Appendix A

Calculus of variations, functional derivatives

The calculus of variations is to multivariable calculus what functions are to vectors. It answers the question of how to differentiate with respect to functions, i.e., objects with an uncountable, infinite number of degrees of freedom. Functional calculus is used to formulate linearized forward models for imaging, as well as higher-order terms in Born series. It is also useful for finding stationary-point conditions of Lagrangians, and gradient descent directions in optimization.

Let X, Y be two function spaces endowed with norms and inner products (technically, Hilbert spaces). A *functional* ϕ is a map from X to $\mathbb{R}$. We denote its action on a function f as $\phi(f)$. An *operator* F is a map from X to Y. We denote its action on a function f as Ff.

We say that a functional ϕ is Fréchet differentiable at $f \in X$ when there exists a linear functional $A : X \mapsto \mathbb{R}$ such that

$$\lim_{h \to 0} \frac{|\phi(f+h) - \phi(f) - A(h)|}{\|h\|} = 0.$$

If this relation holds, we say that A is the *functional derivative*, or Fréchet derivative, of ϕ at f, and we denote it as

$$A = \frac{\delta \phi}{\delta f}[f].$$

It is also called the *first variation* of ϕ. It is the equivalent of the gradient in multivariable calculus. The fact that A is a map from X to $\mathbb{R}$ corresponds to the idea that a gradient maps vectors to scalars when paired with the dot product, to form directional derivatives. If $X = \mathbb{R}^n$ and $f = (f_1, \ldots, f_n)$, we have

$$\frac{\delta \phi}{\delta f}[f](h) = \nabla \phi(f) \cdot h.$$

For this reason, it is is also fine to write $A(h) = \langle A, h \rangle$.

The differential ratio formula for $\frac{\delta \phi}{\delta f}$ is called Gâteaux derivative,

$$\frac{\delta \phi}{\delta f}[f](h) = \lim_{t \to 0} \frac{\phi(f + th) - \phi(f)}{t}, \tag{A.1}$$

which corresponds to the idea of the directional derivative in $\mathbb{R}^n$.

Examples of functional derivatives:

- $\phi(f) = \langle g, f \rangle$,

  $$\frac{\delta \phi}{\delta f}[f] = g, \qquad \frac{\delta \phi}{\delta f}[f](h) = \langle g, h \rangle$$

  Because ϕ is linear, $\frac{\delta \phi}{\delta f} = \phi$. Proof: $\phi(f + th) - \phi(f) = \langle g, f + th \rangle - \langle g, f \rangle = t\langle g, h \rangle$, then use (A.1).

- $\phi(f) = f(x_0)$,

  $$\frac{\delta \phi}{\delta f}[f] = \delta(x - x_0), \qquad \text{(Dirac delta)}.$$

  This is the special case when $g(x) = \delta(x - x_0)$. Again, $\frac{\delta \phi}{\delta f} = \phi$.

- $\phi(f) = \langle g, f^2 \rangle$,

  $$\frac{\delta \phi}{\delta f}[f] = 2fg.$$

  Proof: $\phi(f + th) - \phi(f) = \langle g, (f + th)^2 \rangle - \langle g, f \rangle = t\langle g, 2fh \rangle + O(t^2) = t\langle 2fg, h \rangle + O(t^2)$, then use (A.1).

Nonlinear operators $\mathcal{F}[f]$ can also be differentiated with respect to their input function. We say $\mathcal{F} : X \to Y$ is Fréchet differentiable when there exists a linear operator $F : X \to Y$

$$\lim_{h \to 0} \frac{\|\mathcal{F}[f + h] - \mathcal{F}[f] - Fh\|}{\|h\|} = 0.$$

F is the functional derivative of $\mathcal{F}$, and we write

$$F = \frac{\delta \mathcal{F}}{\delta f}[f].$$

We still have the difference formula

$$\frac{\delta \mathcal{F}}{\delta f}[f]h = \lim_{t \to 0} \frac{\mathcal{F}[f + th] - \mathcal{F}[f]}{t}.$$

Examples:

- $\mathcal{F}[f] = f$. Then
$$\frac{\delta \mathcal{F}}{\delta f}[f] = I,$$
the identity. Proof: $\mathcal{F}$ is linear hence equals its functional derivative. Alternatively, apply the difference formula to get $\frac{\delta \mathcal{F}}{\delta f}[f]h = h$.

- $\mathcal{F}[f] = f^2$. Then
$$\frac{\delta \mathcal{F}}{\delta f}[f] = 2f,$$
the operator of multiplication by $2f$.

Under a suitable smoothness assumption, the Fréchet Hessian of an operator F can also be defined: it takes two functions as input, and returns a function in a linear manner ("bilinear operator"). It is defined through a similar finite-difference formula

$$\langle \frac{\delta^2 \mathcal{F}}{\delta f^2}[f]h_1, h_2 \rangle = \lim_{t \to 0} \frac{\mathcal{F}[f + t(h_2 + h_1)] - \mathcal{F}[f + th_2] - \mathcal{F}[f + th_1] + \mathcal{F}[f]}{t^2}.$$

The Hessian is also called second variation of $\mathcal{F}$. For practical calculations of the Hessian, the notation $\frac{\delta^2 \mathcal{F}}{\delta f^2}$ is too cavalier. Instead, it is useful to view the Hessian as the double directional derivative

$$\frac{\delta^2 \mathcal{F}}{\delta f \delta f'}$$

in two directions f and f', and compute those derivatives one at a time. This formula is the equivalent of the mixed partial $\frac{\partial^2 f}{\partial x_i \partial x_j}$ when the two directions are x_i and x_j in n dimensions.

Functional derivatives obey all the properties of multivariable calculus, such as chain rule and derivative of a product (when all the parties are sufficiently differentiable).

Whenever in doubt when faced with calculations involving functional derivatives, keep track of free variables vs. integration variables — the equivalent of "free indices" and "summation indices" in vector calculus. For instance,

- $\frac{\delta \mathcal{F}}{\delta f}$ is like $\frac{\delta \mathcal{F}_i}{\delta f_j}$, with two free indices i and j;
- $\frac{\delta \mathcal{F}}{\delta f} h$ is like $\sum_j \frac{\delta \mathcal{F}_i}{\delta f_j} h_j$, with one free index i and one summation index j.
- $\frac{\delta^2 \mathcal{F}}{\delta f^2}$ is like $\frac{\delta^2 \mathcal{F}_i}{\delta f_j \delta f_k}$, with three free indices i, j, k.
- $\langle \frac{\delta^2 \mathcal{F}}{\delta f^2} h_1, h_2 \rangle$ is like $\sum_{j,k} \frac{\delta^2 \mathcal{F}_i}{\delta f_j \delta f_k} (h_1)_j (h_2)_k$, with one free index i and two summation indices j and k.

No free index indicates a scalar, one free index indicates a function (or a functional), two free indices indicate an operator, three indices indicate an "object that takes in two functions and returns one", etc.

18.325 Topics in Applied Mathematics: Waves and Imaging
Fall 2015