First Variation of a Functional

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

Gateaux variation

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

$$\delta J(\mathrm{f};h) = \lim_{\varepsilon \to 0} \frac{J(f+\varepsilon h) - J(x)}{\varepsilon}$$ where h is any vector in a vector space (a function space for our purposes), $\mathbb{F}$.

Let us look at the meaning of h and ε geometrically. Note that $\mathrm{f}, h \in \mathbb{F}$. Now, since f is the unknown function to be found so as to minimize (or maximize) a functional, we want to see what happens to the functional $J(f)$ 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 f and look at the value of $J(f+\varepsilon h)$. That is, we look at the perturbed value of the functional due to perturbation εh. Symbolically, this is the shaded area shown in Fig. 1 where the function f is indicated by a thick solid line, h by a thin solid line, and $f+\varepsilon h$ by a thick dashed line. Next, we think of the situation of ε tending to zero. As $\varepsilon \to 0$, we consider the limit of the shaded area divided by ε. If this limit exists, such a limit is called the Gâteaux variation of $J(f)$

at f for an arbitrary but fixed vector h. Note that, we denote it as $\delta J(f;h)$ by including h in defining Gateaux variation.
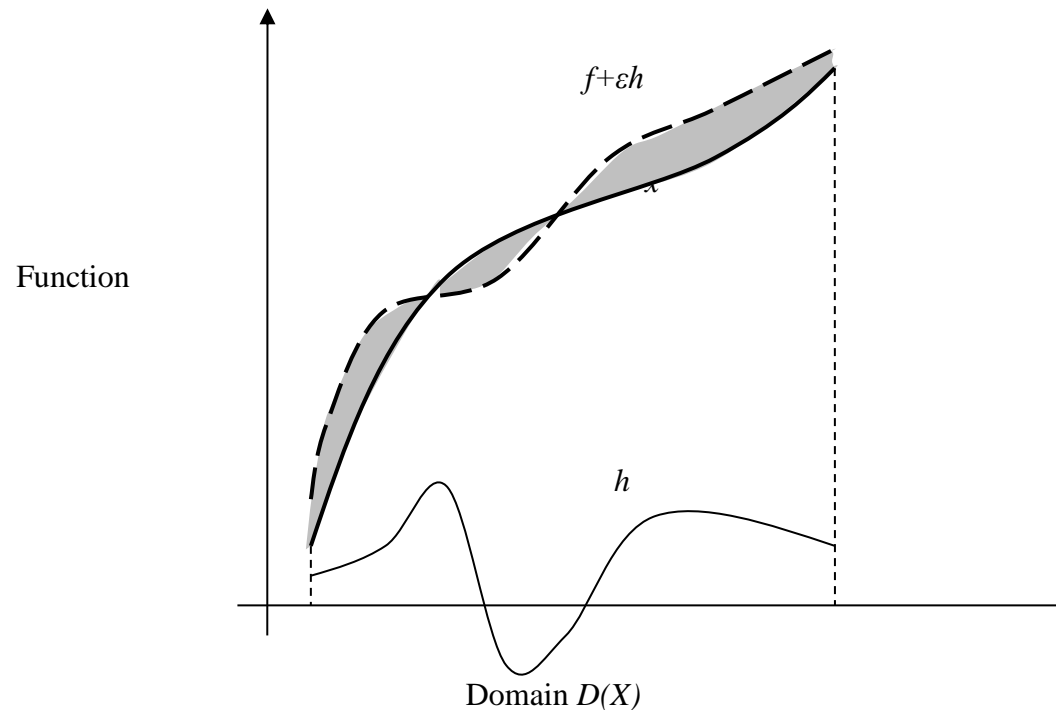

Figure 1. Pictorial depiction of variation εh of a function f

Although the most important developments in calculus of variations happened in 17$^{\text{th}}$ and 18$^{\text{th}}$ centuries, this formalistic concept of variation was put forth by a French mathematician Gateaux at the end of 19$^{\text{th}}$ century. 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 Gateaux variation, let us relate it to the concept of a <u>directional derivative</u> in multi-variable calculus.

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

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

Here the "vector" is the usual notion of "mechanics vector" that you know and not the extended notion of a "vector" in a vector space. We are using the over-bar to indicate that the denoted quantity consists 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 $x_i, 1 \leq i \leq n$. 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. Indeed, $\nabla_{\bar{h}} f(\bar{x}) = \nabla_{\bar{h}} f(\bar{x}) \cdot \bar{h} = \nabla_{\bar{h}} f(\bar{x})^T \bar{h}$. That is, the component of the gradient in the direction of $\bar{h}$.

Now, relate the concept of the directional derivative to Gateaux 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 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 Gateaux variation

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

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

This helps calculate the Gateaux 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.

