Math 5311 – Gateaux differentials and Frechet derivatives

Kevin Long

January 26, 2009

1 Differentiation in vector spaces

Thus far, we've developed the theory of minimization without reference to derivatives. We've even been able to compute minimizers of quadratic forms without using derivatives, by proving that the minimizer of a positive definite quadratic form must be the solution to the algebraic equation $Kx = f$. However, in computations it's convenient and efficient to use derivatives.

We're developing the theory of minimization of functions set in arbitrary vector spaces, so we need to develop differential calculus in that setting. In multivariable calculus, you learned three related concepts: directional derivatives, partial derivatives, and gradients. In arbitrary vector spaces, we will be able to develop a generalization of the directional derivative (called the Gateaux differential) and of the gradient (called the Frechet derivative). We won't go deeply into the theory of these derivatives in this course, but we'll establish the basic differentiation rules.

A reference for differentiation in infinite-dimensional vector spaces, done at the level of this course, is D. R. Smith, *Variational Methods in Optimization*, Dover, 1998. A careful development of the finite-dimensional case can be found in M. Spivak, *Calculus on Manifolds*, J. Munkres, *Analysis on Manifolds*, or Rudin, P*rinciples of Mathematical Analysis*. If you're interested, E. W. Cheney's *Analysis for Applied Mathematics* gives a treatment of Gateaux and Frechet derivatives at a level one notch above the level of this course.

1.1 The Gateaux differential

The Gateaux differential generalizes the idea of a directional derivative.

Definition 1. Let $f : V \to U$ be a function and let $h \neq 0$ and x be vectors in V. The Gateaux differential $d_h f$ is defined

$$d_h f = \lim_{\epsilon \to 0} \frac{f(x + \epsilon h) - f(x)}{\epsilon}.$$

Some things to notice about the Gateaux differential:

- There is not a single Gateaux differential at each point. Rather, at each point x there is a Gateaux differential for each direction h. In one dimension, there are two Gateaux differentials for every x: one directed "forward," one "backward." In two of more dimensions, there are infinitely many Gateaux differentials at each point!
- The Gateaux differential is a one-dimensional calculation along a specified direction h. Because it's one-dimensional, you can use ordinary one-dimensional calculus to compute it. Your old friends such as the chain rule work for Gateaux differentials. Thus, it's usually easy to compute a Gateaux differential even when the space V is infinite dimensional.

1.2 Examples of Gateaux differentials

1.2.1 Linear and quadratic functions

Let f and x be vectors in an inner product space, and define $p(x) = x^T f$. Then

$$d_h p = \lim_{\epsilon \to 0} \frac{x^T f + \epsilon h^T f - x^T f}{\epsilon}$$

$$d_h(x^T f) = h^T f.$$

Next, let K be a symmetric matrix, and define $p(x) = 2x^T f + x^T Kx$. Compute the Gateaux differential

$$\begin{aligned} d_h p &= \lim_{\epsilon \to 0} \frac{2x^T f + \epsilon h^T f + x^T Kx + 2\epsilon h^T Kx + \epsilon^2 h^T Kh - 2x^T f - x^T Kx}{\epsilon} \\ &= h^T f + 2h^T Kx. \end{aligned}$$

1.2.2 The exponential function

There are a number of ways to define the exponential function; most convenient for present purposes is to define it as the infinite series

$$e^x = 1 + x + \frac{x^2}{2} + \cdots + \frac{x^k}{k!} + \cdots .$$

This makes sense even for vector-valued arguments provided multiplication is understood as "elementwise," is in the Matlab "dot-star" operation. In the case of an infinite-dimensional vector space V (for example H^1) whose elements are real-valued functions, the exponential function $e^u : V \to V$ simply maps $u(x)$ pointwise to its exponential $e^{u(x)}$.

