let J be a real-valued functional defined on $\mathbf{A}$. A function $y^* \in \mathbf{A}$ is a *local minimum* of J over $\mathbf{A}$ if there exists $\varepsilon > 0$ such that the inequality

$$J(y^*) \leqslant J(y)$$

holds, for all $y \in \mathbf{A}$ satisfying $\|y - y^*\| < \varepsilon$. ⌟

3.1 First Variation and First-order Necessary Condition

Definition 3.2. [First variation] Let $\mathbf{V}$ be a function space, and let $J : \mathbf{V} \to \mathbb{R}$ be a functional. The linear functional $\delta J|_y : \mathbf{V} \to \mathbb{R}$ is called the *first variational of* J *at* y if, for every $\eta \in \mathbf{V}$ and for every $\alpha \in \mathbb{R}$, the equality

$$J(y + \alpha\eta) = J(y) + \delta J|_y(\eta)\alpha + o(\alpha) , \tag{3.2}$$

where o represents higher-order terms[1]. ⌟

[1] $o(\alpha)/\alpha \to 0$, as $\alpha \to 0$.

The first variation defined in Definition **??** corresponds to the Gateaux derivative of J:

$$\delta J|_y(\eta) = \lim_{\alpha \to 0} \frac{J(y + \alpha\eta) - J(y)}{\alpha} .$$

Definition 3.3. Let $\mathbf{V}$ be a function space, and let $J : \mathbf{V} \to \mathbb{R}$ be a functional. Let y^* be a local minimum of J, the function $\alpha\eta$, where $\alpha \in \mathbb{R}$ and the function $\eta \in \mathbf{V}$, is said to be an *admissible perturbation* (with respect to the subset $\mathbf{A}$) if $y^* + \alpha\eta \in \mathbf{A}$. ⌟

Theorem 3.4. *Let* $\mathbf{V}$ *be a function space, and let* $J : \mathbf{V} \to \mathbb{R}$ *be a functional. If* $y^* \in \mathbf{A}$ *is a local minimum of* J *over* $\mathbf{A}$, *then*

$$\delta J|_{y^*}(\eta) = 0 .$$

⌟

Proof. Theorem **??** claims that

$$\delta J|_{y^*}(\eta) = 0 \, .$$

To show this, suppose that $\delta J|_{y^*}(\eta) \neq 0$. Then, from the higher-order term of equation (**??**), there exists $\varepsilon > 0$ small enough so that the inequality

$$\|\alpha\| < \varepsilon, \alpha \neq 0$$

implies

$$\|o(\alpha)\| < \|\delta J|_{y^*}(\eta)\alpha\| \, .$$

For these values of α, equation (**??**) gives

$$J(y^* + \alpha\eta) - J(y^*) < \delta J|_{y^*}(\eta)\alpha + \|\delta J|_{y^*}(\eta)\alpha\| \, .$$

If α is restricted to have the opposite sign to $\delta J|_{y^*}(\eta)$, the the previous equation becomes $J(y^* + \alpha\eta) - J(y^*) < 0$. But this contradicts the fact that J has a minimum at y^*. Thus, the conclusion holds. □

3.2 Basic Calculus of Variations Problem

Consider a function $L : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. Among all $\mathcal{C}^1$ curves $y : [a, b] \to \mathbb{R}$ satisfying given boundary conditions

$$y(a) = y_0, \quad y(b) = y_1 \tag{3.3a}$$

find (local) minima of the cost functional

$$J(y) = \int_a^b L(x, y(x), y'(x)) \, dx \, . \tag{3.3b}$$

The function L is called *Lagrangian*

3.3 First-order Necessary Condition for Weak Extrema: The Euler- Lagrange Equation

Let y be a test curve in **A**. For an admissible perturbation, the new curve $y + \alpha\eta$ must satisfy the boundary conditions (**??**). This is true if and only if

$$\eta(a) = \eta(b) = 0 \, .$$

Recall the first variation of J. The left-hand side of Equation (**??**) is given by

$$J(y + \alpha\eta) = \int_a^b L(x, y(x) + \alpha\eta(x), y'(x) + \alpha\eta'(x)) \, dx \, . \tag{3.4}$$

The first-order Taylor expansion of J with respect to α is given by expanding L with respect to this term. From the chain rule,

$$\begin{aligned} J(y + \alpha\eta) \quad = \quad & \int_a^b L(x, y(x), y'(x)) \, dx \\ & + \int_a^b \frac{\partial L}{\partial y}(x.y(x), y'(x))\alpha\eta(x) \, dx \\ & + \int_a^b \frac{\partial L}{\partial y'}(x.y(x), y'(x))\alpha\eta'(x) \, dx \\ & + \int_a^b o(\alpha) \, dx \, . \end{aligned}$$

Matching this with the right-hand side of Equation (**??**), one deduces that the first variation is

$$\delta J|_y(\eta) = \int_a^b \left(\frac{\partial L}{\partial y}(x, y(x), y'(x))\eta(x) + \frac{\partial L}{\partial y'}(x, y(x), y'(x))\eta'(x) \right) \, dx \, . \tag{3.5}$$

The dependence of Equation (**??**) on η' can be eliminated using integration by parts in the second term of the right- hand side:

$$\begin{aligned} \delta J|_y(\eta) \quad = \quad & \int_a^b \frac{\partial L}{\partial y}(x, y(x), y'(x))\eta(x) \, dx \\ & + \int_a^b -\eta(x) \frac{\partial^2 L}{\partial x \partial y'}(x, y(x), y'(x)) \, dx \\ & + \frac{\partial L}{\partial y'}(x, y(x), y'(x))\eta(x)|_a^b \, , \end{aligned} \tag{3.6}$$

where the last term is zero due to boundary conditions.