Gateaux variation and the necessary condition for minimization of a functional

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

Consider $J(f)$ where $J(f)$, $f \in D$, is an open subset of a normed vector space $\mathbb{F}$, and $f^* \in D$ and any fixed vector $h \in \mathbb{F}$.

If f^* is a minimum, then

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

must hold for all sufficiently small ε

Now,

for $\varepsilon \geq 0$

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

and for $\varepsilon \leq 0$

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

If we let $\varepsilon \to 0$,

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

This simple derivation proves that the Gateaux 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

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

Based on the foregoing, we note that Gateaux 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. Without a norm, we cannot talk about continuity of a functional

because we cannot judge how close two functions are to each other. Thus, Gateaux variation is not directly related to the continuity of a functional. For this purpose, another differential called <u>Fréchet differential</u> has been put forth.

<mark>Fréchet differential</mark>

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

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

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

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

Some properties of Fréchet differential

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

$$\lim_{h \to \theta} E(f;h) = 0.$$

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

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

ii) $dJ(f; a_1 h_1 + a_2 h_2) = a_1 dJ(f; h_1) + a_2 dJ(f; h_2)$ must hold for any numbers $a_1, a_2 \in K$ and any $h_1, h_2 \in \mathbb{F}$.

This is simply the linearity requirement on the Fréchet differential.

iii) $dJ(f; h) \le c \|h\|$ for all $h \in \mathbb{F}$, where c is a constant.

This is the continuity requirement on the Fréchet differential.

iv) $\left| dJ\left(f;h\right) \right| = J'\left(f\right) h$

Frechet
derivative

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 Gateaux variation and Fréchet differential

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

$$\delta J\left(f;h\right) = dJ\left(f;h\right) \quad \text{for all} \quad h \in \mathbb{F}$$

<u>Here is why:</u>

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

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

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

$$J(f + \varepsilon h) - J(f) - \varepsilon dJ(f;h) = E(f,\varepsilon h)\|h\|\,|\varepsilon| \qquad \text{for any } h \in \mathbb{F}$$

A small rearrangement of terms yields

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

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

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

Note that the latter part of property (i) is once again used in the preceding equation.

Operations using Gateaux variation

Consider a simple general functional of the form shown below.

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

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

Here, $y(x)$ is the unknown function using which the functional is defined. It takes the role of f that we have used so far. We need to have our wits about us to see which symbol is used in what way!

If we want to calculate the Gateaux 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 $J(y + \varepsilon h)$ with respect to ε and evaluating at $\varepsilon = 0$. In

fact, there is an easier route that is almost like a thumb-rule. Let us find that by using the derivative approach for the above simple functional.

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

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

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

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

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 to obtain

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

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

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

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 Gateaux 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.

Variational Derivative

We have studied Gateaux variation and Fréchet differential and the relationship between them. There is one more subtle variant of this, which is called the <u>variational derivative</u>. It is useful in some applications and in proving some theorems. More importantly, it tells us an alternative way of looking at the concept of variation based purely on the techniques of ordinary calculus. In fact, it can be interpreted as the <mark>"partial derivative" equivalent for calculus of variations</mark>. As the history goes, Euler had apparently derived his eponymous necessary condition using this concept. Let us begin with the notation. The variational derivative of a functional $J = \int_{x_0}^{x_f} F(x, y, y') dx$ is denoted as $\dfrac{\delta J}{\delta y}$ and is given by

$$\frac{\delta J}{\delta y} = F_y - \frac{d}{dx}\left(\frac{\partial F}{\partial y'} \right) \qquad\qquad (1)$$

You may observe that it is nothing but the Euler-Lagrange expression that should be zero. When J has a more general form, the expression for $\dfrac{\delta J}{\delta y}$ will be the corresponding expression in the E-L equation that we equate to zero. Let us see what rationale underlies this definition.

