10 Calculus of Variations

In this lecture, we will apply our derivative machinery to a new type of input: neither scalars, nor column vectors, nor matrices, but rather the **inputs will be functions** $u(x)$, which form a perfectly good vector space (and can even have norms and inner products).[12] It turns out that there are lots of amazing applications for differentiating with respect to *functions*, and the resulting techniques are sometimes called the "calculus of variations" and/or "Frechét" derivatives.

10.1 Functionals: Mapping functions to scalars

Example 44

For example, consider functions $u(x)$ that map $x \in [0,1] \to u(x) \in \mathbb{R}$. We may then define the function f:

$$f(u) = \int_0^1 \sin(u(x))\,\mathrm{d}x.$$

Such a function, mapping an input *function* u to an output *number*, is sometimes called a "functional." What is f' or ∇f in this case?

Recall that, given any function f, we always define the derivative as a linear operator $f'(u)$ via the equation:

$$df = f(u + du) - f(u) = f'(u)[du]\,,$$

where now du denotes an arbitrary "small-valued" *function* $du(x)$ that represents a small change in $u(x)$, as depicted in Fig. 12 for the analogous case of a non-infinitesimal $\delta u(x)$. Here, we may compute this via linearization of the integrand:

$$\begin{aligned} df &= f(u + du) - f(u) \\ &= \int_0^1 \sin(u(x) + du(x)) - \sin(u(x))\,dx \\ &= \int_0^1 \cos(u(x))\,du(x)\,dx = f'(u)[du]\,, \end{aligned}$$

where in the last step we took $du(x)$ to be arbitrarily small[13] so that we could linearize $\sin(u + du)$ to first-order in $du(x)$. That's it, we have our derivative $f'(u)$ as a perfectly good linear operation acting on du!

10.2 Inner products of functions

In order to define a gradient ∇f when studying such "functionals" (maps from functions to $\mathbb{R}$), it is natural to ask if there is an inner product on the input space. In fact, there are perfectly good ways to define inner products of functions! Given functions $u(x), v(x)$ defined on $x \in [0,1]$, we could define a "Euclidean" inner product:

$$\langle u, v \rangle = \int_0^1 u(x)v(x)\,\mathrm{d}x.$$

[12] Being fully mathematically rigorous with vector spaces of functions requires a lot of tedious care in specifying a well-behaved set of functions, inserting annoying caveats about functions that differ only at isolated points, and so forth. In this lecture, we will mostly ignore such technicalities—we will implicitly assume that our functions are integrable, differentiable, etcetera, as needed. The subject of *functional analysis* exists to treat such matters with more care.

[13] Technically, it only needs to be small "almost everywhere" since jumps that occur only at isolated points don't affect the integral.

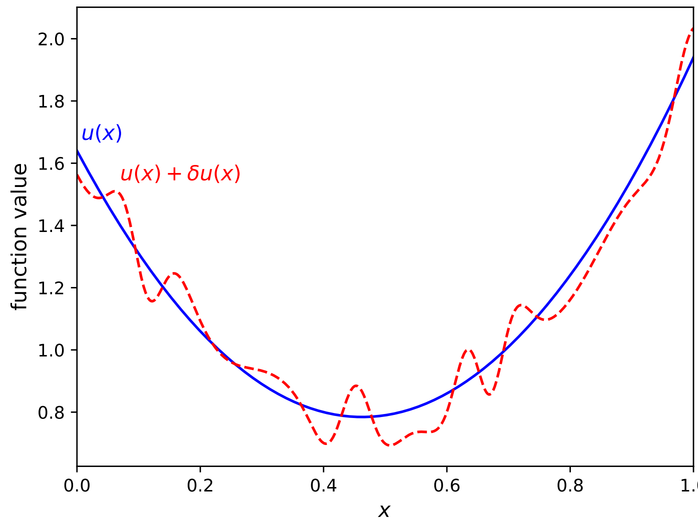

Figure 12: If our $f(u)$'s inputs u are *functions* $u(x)$ (e.g., mapping $[0,1] \mapsto \mathbb{R}$), then the essence of differentiation is linearizing f for small perturbations $\delta u(x)$ that are themselves functions, in the limit where $\delta u(x)$ becomes arbitrarily small. Here, we show an example of a $u(x)$ and a perturbation $u(x) + \delta u(x)$.

