II. Variation of functionals

The derivative of a function being zero is a necessary condition for the extremum of a function in ordinary calculus. Let us now tackle the question of the equivalent of a derivative for functionals because it plays the same crucial role in calculus of variations as the derivative of the ordinary calculus in minimization of functions. Let us begin with a simple but a very important concept called a <u>Gâteaux variation</u>.

Gâteaux variation

The functional $\delta J(x)$ is called the Gâteaux variation of J at x when the limit that is defined as follows exists.

$$\delta J(x;h) = \lim_{\varepsilon \to 0} \frac{J(x+\varepsilon h) - J(x)}{\varepsilon} \text{ where } h \text{ is any vector in } X.$$

Let us look at the meaning of h and ε geometrically. Note that $x, h \in X$. Now, since x is the unknown function to be found so as to minimize (or maximize) a functional, we want to see what happens to the functional $J(x)$ when we perturb this function slightly. For this, we take another function h and multiply it by a small number ε. We add εh to x and look at the value of $J(x+\varepsilon h)$. That is, we look at the perturbed value of the functional due to perturbation εh. This is the shaded area shown in Fig. 1 where the function x indicated by a thick solid line, h by a thin solid line, and $x + \varepsilon h$ by a thick dashed line. Next, we think of the situation of $\in$ tending to zero. As $\in \to 0$, we consider the limit of the shaded area divided by $\in$. If this limit exists, such a limit is called the Gâteaux variation of $J(x)$ at x for an arbitrary but fixed vector h. Note that, we denote it as $\delta J(x;h)$.
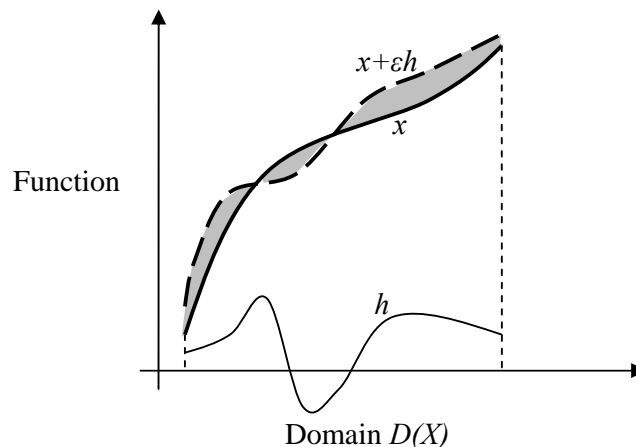

Figure 1. Pictorial depiction of variation εh of a function x

Although the most important developments in calculus of variations happened in 17th and 18th centuries, this formalistic concept of variation was put forth by a French mathematician Gâteaux around the time of the first world war. So, one can say that intuitive and creative thinking leads to new developments and rigorous thinking makes them mathematically sound and completely unambiguous. To reinforce our understanding of the Gâteaux variation

defined as above, let us relate it to the concept of a <u>directional derivative</u> in multi-variable calculus.

A directional derivative of the function $f(x_1, x_2, ..., x_n)$ denoted in a compact form as $f(\bar{x})$ in the direction of a given vector $\bar{h}$ is given by

$$\lim_{\varepsilon \to 0} \frac{f(\bar{x} + \varepsilon \bar{h}) - f(\bar{x})}{\varepsilon}.$$

Here the "vector" is the usual notion that you know and not the extended notion of a "vector" in a vector space. We are using the overbar to indicate that the denoted quantity consist of several elements in an array as in a column (or row) vector. You know how to take the derivative of a function $f(\bar{x})$ with respect to any of its variables, say $x_i, 1 \le i \le n$. It is simply a partial derivative of $f(\bar{x})$ with respect to x_i. You also know that this partial derivative indicates the rate of change of $f(\bar{x})$ in the direction of x_i. What if you want to know the rate of change of $f(\bar{x})$ in some arbitrary direction denoted by $\bar{h}$? This is exactly what a directional derivative gives.

Now, relate the concept of the directional derivative to Gâteaux variation because we want to know how the value of the functional changes in a "direction" of another element h in the vector space. Thus, the Gateaux variation extends the concept of the directional derivative concept of finite multi-variable calculus to infinite dimensional vector spaces, i.e., calculus of functionals.

Gâteaux differentiability

If Gateaux variation exists for all $h \in X$ then J is said to be Gateaux differentiable.
Operationally useful definition of Gâteaux variation

Gateaux variation can be thought of as the following ordinary derivative evaluated at $\varepsilon = 0$

$$\delta J(x;h) = \frac{d}{d\varepsilon} J(x + \varepsilon h) \Big|_{\varepsilon=0}$$

This helps calculate the Gâteaux variation easily by taking an ordinary derivative instead of evaluating the limit as in the earlier formal definition. Note that this definition follows from the earlier definition and the concept of how an ordinary derivative is defined in ordinary calculus if we think of the functional as a simple function of ε.

