Optimal Control, Lecture 8: Calculus of Variations

Anders Hansson

Division of Automatic Control
Linkoping University

Contents

Functionals

Consider a normed linear space of continuously differentiable real-valued functions defined on $\mathcal{D} = [a, b] \subset$ **R**, which we denote by $\mathcal{C}(a, b)$.

The norm $\| \cdot \| : \mathcal{C}(a, b) \to$ **R** is defined as

$$\|y\| = \max_{x \in \mathcal{D}} |y(x)| + \max_{x \in \mathcal{D}} |y'(x)|.$$

Consider real-valued functionals J defined on $\mathcal{C}(a, b)$, *i.e.*, $J : \mathcal{C}(a, b) \to$ **R**.

Example:

$$J[y] = \int_a^b f\left(x, y(x), y'(x)\right) dx \tag{1}$$

for some function $f :$ **R** $\times$ **R** $\times$ **R** $\to$ **R**.

Extrema

Define the *increment* $\Delta J : \mathcal{C}(a,b) \to$ **R** of a functional J for a fixed $y \in \mathcal{C}(a,b)$ as

$$\Delta J[\delta y] = J[y + \delta y] - J[y].$$

In case

$$\Delta J[\delta y] = \delta J[\delta y] + \epsilon \|\delta y\|;$$

where $\delta J : \mathcal{C}(a,b) \to$ **R** is a linear functional and $\epsilon \to 0$ as $\|\delta y\| \to 0$, we say that J is *differentiable*, and we call the functional δJ the *first variation* or *differential* of J.

It can be shown that the differential of a differentiable functional is unique.

Example

The functional in (1) is differentiable if f is a differentiable function in its last two arguments. This follows from a Taylor series expansion:

$$\begin{aligned}\Delta J[\delta y] &= \int_a^b (f\left(x, y(x) + \delta y(x), y'(x) + \delta y'(x)\right) \\ &\quad - f\left(x, y(x), y'(x)\right))dx \\ &= \int_a^b \left(\frac{\partial f\left(x, y(x), y'(x)\right)}{\partial y} \delta y(x) + \frac{\partial f\left(x, y(x), y'(x)\right)}{\partial y'} \delta y'(x) \right. \\ &\quad \left. + h\left(y(x), y'(x)\right) \left\| (\delta y(x), \delta y'(x)) \right\|_2 \right) dx,\end{aligned}$$

where $h : \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ is a function that goes to zero as $(\delta y(x), \delta y'(x)) \to 0$. The latter is implied by $\|\delta y\| \to 0$.

Example ctd.

Define ϵ as

$$\begin{aligned}\epsilon &= \frac{\int_a^b h\left(y(x), y'(x)\right) \|(\delta y(x), \delta y'(x))\|_2 \, dx}{\|\delta y\|} \\ &\le \frac{\int_a^b h\left(y(x), y'(x)\right) \|\delta y\| dx}{\|\delta y\|} = \int_a^b h\left(y(x), y'(x)\right) dx,\end{aligned}$$

which converges to zero as h goes to zero. Hence this functional is differentiable with first variation

$$\delta J[\delta y] = \int_a^b \left(\frac{\partial f\left(x, y(x), y'(x)\right)}{\partial y} \delta y(x) + \frac{\partial f\left(x, y(x), y'(x)\right)}{\partial y'} \delta y'(x) \right) dx. \tag{2}$$

It is left as an exercise to show that this functional is linear, see Exercise 7.2.

Definition of Extremum

We say that a functional J has a *weak extremum* at $y^\star$ if its increment has the same sign for all y in a neighborhood of $y^\star$.[1]

More formally there should exist $\epsilon > 0$ such that $\Delta J[\delta y]$ has the same sign for all δy such that $\|\delta y\| \leq \epsilon$.

If the sign is positive we have a *weak minimum* and if the sign is negative we have a *weak maximum*.

[1] In case there are constraints on y for $x = a$ or $x = b$, δy should be constrained to be zero at those values of x. This is often called that $\delta y(x)$ is *admissible*. We will tacitly assume that we only consider such admissible δy.

Necessary Condition for Extremum

Assume that a differentiable functional J has an extremum at $y^{\star}$. Then the first variation vanishes, *i.e.*, $\delta J[\delta y] = 0$.

Proof: We have

$$\Delta J[\delta y] = \delta J[\delta y] + \epsilon \|\delta y\|,$$

where $\epsilon \to 0$ as $\|\delta y\| \to 0$. Hence for sufficiently small $\|\delta y\|$ the sign of $\Delta J[\delta y]$ will be the same as the sign of $\delta J[\delta y]$. Now assume that $\delta J[\delta y_0] \neq 0$ for some δy_0. Then for any $\alpha > 0$ we have

$$\delta J[-\alpha \delta y_0] = -\delta J[\alpha \delta y_0],$$

since δJ is linear. Hence the increment can be made to have either sign for arbitrary small δy contradicting that J has an extremum.

Second Variation

We say that a functional J is twice differentiable if the increment can be written

$$\Delta J[\delta y] = \delta J[\delta y] + \delta^2 J[\delta y] + \epsilon^2 \|\delta y\|^2$$

where δJ is the first variation, linear in δy, $\delta^2 J$ is the *second variation*, quadratic in δy, and $\epsilon \to 0$ as $\|\delta y\| \to 0$.

It can be shown that also the second variation is unique.

