Lecture Notes

Basic Mathematical Tools for Imaging and Visualization

Calculus of Variations Gâteaux Variation and Euler-Lagrange Basics

Darko Zikic

29 January 2007

1 Disclaimer

Dear Reader. These notes are a supplementary to the lecture *Basic Mathematical Tools for Imaging and Visualization*, offered at the TU München by Prof. Dr. Nassir Navab. The notes are *neither* complete *nor* self contained and meaningful only combined with the visit of the lecture. Comments are gladly accepted at zikic@in.tum.de.

2 Outline

In these notes we briefly present the derviation of the *Euler-Lagrange* equations for the standard model problem of minimizing the fuctional

$$I[y] = \int F(x, y, y')\ dx \quad ,$$

where y is a scalar valued function.

The *Euler-Lagrange* differential equations are derived from necessary conditions for the existence of an extremal point of a given functional. The necessary condition needed for the derivation is that the so called *Gâteaux derivative* - which presents the directional derivative of the functional - must vanish for all test functions at the extremal point. After presenting the necessary notation, we introduce the Gâteaux derivative in Section 4 and present the derivation of the Euler-Lagrange equation for the standard problem in Section 5. The presented derivation follows a book by Meyberg and Vachenauer [1].

3 Notation

- Let V be a $\mathbb{R}$ *vector space*, that is a set with defined operations $+$ and $\cdot$ and some appropriate rules.
- A *function space* is a vector space with functions as vectors. We will assume a function to be $f : \mathbb{R}^n \supseteq B \to \mathbb{R}^m$.
- A *functional* is defined as $I : D \to \mathbb{R}$ with $D \subseteq V$, where V denotes a $\mathbb{R}$ vector space. In words, a functional maps functions from a function space to real numbers.

4 Gâteaux Variation

The so called *Gâteaux derivative* presents the directional derivative of a functional in the direction of a certain test function. By this, it provides us with a possibility to define a necessary condition for an extremal point of a functional. This condition is needed for the derivation of the Euler-Lagrange equations, as presented in Section 5. In this section we show that the Gâteaux derivative, which intuitively can be seen as a directional derivative of a functional $I : D \to \mathbb{R}$ in the direction of a test function $\nu \in D$, has to equal zero for an extremal point of the functional I if defined in the following manner

$$\partial I(y;\nu) = \lim_{t \to 0} \frac{1}{t}[I(y + t\nu) - I(y)] \quad . \tag{1}$$

Given a minimization problem of a functional

$$\hat{y} = \arg\min_y I(y) \quad , \tag{2}$$

for a functional $I : D \to \mathbb{R}$.

In the following we will show that the Gâteaux derivative as defined in (1) has to vanish for an extremal point $\hat{y}$ of the functional I. We assume as known the fact that the derivative of a real-valued function must vanish for an extremum of this function.

In order to link the necessary condition for an extremal point of a functional to the necessary condition for an extremal of a real-valued function, we define a real-valued function $h : \mathbb{R} \to \mathbb{R}$

$$h(\epsilon) = I(y + \epsilon\nu) \tag{3}$$

with $y, \nu \in V$ and $\epsilon \in \mathbb{R}$. Please note that in this definition, only ϵ is a free parameter, y and ν are assumed to be fixed.

The function h has by definition the property that if we chose an extremal point of I as y (that is $y := \hat{y}$), then h will have an extremum for $\epsilon = 0$ for an arbitrary ν. Since this is important we write it once again in equation form

$$\hat{y} = \arg\min_y I(y) \quad \Rightarrow \quad h(\epsilon) = I(\hat{y} + \epsilon\nu) \text{ is extremal for } \epsilon = 0 \quad . \tag{4}$$

Since we assumed to know that for an extremum of a real-valued function, the derivative vanishes, we can state the upper equation equivalently as

$$\hat{y} = \arg\min_y I(y) \quad \Rightarrow \quad 0 = \frac{d}{d\epsilon} h(\epsilon)\bigg|_{\epsilon=0} \quad \text{for } h(\epsilon) = I(\hat{y} + \epsilon\nu) \quad . \tag{5}$$

And this equation gives us a necessary condition for the existence of the extremal point of I.

If y is an extremal point of I, then it is necessary that $0 = \frac{d}{d\epsilon} I(\hat{y} + \epsilon\nu)\big|_{\epsilon=0}$ holds.

