Function spaces and variation of functionals

Banach space

A complete normed vector space is called a *Banach space*.

A normed vector space X is *complete* if every *Cauchy sequence* from X has a limit in X

A sequence $\{x_n\}$ in a normed vector space is said to be *Cauchy* (or fundamental) *sequence* if $\|x_n - x_m\| \to 0 \quad$ as $n, m \to \infty$

In other words, given $\varepsilon > 0$ there is an integer N such that $\|x_n - x_m\| < \varepsilon$ for all $m, n > N$

$x \in X$ is called a *limit* of a convergent sequence $\{x_n\}$ in a normed vector space if the sequence $\{\|x - x_n\|\}$ converges to zero. In other words, $\lim_{n\to\infty} \|x - x_n\| = 0$.

Verifying if a given normed vector space is a Banach space requires an investigation into the limit of all Cauchy sequences. This needs tools of *real analysis*. We are not going to discuss them here. But let us try to relate to these sequences from a practical viewpoint and why we should worry about them.

In the context of structural optimization, we can imagine the *sequences* (that may or may not be Cauchy sequences) as candidate designs that we obtain in a sequence in iterative numerical optimization. As you may be aware any numerical optimization technique needs an initial guess which is improved in each iteration. Thus, we get a sequence of "vectors" (functions in our study). Whether such a sequence converges at all, or converges to a limit within the space we are concerned with, are important practical questions. The abstract notion of a complete normed vector space helps us in this regard. So, it is useful to know the properties of a function space that we are dealing with. It is one way of knowing if numerical optimization would converge to a limit, which will be our optimal solution.

Hilbert space

A complete inner product space is called a *Hilbert space*.

An *inner product space* (or *pre-Hilbert space*) is a vector space X with an inner product defined on it.

An *inner product* on a vector space X is a mapping $X \times X$ into a scalar field K of X denoted as $\langle x, y \rangle, \quad x, y \in X$ and satisfies the following properties:

(i) $\langle x+y,z\rangle = \langle x,z\rangle + \langle y,z\rangle$

(ii) $\langle \alpha x, y\rangle = \alpha \langle x,y\rangle$

(iii) $\langle x,y\rangle = \overline{\langle y,x\rangle}$ The over bar denotes conjugation and is not necessary if x, y are real.

(iv) $\langle x,x\rangle \geq 0$ and

$\langle x,x\rangle = 0$ if and only if $x=\theta$

Note the following relationship between a norm and an inner product.

$$\|x\| = \sqrt{\langle x,x\rangle}$$

Note also the relationship between a metric and an inner product.

$$d(x,y) = \|x-y\| = \sqrt{\langle x-y, x-y\rangle}$$

As an example, for $C^0[a,b]$, the norm and inner product defined as follows.

$$\|x\| = \sqrt{\int_a^b x^2(t)\,dt} = \sqrt{\langle x,x\rangle}$$

$$\langle x,y\rangle = \int_a^b x(t)y(t)\,dt$$

Thus, inner product spaces are normed vector spaces. Likewise, Hilbert spaces are Banach spaces.

Normed vector spaces give us the tools for algebraic operations to be performed on vector spaces because we have the notion of how close things ("vectors") are to each other by way of norm. Inner product spaces enable us to do more—study the geometric aspects. As an example, consider that orthogonality (or perpendicularity) or lack of it is easily noticeable from the inner product.

For $x, y \in X$, if $\langle x,y\rangle$=0 , then x is said to be orthogonal to y

Banach and Hilbert spaces are classes of useful function spaces (again remember that a function space is only one type of the more general concept of a vector space). There are also some specific function spaces that we should be familiar with as they are the spaces to which the design spaces that we consider in structural optimization actually belong.

Lebesgue space

A *Lebesgue space* defined as below is a Banach space.

$$L^q(\Omega) = \left\{ v : v \text{ is defined on } \Omega \text{ and } \|v\|_{L^q(\Omega)} < \infty \right\} \text{ where } \|v\|_{L^q(\Omega)} = \left(\int_\Omega \left| v(x)^q \right| dx \right)^{1/q} \qquad 1 \leq q \leq \infty$$

The case of $q=2$ gives $L^2(\Omega)$ consisting of all square-integrable functions. The integration of square of a function is important for us as it often gives the energy of some kind. Think of kinetic energy which is a scalar multiple of the square of the velocity. On many occasions, we also have other energies (usually potential energies or strain energies) that are squares of derivatives of functions. This gives us a number of energy spaces. The Sobolev space gives us exactly that.

Sobolev space

$$W^{r,q}(\Omega)=\left\{v\in L^1(\Omega):\|v\|_{W^{r,q}(\Omega)}<\infty\right\},\qquad 1\le q\le\infty$$

where

$$\left.\|v\|_{W^{r,q}(\Omega)}=\left(\sum_{|\alpha|\le r}\|D^\alpha v\|_{L^q(\Omega)}^q\right)^{1/q}\right\}\text{ is the } \textit{Sobolev norm}$$

$$L^1(\Omega)=\left\{v: v\in L^1(K)\text{ for any compact } K \text{ inside } \Omega\right\}$$

D^α used above denoted the derivative of order α. Sobolev space is a Banach space.

Note: We have used the qualifying word "compact" for K above. A closed and bounded set is called a compact set. We will spare us from the definitions of closedness and boundedness of a set because we have already deviated from our main objective of knowing what a functional is. Let us return to functionals now.