Gâteaux variation and the necessary condition for minimization of a functional

Gâteaux variation provides a necessary condition for a minimum of a functional.

Consider   where $J(x)$, $x \in D$, is an open subset of a normed vector space X and $x^* \in D$ and any fixed vector $h \in X$

If x^* is a minimum, then

$$J\left(x^* + \varepsilon h\right) - J\left(x^*\right) \geq 0$$

must hold for all sufficiently small ε

Now,
for $\varepsilon \geq 0$

$$\frac{J\left(x^* + \varepsilon h\right) - J\left(x^*\right)}{\varepsilon} \geq 0$$

and for $\varepsilon \leq 0$

$$\frac{J\left(x^* + \varepsilon h\right) - J\left(x^*\right)}{\varepsilon} \leq 0$$

If we let $\varepsilon \to 0$,

$$\left. \begin{array}{c} \lim\limits_{\substack{\varepsilon \to 0 \\ \varepsilon > 0}} \dfrac{J\left(x^* + \varepsilon h\right) - J\left(x^*\right)}{\varepsilon} \geq 0 \\[2em] \text{and} \quad \lim\limits_{\substack{\varepsilon \to 0 \\ \varepsilon < 0}} \dfrac{J\left(x^* + \varepsilon h\right) - J\left(x^*\right)}{\varepsilon} \leq 0 \end{array} \right\} \quad \underbrace{\lim\limits_{\varepsilon \to 0} \dfrac{J\left(x^* + \varepsilon h\right) - J\left(x^*\right)}{\varepsilon}}_{\substack{\text{existence of Gâteaux variation} \\ \text{ensures the existence of this limit}}} = \delta J(x;h) = 0$$

This simple derivation proves that the Gâteaux variation being zero is the necessary condition for the minimum of a functional. Likewise we can show (by simply reversing the inequality signs in the above derivation) that the same necessary condition applies to maximum of a functional.

Now, we can state this as a theorem since it is a very important result.
Theorem: necessary condition for a minimum of a functional

$$\delta J\left(x^*; h\right) = 0 \quad \text{for all } h \in X$$

Based on the foregoing, we note that the Gâteaux variation is very useful in the minimization of a functional but the existence of Gateaux variation is a weak requirement on a functional since this variation does not use a norm in X. Thus, it is not directly related to the continuity of a functional. For this purpose, another differential called <u>Fréchet differential</u> has been put forth.

Frechet differential

$$\lim_{\|h\| \to 0} \frac{\left\| J\left(x+h\right) - J\left(x\right) - dJ\left(x;h\right) \right\|}{\|h\|} = 0$$

If the above condition holds and $dJ\left(x;h\right)$ is a linear, continuous functional of h, then J is said to be Fréchet differentiable at x with "increment" h.

$dJ(x;h)$ is called the Fréchet differential.

If J is differentiable at each $x \in D$ we say that J is Fréchet differentiable in D.

Some properties of Fréchet differential

i) $J(x+h) = J(x) + dJ(x;h) + E(x;h)\|h\|$ for any small non-zero $h \in X$ has a limit zero at the zero vector in X. That is,

$$\lim_{\substack{h \to 0 \text{ in } X}} E(x;h) = 0.$$

Based on this, sometimes the Fréchet differential is also defined as follows.

$$\lim_{h \to 0} \frac{\|J(x+h) - J(x) - dJ(x;h)\|}{\|h\|} = 0.$$

ii) $dJ(x; a_1 h_1 + a_2 h_2) = a_1 dJ(x;h_1) + a_2 dJ(x;h_2)$ must hold for any numbers $a_1, a_2 \in K$ and any $h_1, h_2 \in X$.

This is simply the linearity requirement on the Fréchet differential.

iii) $dJ(x;h) \leq \text{constant } \|h\|$ for all $h \in X$

This is the continuity requirement on the Fréchet differential.

iv) $|dJ(x;h)| = \underbrace{J'(x)}_{\substack{\text{Frechet} \\ \text{derivative}}} h$

This is to say that the Fréchet differential is a linear functional of h. Note that it also introduces a new definition: Fréchet derivative, which is simply the coefficient of h in the Fréchet differential.