With e^x defined as above, it can be shown (provided multiplication of members of V is commutative!) by straightforward calculation and rearrangement that $e^{a+b} = e^a e^b$ (see, e.g., blue or green Rudin. If you're interested, see a reference on Lie algebras for information on the exponentiation of non-commutative objects; that turns out to be useful in fields such as aircraft control and quantum physics.)

With the identity $e^{a+b} = e^a e^b$ and the series defining e^x, we can compute the Gateaux derivative

$$\begin{aligned} d_h(e^u) &= \lim_{\epsilon \to 0} \frac{e^u e^{\epsilon h} - e^u}{\epsilon} \\ &= e^u \lim_{\epsilon \to 0} \frac{e^{\epsilon h} - 1}{\epsilon} \\ &= he^u. \end{aligned}$$

1.2.3 The absolute value function in $\mathbb{R}$

Let $f(x) = |x|$. Calculation of the limit gives

$$d_h f = \begin{cases} h \frac{x}{|x|} & x \neq 0 \\ |h| & x = 0. \end{cases}$$

Notice that the Gateaux differentials of $|x|$ do exist at zero; however, at zero, the Gateaux differentials depend on h in a nonlinear way.

1.2.4 A spatial derivative

$$
\begin{aligned}
d_h\left(\frac{du}{dx}\right) &= \lim_{\epsilon\to 0}\frac{u_x+\epsilon h_x - u_x}{\epsilon} \\
&= \frac{dh}{dx}
\end{aligned}
$$

1.2.5 A functional

Let $J : H^1(\Omega) \to \mathbb{R}$ be

$$
J[u] = \int_\Omega \left[\frac{1}{2}u_x^2 + \frac{1}{2}u^2\right]\, dx.
$$

Then

$$
\begin{aligned}
d_h J &= \lim_{\epsilon\to 0}\frac{\int_\Omega \left[\frac{1}{2}u_x^2 + \frac{1}{2}u^2 + \epsilon u h + \epsilon u_x h_x + \frac{1}{2}\epsilon^2 h_x^2 + \frac{1}{2}\epsilon h^2 - \frac{1}{2}u_x^2 - \frac{1}{2}u^2\right]\, dx}{\epsilon} \\
d_h J &= \int_\Omega \left[u h + u_x h_x\right]\, dx
\end{aligned}
$$

Note: it's routine in infinite-dimensional optimization problems to exchange integration and Gateaux differentiation without comment. The integration variable x and the differentiation variable u are different. If the limit $\epsilon \to 0$ can be taken by means of a sequence $\{\epsilon_n\}$ for which the difference quotient of the integrand is dominated by a Lebesgue-integrable function, the conditions of the Lebesgue dominated convergence theorem hold and the limit and integral can be interchanged safely. For the section on optimization, we'll assume this to be true throughout. However, later in the semester we'll see some examples with Fourier series and transforms where differentiation of a convergent series or integral results in divergence at one or more points.

There is, however, an important real-world case where this process requires a little more care: the case where the *limits* of integration depend on the variable u. This arises in the field of *shape optimization*, where the "design variable" u determines the size and shape of a domain Ω.

1.3 Rules for Gateaux differentials

Computing from the definition quickly gets dull, so as with ordinary calculus, let's work out rules for Gateaux differentials.

1.3.1 Differential of a constant

The Gateaux differential of a constant is zero: $d_h c = 0$. The proof follows immediately from the definition.

1.3.2 Sum rule

Gateaux differentiation distributes over sums: $d_h\,(f \pm g) = d_h f \pm d_h g$. The proof follows immediately from the definition.

1.3.3 Product rule and quotient rule

The Gateaux differential of an elementwise product fg is $d_h(fg) = (d_h f)g + f(d_h g)$.

The Gateaux differential of an inner product $\langle f, g\rangle$ (or $f^T g$) is $d_h\,\langle f, g\rangle = \langle f, d_h g\rangle + \langle d_h f, g\rangle$. With transpose notation, this is $d_h(f^T g) = f^T d_h g + (d_h f)^T g$.