We have defined a functional as a particular case of an operator whose range is a real (or complex) number set. Let us also consider another definition which says the same thing but in a different way as we have talked much about vector spaces and fields.

Functional—another definition

A *functional* f is a transformation from a vector space to its coefficient field $f: X\to K$.

Let us now look at certain types of functionals that are of main interest to us.

A *linear functional* is one for which
$f(x+y)=f(x)+f(y)$ for all $x,y\in X$ and $f(\alpha x)=\alpha f(x)$ for all $\alpha\in K,\ x\in X$ hold good. Some people write the above two linearity properties as a single property as follows.
$f(\alpha x+\beta y)=\alpha f(x)+\beta f(y)$ for all $x,y\in X;\ \alpha,\beta\in K$

A definite integral is a linear functional. We will deal with a lot of definite integrals in calculus of variations as well as variational methods and structural optimization.

A *bounded functional* is one when there exists a real number c such that $|f(x)|\le c\|x\|$ where $|\cdot|$ is the norm in K; $\|\cdot\|$ is the norm in X.

Continuous functional

Now, we have discussed in which function spaces our functions reside. In calculus of variations, our unknowns are functions. Our objective is a functional. Just as in ordinary finite-variable optimization, in calculus of variations too we need to take derivatives of functionals. What is the equivalent of a derivative for a functional? Before we define such a thing, we need to understand the concept of continuity for a functional. We do that next.

A functional J is said to be continuous at x in D (an open set in a given normed vector space X) if J has the limit $J(x)$ at x . Or symbolically, $\lim_{y \to x \in X} J(y) = J(x)$.

J is said to be *continuous* on D if J is continuous at each vector in D

J has the limit L at x if for every positive number ε there is a ball $B_r(x)$ (with radius r) contained in D such that $\left|L - J(y)\right| < \varepsilon$ for all $y \in B_r(x)$. Or symbolically, $\lim_{y \to x \in X} J(y) = L$.

Since 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. Let us begin with a simple but very important concept called a Gâteaux variation.

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 $\delta J(x;h)$.

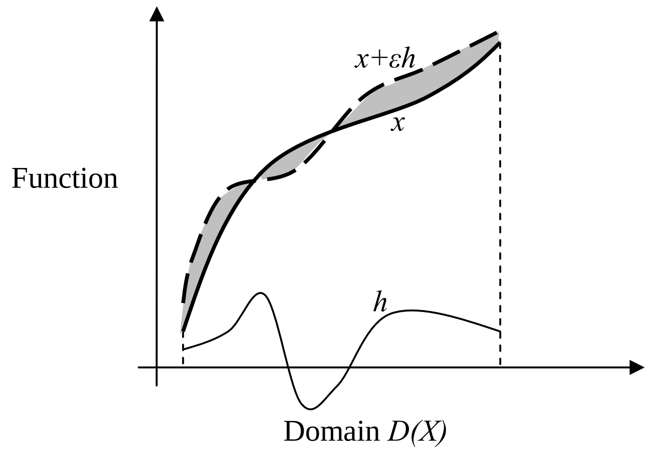

Figure 1. Pictorial depiction of variation εh of a function x

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 directional derivative in multi-variable calculus.

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

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

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 $f\left(\overline{x}\right)$ with respect to any of its variables, say $x_i, 1 \le i \le n$. It is simply a partial derivative of $f\left(\overline{x}\right)$ with respect to x_i. You also know that this partial derivative indicates the rate of change of $f\left(\overline{x}\right)$ in the direction of x_i. What if you want to know the rate of change of $f\left(\overline{x}\right)$ in some arbitrary direction denoted by $\overline{h}$? 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 $\varepsilon = 0$

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

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 $J(x), \quad x \in D,$ 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

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

must hold for all sufficiently small ε

Now,
for $\varepsilon \geq 0$

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

and for $\varepsilon \leq 0$

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

If we let $\varepsilon \to 0,$

$$\left.\begin{aligned} &\lim_{\substack{\varepsilon\to 0\\ \varepsilon>0}} \frac{J(x^* + \varepsilon h) - J(x^*)}{\varepsilon} \geq 0 \\ \text{and} \quad &\lim_{\substack{\varepsilon\to 0\\ \varepsilon<0}} \frac{J(x^* + \varepsilon h) - J(x^*)}{\varepsilon} \leq 0 \end{aligned}\right\} \underbrace{\lim_{\varepsilon\to 0} \frac{J(x^* + \varepsilon h) - J(x^*)}{\varepsilon}}_{\substack{\text{existence of Gâteaux variation}\\ \text{ensures the existence of this limit}}} = \delta J(x;h) = 0$$

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 Fréchet differential 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.

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

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

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

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

ii) $dJ\left(x;a_1h_1+a_2h_2\right)=a_1dJ\left(x;h_1\right)+a_2dJ(x;h_2)$ 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) $dJ\left(x;h\right)\le \text{constant } \|h\|$ for all $h\in X$
This is the continuity requirement on the Fréchet differential.

Ananthasuresh, IISc

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

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.

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,

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

Here is why:

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

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

A small rearrangement of terms yields

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

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

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

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

<u>Operations using Gateaux variation</u>

Consider a simple general functional of the form shown below.

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

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

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

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

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

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

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.

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

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 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.