Because we want to use only the techniques of ordinary calculus, let us "discretize" $y(x)$ and consider finitely many discrete points $x_k \left(k = 1, 2, ..., N \right)$ within the interval $\left(x_0, x_f \right)$. See Fig. 1. As can be seen in this figure, by way of discretization, we are approximating the continuous curve of $y(x)$ by a polygon.
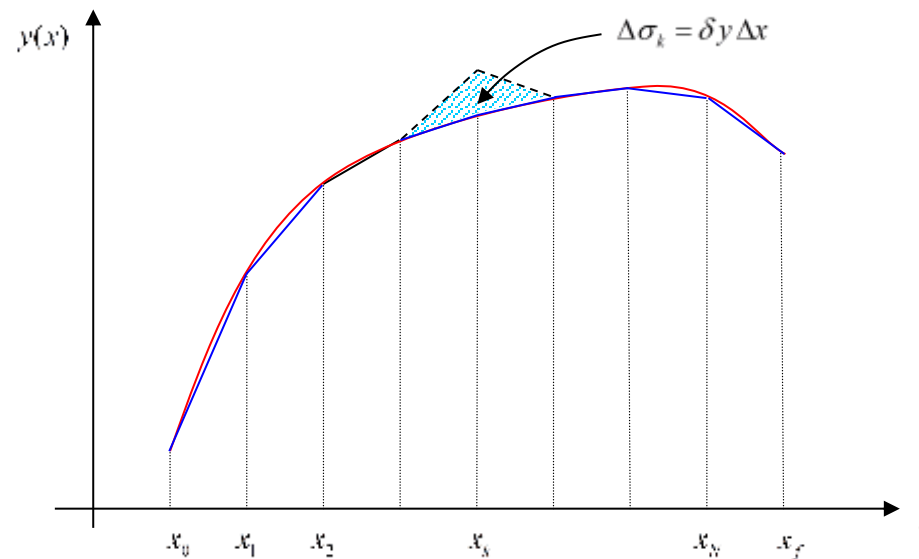
Figure 1. Discretization of a continuous curve $y(x)$ by a polygon. All subdivisions on the x-axis are equal to Δx. A local perturbation at x_k is considered and its effect is shown with the dashed lines.

Now, the functional can be approximated as follows.

$$\sum_{k=1}^{N} \quad J \approx J_N = \sum_{k=1}^{N} F\left(x_k, y_k, \frac{(y_{k+1} - y_k)}{(x_{k+1} - x_k)}\right)(x_{k+1} - x_k) = \sum_{k=1}^{N} F\left(x_k, y_k, \frac{(y_{k+1} - y_k)}{\Delta x}\right)\Delta x \quad (2)$$

where in the last step we have assumed that all subdivisions along the x-axis are equal to Δx. Our variables to minimize J_N are now $\{y_1, y_2, \cdots, y_N\}$. Consider the partial derivative of J_N with respect to y_k.

$$\frac{\partial J_N}{\partial y_k} = F_y(x_k, y_k, \frac{(y_{k+1} - y_k)}{\mathrm{D}x})\mathrm{D}x + F_{y'}(x_{k-1}, y_{k-1}, \frac{(y_k - y_{k-1})}{\mathrm{D}x}) - F_{y'}(x_k, y_k, \frac{(y_{k+1} - y_k)}{\mathrm{D}x})$$

(3)

Here, we have just used the chain rule of differentiation. As $\Delta x \to 0$, the RHS of Eq. (3) goes to zero. Now, divide the LHS and RHS of Eq. (3) by Δx to get

$$\frac{\partial J_N}{\partial y_k \Delta x} = F_y(x_k, y_k, \frac{(y_{k+1} - y_k)}{\Delta x}) + \frac{F_{y'}(x_{k-1}, y_{k-1}, \frac{(y_k - y_{k-1})}{\Delta x}) - F_{y'}(x_k, y_k, \frac{(y_{k+1} - y_k)}{\Delta x})}{\Delta x}$$

(4)

When $\Delta x \to 0$, $\partial y_k \Delta x$, which can be interpreted as the shaded area in Fig. 1, also tends to zero. In fact, we then denote $\partial y_k \Delta x$ as $\Delta \sigma_k$ or, in general, simply as δy evaluated at $x = x_k$. Furthermore, as $\Delta x \to 0$, $J_N \to J$. We take the limit of Eq. (4) as $\Delta x \to 0$.

$$\lim_{\Delta x \to 0} \frac{\partial J_N}{\partial y_k \Delta x} = \frac{\delta J}{\delta y} = F_y - \frac{d}{dx}\left(F_{y'}\right) \qquad (5)$$

Notice how we defined the variational derivative in Eq. (5). We can think of $\dfrac{\delta J}{\delta y}$ as the limiting case of $\dfrac{J(y+h) - J(y)}{\Delta \sigma}$ where h is the perturbation (i.e., variation) of y at some $\hat{x}$ and $\Delta \sigma$ is the extra area under $y(x)$ due to that perturbation. Therefore, we write

$$\Delta J = J(y+h) - J(y) = \left\{ \frac{\delta J}{\delta y}\bigg|_{x=\hat{x}} + \varepsilon \right\} \Delta \sigma \qquad (6)$$

where ε is a small discretization error. When the discretization error is insignificantly small, we can write

$$\Delta J \approx \frac{\delta J}{\delta y}\bigg|_{x=\hat{x}} \Delta \sigma \qquad (7)$$

Thus, the variational derivative helps us get the first order change in the value of the functional for a local perturbation of $y(x)$ at $x = \hat{x}$. Think of Taylor series of expansion of a function of many variables and try to relate this concept of first order change in the value of the function.