Lecture 6 Part 2: Calculus of Variations and Gradients of Functionals

MIT 18.S096 Matrix Calculus for Machine Learning and Beyond

March 15, 2024

1 Calculus of variations

In previous lectures, we dealt with **vectors** (in an abstract sense) of real values (e.g., real-valued vectors and real-valued matrices). However, there are also other kinds of **vectors**. For example, certain classes of functions can be considered as vectors and form vector spaces as well.

As in previous lectures, here we are interested in expressing df as a linear operator on du, i.e.,

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

except that now $u \in$ some vector space of functions. Usually, $f(u) \in \mathbb{R}$, i.e., f is a function that maps from functions to scalars. Once we express df as a linear operator on du, we can use the Riesz representation theorem[1] (i.e., linear functions in Hilbert/Banach spaces can be written as dot products) to find the gradient. But first, how should we define the dot product for functions?

My thoughts. It's helpful to think of functions as infinite-dimensional real-valued vectors.

2 Dot product of functions

Suppose we have $u(x)$ and $v(x)$ on $x \in [0, 1]$. Then, a natural "Euclidean" inner product for functions can be defined as

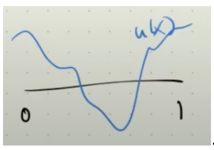
$$u \cdot v \triangleq \int_0^1 u(x) v(x)\, dx.$$

It follows that the norm can be defined as

$$\|u\| \triangleq \sqrt{\int_0^1 u(x)^2\, dx}.$$

3 A simple example

Let u be a function that maps from $[0, 1]$ to $\mathbb{R}$. For example:



.

Define f to be

$$f(u) = \int_0^1 \sin(u(x))\, dx.$$

1. I'm no expert in functional analysis so I'm only using the language I heard in the lecture. If my interpretation is wrong, please let me know.

Linearize f:

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

Treating $u(x)$ as a variable, we know that

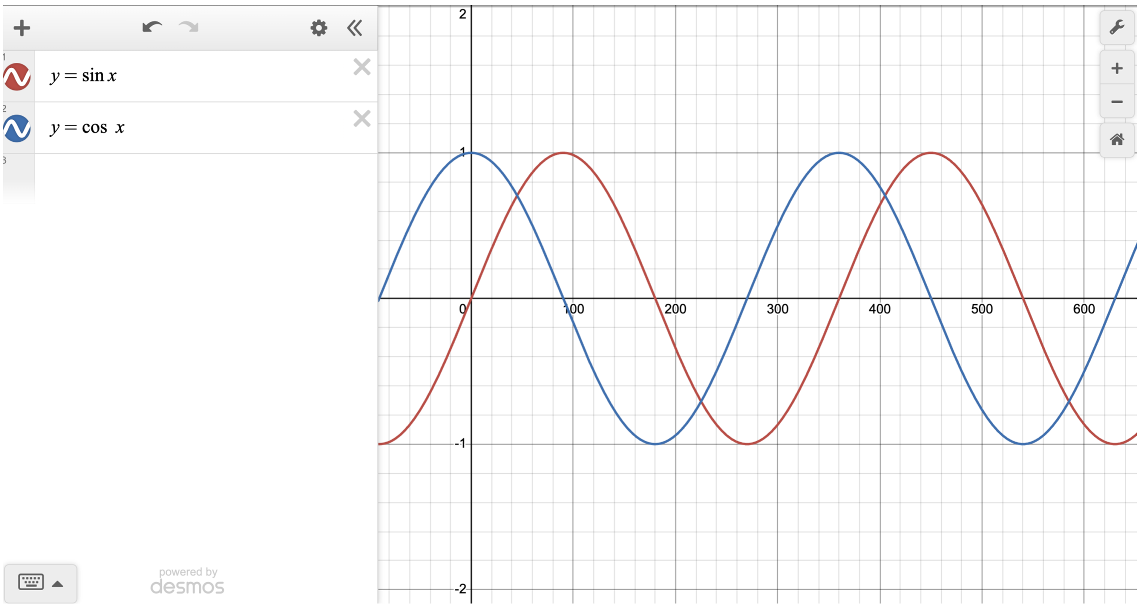
$$\begin{aligned} \sin(u(x)+du(x))-\sin(u(x)) &= \cos(u(x))\,du(x)\\ &\Downarrow\\ \sin(u(x)+du(x)) &= \sin(u(x))+\cos(u(x))\,du(x). \end{aligned}$$

Using this result, obtain

$$\begin{aligned} &= \int_0^1 \Big[\cos(u(x))\,du(x)\Big]\,dx\\ &= \cos(u)\cdot du\\ &\Downarrow\\ \nabla f &= \cos(u) \end{aligned}$$

My thoughts. How to make sense of this result? Here's my take. Just as the gradient of a $\mathbb{R}^n \to \mathbb{R}$ function is in $\mathbb{R}^n$, the gradient of a $\mathbb{R}^\infty$ (a function) $\to\mathbb{R}$ should be in $\mathbb{R}^\infty$ (a function) as well (because they should be over the same input domain).