Relationship between Gâteaux variation and Fréchet differential

If a functional J is Fréchet differentiable at x then the Gateaux variation of J at x exists and is equal to the frechet differential. That is,

$$\delta J(x;h) = dJ(x;h) \quad \text{for all} \quad h \in X$$

<u>Here is why:</u>

Due to the linearity property of $dJ(x;h)$, we can write

$$dJ(x;\varepsilon h) = \varepsilon dJ(x;h)$$

Substituting the above result into property (i) of the Fréchet differential noted earlier, we get

$$J\left(x+\varepsilon h\right)-J\left(x\right)-\varepsilon dJ\left(x;h\right)=E\left(x,\varepsilon h\right)\left\|h\right\|\left|\varepsilon\right| \quad \text{for any } h \in X$$

A small rearrangement of terms yields

$$\frac{J\left(x+\varepsilon h\right)-J\left(x\right)}{\varepsilon}=dJ\left(x;h\right)+E\left(x,\varepsilon h\right)\left\|h\right\|\frac{\left|\varepsilon\right|}{\varepsilon}$$

When limit $\varepsilon \to 0$ is taken, the above equation gives what we need to prove:

$$\lim_{\varepsilon\to0}\frac{J\left(x+\varepsilon h\right)-J\left(x\right)}{\varepsilon}=\delta J\left(x;h\right)=dJ\left(x;h\right) \quad \text{because} \quad \lim_{\varepsilon\to0}E\left(x,\varepsilon h\right)\left\|h\right\|\frac{\left|\varepsilon\right|}{\varepsilon}=0$$

Note that the latter part of property (i) is once again used above.

Operations using Gateaux variation

Consider a simple general functional of the form shown below.

$$J\left(y\right)=\int_{x_1}^{x_2}F\left(x,\ y\left(x\right),\ y'\left(x\right)\right)\ dx$$

where $y'\left(x\right)=\dfrac{dy}{dx}$

Note our sudden change of using x. It is no longer a member (element, vector) of a normed vector space X. It is now an independent variable and defines the domain of $y(x)$, which is a member of a normed vector space. Now, $y(x)$ is the unknown function using which the functional is defined.

If we want to calculate the Gâteaux variation of the above functional, instead of using the formal definition that needs an evaluation of the limit we should use the alternate operationally useful definition—taking the ordinary derivative of $J\left(y+\varepsilon h\right)$ with respect to ε and evaluating at $\varepsilon=0$. In fact, there is even easier route that almost like a thumb-rule. Let us find that by using the derivative approach for the above simple functional.

$$J\left(y+\varepsilon h\right)=\int_{x_1}^{x_2}F\left(x,\ y\left(x\right)+\varepsilon h\left(x\right),\ y'\left(x\right)+\varepsilon h'\left(x\right)\right)\ dx$$

Recalling that $\delta J\left(x;h\right)=\dfrac{d}{d\varepsilon}J\left(x+\varepsilon h\right)\bigg|_{\varepsilon=0}$, we can write

$$\frac{d}{d\varepsilon}J(x+\varepsilon h) = \frac{d}{d\varepsilon}\left\{\int_{x_1}^{x_2} F(x,\ y+\varepsilon h,\ y'+\varepsilon h')\ dx\right\}$$

$$= \int_{x_1}^{x_2}\frac{\partial}{\partial\varepsilon}\left\{F(x,\ y+\varepsilon h,\ y'+\varepsilon h')\right\}\ dx$$

Please note that the order of differentiation and integration have been switched above. It is a legitimate operation. By using chain-rule of differentiation for the integrand of the above functional, we can further simplify it as to obtain

$$\delta J(x;h) = \int_{x_1}^{x_2}\left(\frac{\partial F}{\partial(y+\varepsilon h)}h + \frac{\partial F}{\partial(y'+\varepsilon h')}h'\right)\Bigg|_{\varepsilon=0} = \int_{x_1}^{x_2}\left(\frac{\partial F}{\partial y}h + \frac{\partial F}{\partial y'}h'\right)dx\ .$$

What we have obtained above is a general result in that for any functional, be it of the form $J(x,y,y',y'',y''',\cdots)$, we can write the variation as follows.

$$\delta J(x;h) = \int_{x_1}^{x_2} F(x,y,y',y'',y''',\cdots)dx = \int_{x_1}^{x_2}\left(\frac{\partial F}{\partial y}h + \frac{\partial F}{\partial y'}h' + \frac{\partial F}{\partial y''}h'' + \frac{\partial F}{\partial y'''}h''' + \cdots\right)dx\ .$$

Note that in taking partial derivatives with respect to y and its derivatives we treat them as independent. It is a thumb-rule that enables us to write the variation rather easily by inspection and using rules of partial differentiation of ordinary calculus.

We have now laid the necessary mathematical foundation for deriving the Euler-Lagrange equations that are the necessary conditions for the extremum of a function. Note that the Gâteaux variation still has an arbitrary function h. When we get rid of this, we get the Euler-Lagrange equations. For that we need to talk about fundamental lemmas of calculus of variations.