The necessity of the condition can be seen by the negation of the upper statement: The condition $0 = \frac{d}{d\epsilon} I(\hat{y} + \epsilon\nu)\big|_{\epsilon=0}$ must hold, since otherwise y cannot be an extremal point of I. Hence, the condition $0 = \frac{d}{d\epsilon} I(\hat{y} + \epsilon\nu)\big|_{\epsilon=0}$ is necessary.

The resulting term $0 = \frac{d}{d\epsilon} I(\hat{y} + \epsilon\nu)\big|_{\epsilon=0}$ from which must vanish at an extremal point of a functional is called the *Gâteaux derivative* or *Gâteaux variation*. It can be seen as a derivative of the functional $I : D \to \mathbb{R}$ in the direction of a certain function $\nu \in D$.

We denote the Gâteaux derivative of the functional $I : D \to \mathbb{R}$ in the direction of $\nu \in D$ by

$$\partial I(y; \nu) = \frac{d}{d\epsilon} I(y + \epsilon\nu)\bigg|_{\epsilon=0} \quad . \tag{6}$$

Side note: The Gâteaux derivative presents an analogon to the directional derivative in the case of extremal values in $\mathbb{R}^n$.

It is important to remember that the vanishing of the Gâteaux derivative is only a necessary condition for an extremal point of a given functional I. It is in general not sufficient. Meaning that if $\partial I(y; \nu) = 0$ for a certain y we can only say that y is a candidate for an extremal point of I. However we can explicitly not guarantee that y is actually an extremal point of y. In practice, however, almost always only the necessary condition is enforced, and the sufficient conditions are assumed to be fulfilled.

Deriving a necessary condition for existence of an extremal point of a given functional is the most important step of this section. In the remainder of the section we derive an equivalent form for the Gâteaux derivative, that is we show that the follow equation holds.

$$\partial I(y; \nu) = \frac{d}{d\epsilon} I(y + \epsilon\nu)\bigg|_{\epsilon=0} = \lim_{t \to 0} \frac{1}{t} [I(y + t\nu) - I(y)] \quad . \tag{7}$$

The second formulation is perhaps more intuitive when interpreted as a directional derivative. The derivation is however only technical in nature and provides no further crucial results.

$$
\begin{aligned}
\frac{d}{d\epsilon}h(\epsilon)\bigg|_{\epsilon=0} &= \lim_{t\to 0}\frac{1}{t}[h(\epsilon+t)-h(\epsilon)]\bigg|_{\epsilon=0} && (8)\\
&= \lim_{t\to 0}\frac{1}{t}[I(y+(\epsilon+t)\nu)-I(y+\epsilon\nu)]\bigg|_{\epsilon=0} && (9)\\
&= \lim_{t\to 0}\frac{1}{t}[I(y+\epsilon\nu+t\nu)-I(y+\epsilon\nu)]\bigg|_{\epsilon=0} && (10)\\
&= \lim_{t\to 0}\frac{1}{t}[I(y+t\nu)-I(y)] \quad . && (11)
\end{aligned}
$$

5 Euler-Lagrange Equations

In the following we derive the Euler-Lagrange differential equation for the model problem

$$\hat{y} = \arg\min_y I[y] = \int_a^b F(x,y,y')\ dx \quad ,$$

Here y is a real-valued function from the domain D, and the functional is defined as $I : D \to \mathbb{R}$. The domain D is defined as

$$D = \left\{ y \in C^2[a,b] \mid y(a) = y_a,\ y(b) = y_b,\ a,b \in \mathbb{R} \right\} \quad . \tag{12}$$

Please note that the definition of the domain D encodes boundary functions for admissible functions y. The first and second derivative of y are denoted by y' and y'' respectively.

We will show that using the fact that the Gâteaux derivative vanishes at an extremal point we can derive from the model problem the following equation

$$0 = \frac{\partial F}{\partial y}(x,y,y') - \frac{d}{dx}\frac{\partial F}{\partial y'}(x,y,y') \quad . \tag{13}$$

This equation must hold for $y = \hat{y}$, that is, an extremal point of the model problem.

We define a function $y_\epsilon \in$, based on the assumption that the extremal $\hat{y}$ is known as

$$y_\epsilon = \hat{y} + \epsilon\nu \quad , \tag{14}$$

with real-valued $\nu \in C^2[a,b]$ with $\nu(a) = \nu(b) = 0$ and $\epsilon \in \mathbb{R}$.