The proofs are similar to what you would do in $\mathbb{R}$, and are left as exercises; the little-o notation below will be useful.

1.3.4 Chain rule

It will be useful to introduce "little-o notation," which may be familiar from analysis. To say that a quantity q "is $o(\epsilon)$" means

$$\lim_{\epsilon \to 0} \frac{q}{\epsilon} = 0.$$

For example, ϵ^{1+a} is $o(\epsilon)$ whenever $a > 0$, but $\sin(\epsilon a)$ is not $o(\epsilon)$ when $a \neq 0$. Notice that if q is $o(\epsilon)$ then certainly $\lim_{\epsilon \to 0} q = 0$.

If a function f is Gateaux differentiable, we have, by definition

$$d_h f = \lim_{\epsilon \to 0} \frac{f(x + \epsilon h) - f(x)}{\epsilon}.$$

From this and the definition of $o(\epsilon)$ it follows that

$$f(x + \epsilon h) = f(x) + \epsilon d_h f(x) + o(\epsilon).$$

Now, assume $g : V \to U$ is Gateaux differentiable at $x \in V$, and that $f : U \to W$ is Gateaux differentiable at $g(x)$. We want to compute the Gateaux differential of their composition $(f \circ g)(x) = f(g(x))$. Start with the difference quotient, and use the identity $f(x + \epsilon h) = f(x) + \epsilon d_h f(x) + o(\epsilon)$ twice:

$$d_h (f \circ g)(x) = \lim_{\epsilon \to 0} \frac{f(g(x + \epsilon h)) - f(g(x))}{\epsilon}$$

$$= \lim_{\epsilon \to 0} \frac{f(g(x) + \epsilon d_h g(x) + o(\epsilon)) - f(g(x))}{\epsilon}$$

$$= \lim_{\epsilon \to 0} \frac{f(g(x) + \epsilon(d_h g(x) + \epsilon^{-1} o(\epsilon))) - f(g(x))}{\epsilon}$$

$$= \lim_{\epsilon \to 0} \frac{f(g(x)) + \epsilon d_{(d_h g + \epsilon^{-1} o(\epsilon))} f(g) - f(g(x))}{\epsilon}$$

$$= d_{d_h g} f(g).$$

This may not be immediately recognizable as the chain rule: where's the multiplication? Let's do an example to see how it works.

Chain rule example Compute the Gateaux differential of $F(x) = (x^T x)^2$. Let $f(u) = u^2$. This is a mapping $f : \mathbb{R} \to \mathbb{R}$. Let $g(x) = x^T x$. This is a mapping from some vector space V to the reals. Our function $F(x)$ is the composition of these two: $F(x) = f(g(x))$. Now, from the product rule for inner products we know that $d_h(x^T x) = 2h^T x$, and from the product rule for elementwise products we know that $d_k(u^2) = 2ku$. The chain rule tells us that

$$d_h F(x) = d_{d_h g} f(g)$$

which is, given our f and g,

$$d_h F = d_{2h^T x}(g^2)$$

$$= 2(2h^T x)g$$

$$= 4(h^T x)(x^T x).$$

You should check by comparing to $d_h F$ computed directly from the definition of Gateaux differential. Also, compare to the 1D case $F(x) = (x^2)^2$.

The multiplicative structure of the chain rule is buried in the use of $d_h g$, the differential of the "inner" function g as the direction for the differential of the outer function f. Usually – but **not** always – the direction will appear linearly in the differential, recovering in the usual case the expected form of the chain rule. See below for discussion of exactly when the direction will appear linearly.

1.3.5 Transcendental functions

Gateaux differentiation of the exponential function has been shown above. A similar computation can be done for any other function that can be defined as a power series, for example, the trigonometric functions.