Notice that this implies

$$\|u\| := \sqrt{\langle u, u \rangle} = \sqrt{\int_0^1 u(x)^2 dx}\,.$$

Recall that the gradient ∇f is *defined* as whatever we take the inner product of du with to obtain df. Therefore, we obtain the gradient as follows:

$$df = f'(u)[du] = \int_0^1 \cos(u(x))\, du(x)\, dx = \langle \nabla f, du \rangle \implies \nabla f = \cos(u(x))\,.$$

The two infinitesimals du and dx may seem a bit disconcerting, but if this is confusing you can just think of the $du(x)$ as a small non-infinitesimal function $\delta u(x)$ (as in Fig. 12) for which we are dropping higher-order terms.

The gradient ∇f is just another function, $\cos(u(x))$! As usual, ∇f has the same "shape" as u.

Remark 45. *It might be instructive here to compare the gradient of an integral, above, with a discretized version where the integral is replaced by a sum. If we have*

$$f(u) = \sum_{k=1}^{n} \sin(u_k)\Delta x$$

where $\Delta x = 1/n$, for a vector $u \in \mathbb{R}^n$, related to our previous $u(x)$ by $u_k = u(k\Delta x)$, which can be thought of as a "rectangle rule" (or Riemann sum, or Euler) approximation for the integral. Then,

$$\nabla_u f = \begin{pmatrix} \cos(u_1) \\ \cos(u_2) \\ \vdots \end{pmatrix} \Delta x\,.$$

Why does this discrete version have a Δx multiplying the gradient, whereas our continuous version did not? The reason is that in the continuous version we effectively included the dx in the definition of the inner product $\langle u, v \rangle$ (which was an integral). In discrete case, the ordinary inner product (hence the ordinary gradient) is just a sum.

However, if we define a weighted *discrete inner product* $\langle u, v\rangle = \sum_{k=1}^{n} u_k v_k \Delta x$*, then, according to Sec. 5, this changes the definition of the gradient, and in fact will remove the* Δx *term to correspond to the continuous version.*

10.3 Example: Minimizing arc length

We now consider a more tricky example with an intuitive geometric interpretation.

Example 46

Let u be a differentiable function on $[0, 1]$ and consider the functional

$$f(u) = \int_0^1 \sqrt{1 + u'(x)^2}\, dx.$$

Solve for ∇f when $u(0) = u(1) = 0$.

Geometrically, you learned in first-year calculus that this is simply the **length of the curve** $u(x)$ from $x = 0$ to $x = 1$. To differentiate this, first notice that ordinary single-variable calculus gives us the linearization

$$d\left(\sqrt{1+v^2}\right) = \sqrt{1+(v+dv)^2} - \sqrt{1+v^2} = \left(\sqrt{1+v^2}\right)' dv = \frac{v}{\sqrt{1+v^2}} dv\,.$$

Therefore,

$$\begin{aligned} df &= f(u+du) - f(u) \\ &= \int_0^1 \left(\sqrt{1+(u+du)'^2} - \sqrt{1+u'^2}\right) dx \\ &= \int_0^1 \frac{u'}{\sqrt{1+u'^2}}\, du' dx. \end{aligned}$$

However, this is a linear operator on du' and not (directly) on du. Abstractly, this is fine, because du' is itself a linear operation on du, so we have $f'(u)[du]$ as the composition of two linear operations. However, it is more revealing to rewrite it explicitly in terms of du, for example in order to define ∇f. To accomplish this, we can apply *integration by parts* to obtain

$$f'(u)[du] = \int_0^1 \frac{u'}{\sqrt{1+u'^2}}\, du' dx = \left.\frac{u'}{\sqrt{1+u'^2}} du\right|_0^1 - \int_0^1 \left(\frac{u'}{\sqrt{1+u'^2}}\right)' du\, dx\,.$$