Using the fact that for $\hat{y}$ the Gâteaux derivative must be zero, that is

$0 = \partial I(\hat{y};\nu)$, we get the following derivation

$$
\begin{aligned}
0 &= \partial I(\hat{y};\nu) && (15)\\
&= \frac{d}{d\epsilon}\int_a^b F(x, y_\epsilon, y'_\epsilon)\ dx\Big|_{\epsilon=0} && (16)\\
&= \int_a^b \frac{d}{d\epsilon}F(x, y_\epsilon, y'_\epsilon)\ dx\Big|_{\epsilon=0} && (17)\\
&= \int_a^b \frac{\partial}{\partial y}F(x, y_\epsilon, y'_\epsilon)\nu(x) + \frac{\partial}{\partial y'}F(x, y_\epsilon, y'_\epsilon)\nu'(x)\ dx\Big|_{\epsilon=0} && (18)\\
&= \int_a^b \frac{\partial}{\partial y}F(x, \hat{y}, \hat{y}')\nu(x) + \underbrace{\frac{\partial}{\partial y'}F(x, \hat{y}, \hat{y}')\nu'(x)}_{*}\ dx\ . && (19)
\end{aligned}
$$

Side note: the step (18) is obtained by the application of the chain rule, from (18) to (19) we get by evaluating for $\epsilon = 0$. In order to get rid of the term ν' we take a closer look at the (*) term and perform integration by parts. We get

$$
\begin{aligned}
\int_a^b \frac{\partial}{\partial y'}F(x, \hat{y}, \hat{y}')\nu(x)'\ dx &= \left[\nu(x)\frac{\partial}{\partial y'}F(x, \hat{y}, \hat{y}')\right]_a^b\\
&\quad - \int_a^b \nu(x)\frac{d}{dx}\frac{\partial}{\partial y'}F(x, \hat{y}, \hat{y}')\ dx && (20)\\
&= -\int_a^b \nu(x)\frac{d}{dx}\frac{\partial}{\partial y'}F(x, \hat{y}, \hat{y}')\ dx && (21)
\end{aligned}
$$

Side note: The first term vanishes from step (20) to (21) because the first term in the equations evaluates to zero because of the boundary condition imposed on ν that $\nu(a) = \nu(b) = 0$.

By substituting (21) in (19) we get

$$
0 = \int_a^b \left[\frac{\partial}{\partial y}F(x, \hat{y}, \hat{y}') - \frac{d}{dx}\frac{\partial}{\partial y'}F(x, \hat{y}, \hat{y}')\right]\nu(x)\ dx\ . \qquad (22)
$$

And finally, using the Theorem (5.1), we get the Euler-Lagrange equation for the model problem

$$
0 = \frac{\partial}{\partial y}F(x, \hat{y}, \hat{y}') - \frac{d}{dx}\frac{\partial}{\partial y'}F(x, \hat{y}, \hat{y}')\ . \qquad (23)
$$

This differential equation presents a necessary condition for an extremal point $\hat{y}$ of the functional from the model problem.

By remembering that y and y' are functions of x, applying the chain rule to

the second term, dropping the arguments for F we get the alternative notation

$$0 = \frac{\partial}{\partial y} F(x, \hat{y}, \hat{y}') - \frac{d}{dx} \frac{\partial}{\partial y'} F(x, \hat{y}, \hat{y}') \tag{24}$$

$$= \frac{\partial}{\partial y} F - \left[\frac{d}{dx} \frac{\partial}{\partial y'} F + \frac{\partial}{\partial y} \frac{\partial}{\partial y'} F \cdot y' + \frac{\partial}{\partial y'} \frac{\partial}{\partial y'} F \cdot y'' \right] \quad . \tag{25}$$

Theorem 5.1 *Let $f : [a, b] \to \mathbb{R}$ be continuous. If for an arbitrary $g \in C^2[a, b]$ with $g(a) = g(b) = 0$ it holds that*

$$0 = \int_a^b f(x) g(x) \; dx \quad , \tag{26}$$

then it follows that

$$f = 0 \quad . \tag{27}$$

The proof for the theorem is straight forward and can be done by assuming that $f \neq 0$. Then, it is easy to find a g, such that $\int_a^b f(x)g(x) \; dx \neq 0$, which presents a contradiction to the basic assumption of the theorem. If $f \in C^2[a, b]$ with $f(a) = f(b) = 0$ and $f \neq 0$ in a certain subinterval of $[a, b]$, then f itself is a suitable candidate for g, such that $\int_a^b f(x)g(x) \; dx \neq 0$.

References

[1] Kurt Meyberg and Peter Vachenauer. *Höhere Mathematik 2*. Springer, 4 edition, 2001.