We see that in order for ∇f to be zero for all $x \in [0,1]$, u needs to be a flat line $y = 90° \pm 360°n$ for $n \in \mathbb{Z}$ (maxima) or $y = 270° \pm 360°n$ for $n \in \mathbb{Z}$ (minima).



4 A more tricky example

The arc length of $u(x)$ from $x=a$ to $x=b$ is

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

Linearize f:

$$\begin{aligned}
df &= f(u+du)-f(u) \\
&= \int_a^b \left[\sqrt{1+(u(x)+du(x))'^2} - \sqrt{1+u'(x)^2} \right] dx \\
&= \int_a^b \left[\sqrt{1+(u'(x)+du'(x))^2} - \sqrt{1+u'(x)^2} \right] dx \quad \text{(linearity of the derivative)}
\end{aligned}$$

Treating $u'(x)$ as a variable, we can show that

$$\begin{aligned}
\sqrt{1+(u'(x)+du'(x))'^2} - \sqrt{1+u'(x)^2} &= \frac{1}{2}(1+(u(x))'^2)^{-1/2} \times (2(u(x))') \, {\color{red} du'(x)} \\
&= \frac{u'(x)^2}{\sqrt{1+u'(x)^2}} \, {\color{red} du'(x)}.
\end{aligned}$$

Using this result, obtain

$$df = \int_a^b \left[\frac{u'(x)}{\sqrt{1+u'(x)^2}} du'(x) \right] dx.$$

But this is a linear operator on du', not du, so it can't be written as a dot product.

Integration by parts for definite integrals (review).

$$\int_a^b f(x) g'(x) = f(x) g(x) \Big|_a^b - \int_a^b f'(x) g(x) \, dx$$

To get du into the integral, use integration by parts:

$$df = \int_a^b \underbrace{\frac{u'(x)}{\sqrt{1+u'(x)^2}}}_{f(x)} \underbrace{du'(x)}_{g'(x)} dx = \underbrace{\frac{u'(x)}{\sqrt{1+u'(x)^2}}}_{f(x)} \underbrace{du(x)}_{g(x)} \Big|_{x=a}^{x=b} - \int_a^b \underbrace{\left(\frac{u'(x)}{\sqrt{1+u'(x)^2}} \right)'}_{f'(x)} \underbrace{du\,(x)}_{g(x)} \, dx.$$

Suppose we fix $u(a)=A$ and $u(b)=B$ and also fix $du(a)=du(b)=0$ (perturbations must not change the endpoints). Then, the first term would be zero, which gives us

$$df = \int_a^b \left(\frac{u'(x)}{\sqrt{1+u'(x)^2}} \right)' du\,(x) \, dx.$$

The gradient:

$$\nabla f = \left(\frac{u'(x)}{\sqrt{1+u'(x)^2}} \right)' \overset{(*)}{=\!=} \frac{u''(x)}{1+u'(x)^{3/2}}$$

In order for this to be zero over $[a,b]$, we must have $u''(x)=0$ (i.e., a straight line).

$(*)$ Use quotient rule:

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

5 Generalization of the tricky example

Generally, we might define f as the following integral:

$$f(u) = \int_a^b F(u, u', x)\, dx.$$

Linearize f:

$$\begin{aligned}
df &= f(u+du) - f(u) && (2)\\
&= \int_a^b \Bigg[\underbrace{\frac{\partial F}{\partial u} du}_{\text{change wrt 1st arg}} + \underbrace{\frac{\partial F}{\partial u'} du'}_{\text{change wrt 2nd arg}} \Bigg] dx && (3)\\
&= \underbrace{\frac{\partial F}{\partial u'} du \Big|_a^b}_{\substack{\text{0 if endpoints are fixed}\\ (du(a)=du(b)=0)}} + \int_a^b \underbrace{\left[\frac{\partial F}{\partial u} - \left(\frac{\partial F}{\partial u'} \right)' \right]}_{\nabla f} du\, dx. && (4)
\end{aligned}$$

Derivation notes:

- To get from Equation 2 to 3: $F(u+du) = f(u) + \frac{\partial F}{\partial u} du + \frac{\partial F}{\partial u'} du'$ (18.01 stuff)
- To get from Equation 3 to 4: apply integration by parts to the second part only ($\frac{\partial F}{\partial u'} du'$)

At extrema, we have

$$\frac{\partial F}{\partial u} - \left(\frac{\partial F}{\partial u'} \right)' = 0,$$

which is a 2nd order differential equation in u called the *Euler-Lagrange* equation.

6 Lingering questions

du is a small perturbation to u, but not all perturbations are allowed for functions right? (unlike for real vectors) Maybe we want smooth & differentiable perturbations?