Notice that up until now we did not need utilize the "boundary conditions" $u(0) = u(1) = 0$ for this calculation. However, if we want to restrict ourselves to such functions $u(x)$, then our perturbation du cannot change the endpoint values, i.e. we must have $du(0) = du(1) = 0$. (Geometrically, suppose that we want to find the u that minimizes arc length between $(0, 0)$ and $(1, 0)$, so that we need to fix the endpoints.) This implies that the boundary term in the above equation is zero. Hence, we have that

$$df = -\int_0^1 \underbrace{\left(\frac{u'}{\sqrt{1+u'^2}}\right)'}_{\nabla f} du\, dx = \langle \nabla f, du\rangle\,.$$

Furthermore, note that the u that minimizes the functional f has the property that $\left.\nabla f\right|_u = 0$. Therefore, for a u that minimizes the functional f (the *shortest curve*), we must have the following result:

$$\begin{aligned} 0 = \nabla f &= -\left(\frac{u'}{\sqrt{1+u'^2}}\right)' \\ &= -\frac{u''\sqrt{1+u'^2} - u'\frac{u''u'}{\sqrt{1+u'^2}}}{1+u'^2} \\ &= -\frac{u''(1+u'^2) - u''u'^2}{(1+u'^2)^{3/2}} \\ &= -\frac{u''}{(1+u'^2)^{3/2}}. \end{aligned}$$

Hence, $\nabla f = 0 \implies u''(x) = 0 \implies u(x) = ax + b$ for constants a, b; and for these boundary conditions $a = b = 0$. In other words, u is the horizontal straight line segment!

Thus, we have recovered the familiar result that straight line segments in $\mathbb{R}^2$ are the shortest curves between two points!

Remark 47. *Notice that the expression* $\frac{u''}{(1+u'^2)^{3/2}}$ *is the formula from multivariable calculus for the curvature of the curve defined by* $y = u(x)$. *It is not a coincidence that the gradient of arc length is the (negative) curvature, and the minimum arc length occurs for zero gradient = zero curvature.*

10.4 Euler–Lagrange equations

This style of calculation is part of the subject known as the **calculus of variations**. Of course, the final answer in this example above (a straight line) may have been obvious, but a similar approach can be applied to many more interesting problems. We can generalize the approach as follows:

Example 48

Let $f(u) = \int_a^b F(u, u', x)\,dx$ where u is a differentiable function on $[a, b]$. Suppose the endpoints of u are fixed (i.e. its values at $x = a$ and $x = b$ are constants). Let us calculate df and ∇f.

We find:

$$\begin{aligned} df &= f(u + du) - f(u) \\ &= \int_a^b \left(\frac{\partial F}{\partial u} du + \frac{\partial F}{\partial u'} du'\right) dx \\ &= \underbrace{\frac{\partial F}{\partial u'} du\Big|_a^b}_{=0} + \int_a^b \left(\frac{\partial F}{\partial u} - \left(\frac{\partial F}{\partial u'}\right)'\right) du\,dx\,, \end{aligned}$$

where we used the fact that $du = 0$ at a or b if the endpoints $u(a)$ and $u(b)$ are fixed. Hence,

$$\nabla f = \frac{\partial F}{\partial u} - \left(\frac{\partial F}{\partial u'}\right)',$$

which equals zero at extremum. Notice that this gives rise to a 2nd-order differential equation in u, known as the Euler–Lagrange equations!

Remark 49. *The notation* $\partial F/\partial u'$ *is a notoriously confusing aspect of the calculus of variations—what does it mean to take the derivative "with respect to* u'*" while holding* u *fixed? A more explicit, albeit more verbose, way* *of expressing this is to think of* $F(u, v, x)$ *as a function of three* unrelated *arguments, for which we only substitute* $v = u'$ after *differentiating with respect to the second argument* v*:*

$$\frac{\partial F}{\partial u'} = \left.\frac{\partial F}{\partial v}\right|_{v=u'} .$$

There are many wonderful applications of this idea. For example, search online for information about the "brachistochrone problem" (animated here) and/or the "principle of least action". Another example is a catenary curve, which minimizes the potential energy of a hanging cable. A classic textbook on the topic is *Calculus of Variations* by Gelfand and Fomin.