Thus, if y is an extremum, then one must have (cf. Theorem **??**)

$$\int_a^b \left(\frac{\partial L}{\partial y}(x, y(x), y'(x)) - \frac{\partial^2 L}{\partial x \partial y'}(x, y(x), y'(x)) \right) \eta(x) \, dx = 0, \tag{3.7}$$

for all $\mathcal{C}^1$ curves η vanishing at endpoints $x = a$ and $x = b$.

Lemma 3.5. If a continuous function $\zeta : [a, b] \to \mathbb{R}$ is such that

$$\int_a^b \zeta(x)\eta(x) \, dx = 0 \, ,$$

for all $\mathcal{C}^1$ functions $\eta : [a, b] \to \mathbb{R}$ with $\eta(a) = \eta(b) = 0$, then $\zeta \equiv 0$. ⌟

From Equation (**??**) and Lemma (**??**), y is an extremum if the equality Equations!Euler-Lagrange

$$\frac{\partial L}{\partial y}(x, y(x), y'(x)) = \frac{\partial^2 L}{\partial x \partial y'}(x, y(x), y'(x)) \tag{Euler-Lagrange}$$

holds, for every $x \in [a, b]$.

3.3.1 Hamilton's Canonical Equations

The quantity

$$p = \frac{\partial^2 L}{\partial x \partial y'}(x, y, y') \tag{momentum}$$

is called *momentum*. It is usually regarded as a function of x associated to a given curve $y = y(x)$. The second object is the *Hamiltonian*

$$H(x, y, y', p) = p \cdot y' - L(x, y, y') \, , \tag{Hamiltonian}$$

which is written as a function of four variables but becomes a function of x, when evaluated along a curve.

The variables p and y are said to be *canonical variables*. When y satisfies the (**??**) equations, the differential equations describing the evolution of y and p along this extremal, when written in terms of the Hamiltonian H, takes the form

$$\frac{dy}{dx} = y'(x) = \frac{\partial H}{\partial p}(x, y(x), y'(x)) \, .$$

For p, we have

$$\frac{dp}{dx} = \frac{dL}{dx}\frac{\partial L}{\partial y'}(x, y(x), y'(x)) = \frac{\partial L}{\partial y}(x, y(x), y'(x)) = -\frac{\partial H}{\partial y}(x, y(x), y'(x))$$

In other words, the two previous conditions can be stated as

$$\begin{cases} y' &= \dfrac{\partial H}{\partial p} \\ p' &= -\dfrac{\partial H}{\partial y}\Omega \end{cases} \tag{3.8}$$

An important additional observation is that,

$$\frac{\partial H}{\partial y'}(x, y, y', p) = p - \frac{\partial L}{\partial y'}(x, y, y') = 0 \tag{3.9}$$

This suggests that, in addition to the canonical equations (**??**), another necessary condition for optimality should be that H has a stationary point as a function y' along an optimal curve. To see this fact, relabel y' as z and note that, from the previous equation,

$$H(z) = \frac{\partial L}{\partial y'}L(x, y(x), y'(x)) \cdot z - L(x, y(x), z)$$

we claim that this function has a stationary point, when z equals $y'(x)$. This implies that

$$\frac{dH}{dz}(y'(x)) = \frac{\partial L}{\partial y'}(x, y(x), y'(x)) - \left.\frac{\partial L}{\partial z}(x, y(x), z)\right|_{z=y'(x)} = 0$$

4 | Bayesian Optimization

- Eric Brochu, Vlad M. Cora, Nando de Freitas: A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. CoRR abs/1012.2599 (2010). arXiv:1012.2599. This is a good paper on the subject

- Jonas Mockus (2013). Bayesian approach to global optimization: theory and applications. Kluwer Academic

A | Bibliography

[CalafioreGhaoui2014] is a self-contained book. It presents the concepts of linear algebra used in the book. The book starts by with linear optimisation moving to cone and semidefinite optimisation. It also contains an introduction to solving algorithms and applications to machine learning, finance, control and engineering;

[Clarke:2013] is a more theoretical book. It contains elements of functional analysis, nonsmooth analysis and optimisation (generalised gradients). The generality of the optimisation formulation is achieved with the use of calculus of variations;

[Liberzon2012] is a comprehensive book on the optimisation. It starts the book by introducing finite and infinite-dimensional optimisation problems. The next subject is the calculus of variations, and optimal control.

[VandenbergheBoyd1996]