1.4 The Frechet derivative

The Frechet derivative Df of $f : V \to U$ is defined implicitly by

$$f(x+k) = f(x) + (Df)\,k + o(\|k\|).$$

To establish the relationship to the Gateaux differential, take $k = \epsilon h$ and write

$$f(x+\epsilon h) = f(x) + \epsilon(Df)h + h\,o(\epsilon).$$

In the limit $\epsilon \to 0$, we have $(Df)\,h = d_h f$. Then, if $d_h f$ has the form $A\,h$, then we can identify $Df = A$.

1.4.1 Existence and uniqueness of the Frechet derivative

I'll cite two theorems without proof

Theorem 2. *The Frechet derivative exists at x=a iff all Gateaux differentials are continuous functions of x at $x = a$.*

Proof. See, for example, Munkres or Spivak (for R^N) or Cheney (for any normed vector space). □

Theorem 3. *If it exists for a function f at a point x, the Frechet derivative is unique.*

Proof. Assume otherwise, then construct a contradiction. See, for example, Munkres or Spivak (for $\mathbb{R}^N$) or Cheney (for any normed vector space). □

1.4.2 Examples of Gateaux differentiable functions that are not Frechet differentiable

Example: $f(x) = |x|$ The Gateaux differential at $x = 0$ is

$$d_h f = \lim_{\epsilon\to 0} \frac{|\epsilon h|}{\epsilon} = |h|\,.$$

This is not a linear function of h. Therefore, the Frechet derivative does not exist.

Example: $f(x,y) = 2xy/\sqrt{x^2+y^2}$. This function is indeterminate at (0,0). However, we can still construct a function that is continuous at (0,0) provided that $\lim_{(x,y)\to(0,0)} f(x,y)$ exists and is independent of the path along which the limit is taken. Let $(x,y) = (\epsilon h, \epsilon k)$ with $h \neq 0,\ k \neq 0$, so that $f = 2\epsilon hk/\sqrt{h^2+k^2}$. Then $\lim_{\epsilon\to 0} f = 0$ independently of (h,k). By defining

$$f(x,y) = \begin{cases} \frac{2xy}{\sqrt{x^2+y^2}} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0). \end{cases}$$

we have constructed a continuous function.

Compute the Gateaux derivative at $(0,0)$:

$$d_h f(0,0) = \lim_{\epsilon\to 0} \frac{2\epsilon h_x h_y/\sqrt{h_x^2+h_y^2}}{\epsilon}$$

$$= \frac{2\,h_x h_y}{\sqrt{h_x^2+h_y^2}}.$$

This is not a linear function of $h = (h_x, h_y)$, so $f(x,y)$ is not Frechet differentiable at $(0,0)$.

1.5 The chain rule for Frechet differentiation

Recall the chain rule for Gateaux differentials: $d_h(f \circ g) = d_{d_h g(x)} f(g)$. When both f and g are Frechet differentiable, then $d_h g = (Dg)\,h$ and $d_k f = (Df)\,k$. Thus, $d_h(f \circ g) = (Df)\,(Dg)\,h$. This has the expected form of a product of derivatives.

1.6 The Frechet derivative defines a tangent hyperplane

Consider a real-valued function, $f : V \to \mathbb{R}$. When the Frechet derivative Df exists at a point $x = a$, f can be approximated to first order in $\|h\|$ by

$$f(a+h) \approx f(a) + (Df)\,h.$$

Note that $h \in V$ and $Df : V \to V$. We'll write Df as ∇f, and call it the gradient. The term $(\nabla f)\,h$ is an inner product, $\langle \nabla f, h \rangle$ which is of course commutative; we'll often write the linear term as $h^T \nabla f$.

Now, define $p(h) = f(a) + h^T \nabla f(a)$. This is a linear real-valued function in h, and as a linear function, it defines a hyperplane in the space V. Clearly $p(0) = f(a)$ and $Dp(0) = Df(a)$: the function p defines the tangent hyperplane to f.