Example

The functional in (1) is twice differentiable if f is a twice differentiable function in its last two arguments (proof by Taylor series expansion) The second variation is

$$\delta^2 J[\delta y] = \frac{1}{2} \int_a^b \begin{bmatrix} \delta y(x) \\ \delta y'(x) \end{bmatrix}^T \begin{bmatrix} \frac{\partial^2 f(x,y(x),y'(x)))}{\partial y^2} & \frac{\partial^2 f(x,y(x),y'(x)))}{\partial y \partial y'} \\ \frac{\partial^2 f(x,y(x),y'(x)))}{\partial y \partial y'} & \frac{\partial^2 f(x,y(x),y'(x)))}{\partial y'^2} \end{bmatrix} \times \begin{bmatrix} \delta y(x) \\ \delta y'(x) \end{bmatrix} dx.$$

Second Order Conditions for Extremum

It can be proven that a necessary condition for $y^\star$ to be a minimum for J is that $\delta^2 J[\delta y] \geq 0$ for all δy.

This is not a sufficient condition.

We say that the second variation is *strongly positive* if there exists a constant $k > 0$ such that $\delta^2 J[\delta y] \geq k\|\delta y\|^2$ for all y and δy.

A sufficient condition for $y^\star$ to be optimal for J is that its first variation vanishes and that its second variation is strongly positive.

This is not a necessary condition.

Constrained Problem

$$\text{minimize} \quad J[y] \tag{3}$$
$$\text{subject to} \quad K[y] = 0 \tag{4}$$

with variable $y \in \mathcal{C}(a, b)$, where J and K are functionals from $\mathcal{C}(a, b)$ to **R** and $\mathbf{R}^p$, respectively.

Assume that K is affine and that J is twice differentiable with a strongly positive second variation.

Sufficient Condition for Optimality

Define the Lagrangian functional $L : \mathcal{C}(a, b) \times \mathbf{R}^p \to \mathbf{R}$ as

$$L[y, \mu] = J[y] + \mu^T K[y].$$

Assume that $\bar{y} \in \mathcal{C}(a, b)$ and $\bar{\mu} \in \mathbf{R}^p$ satisfies

$$\begin{aligned} \delta L[\delta y] = \delta J[\delta y] + \bar{\mu}^T \delta K[\delta y] &= 0 \\ K[\bar{y}] &= 0. \end{aligned}$$

It then follows that $\bar{y}$ is optimal for (3–4).

Proof: By the first condition above $\bar{y}$ minimizes $L[y, \bar{\mu}]$, since L is twice differentiable with a strongly positive second variation. Now assume that $\bar{y}$ is not optimal for the above optimization problem, but that $\tilde{y}$ is optimal. Then $J[\tilde{y}] < J[\bar{y}]$ and $K[\tilde{y}] = 0$ implies that

$$L[\tilde{y}, \bar{\mu}] = J[\tilde{y}] < J[\bar{y}],$$

which contradicts that $\bar{y}$ minimizes $L[y, \mu]$.

Du Bois-Reymond Lemma

If $y : [a, b] \to \mathbf{R}$ is a continuous function and

$$\int_a^b y(x)h(x)dx = 0$$

for all $h \in \mathcal{C}(a, b)$ such that $h(a) = h(b) = 0$, then $y(x) = 0$ for all $x \in [a, b]$.

Proof: Suppose the function y is positive at some point $c \in [a, b]$. By continuity it is also positive in some interval $[x_1, x_2] \subset [a, b]$. Let

$$h(x) = (x - x_1)(x_2 - x)$$

for $x \in [x_1, x_2]$ and zero otherwise. Then it follows that

$$\int_a^b y(x)h(x)dx > 0$$

which is a contradiction. In case the function is negative at some point a similar argument can be done. The lemma still holds true if we do not constrain h at a and/or b.

Example

Consider the problem in (3–4) where

$$J[y] = \int_a^b f(y(x))dx$$
$$K[y] = \int_a^b xy(x)dx - k$$

for some constant k, where $f(y) = y \log y$. We have $f'(y) = \log y + 1$ and $f''(y) = 1/y$ and hence from (2) and from the examples the variations of J are

$$\delta J[\delta y] = \int_a^b (\log y(x) + 1)\, \delta y(x) dx$$
$$\delta^2 J[\delta y] = \frac{1}{2} \int_a^b \frac{1}{y(x)} \delta y(x)^2 dx.$$

The first variation of K is $\delta K[\delta y] = \int_a^b x\delta y(x)dx$ and the second variation of K is zero.

Example ctd.

We define the Lagrangian as $L[y, \mu] = J[y] + \mu K[y]$ and its first variation is

$$\delta L[\delta y] = \int_a^b (\log y(x) + 1 + \mu x)\, \delta y(x) dx.$$

By the du-Bosi-Reymond lemma it holds that if the first variation is zero, then

$$\log y(x) + 1 + \mu x = 0,$$

and hence we have that $y(x) = \exp(-1 - \mu x)$. The constraint $K[y] = 0$ can be used to determine μ in terms of a, b, and k.

Generalizations

Most of what has been discussed in this section generalizes to $\mathcal{D} \subseteq \mathbf{R}^n$.

We will also consider $y(x)$ to be vector valued, *i.e.*, $y : \mathcal{D} \to \mathbf{R}^n$. To this end we just define the norm as

$$\|y\| = \sup_{x \in \mathcal{D}} \|y(x)\|_2 + \sup_{x \in \mathcal{D}} \|y'(x)\|_2$$

where $\| \cdot \|_2$ is the Euclidean vector norm.

From now on $\mathcal{C}^n(a, b)$ is the normed linear space of differentiable functions $y : \mathcal{D} \to \mathbf{R}^n$ with the above norm, where $\mathcal{D} = [a, b]$.