wj32

information when you need it

Differentiation done correctly: 1. The derivative

Posted on February 21, 2013 by wj32 / 1 Comment

Navigation: **1. The derivative** | 2. Higher derivatives | 3. Partial derivatives | 4. Inverse and implicit functions | 5. Maxima and minima

In multivariable calculus courses, one often encounters nonsensical equations such as the following:

> (Chain rule.) If $F = F(x, y)$ and $x = x(t), y = y(t)$ then
>
> $$\frac{dF}{dt} = \frac{dF}{\partial x}\frac{dx}{dt} + \frac{dF}{\partial y}\frac{dy}{dt}.$$

(What is "F" on the left, and what is "F" on the right?) A partial derivative of a function $f: \mathbb{R}^2 \to \mathbb{R}$ is denoted by an assortment of vague expressions like

$$\frac{\partial f}{\partial x} \quad \text{or} \quad \frac{\partial f(x, y)}{\partial x} \quad \text{or} \quad \frac{\partial f}{\partial x}(x, y) \quad \text{or} \quad f_x,$$

which makes it hard to distinguish between the partial derivative as a function and the partial derivative at a point (u, v):

$$\frac{\partial f}{\partial x}(u, v) \quad \text{or} \quad \left.\frac{\partial f(x, y)}{\partial x}\right|_{(x,y)=(u,v)} \quad \text{or} \quad f_x(u, v) \quad ?$$

Furthermore, it isn't even clear what the difference between

$$\frac{df}{dx} \quad \text{and} \quad \frac{\partial f}{\partial x}$$

is!

Partial derivatives are usually introduced first because they provide an extension of differentiation in one variable that is convenient for computation. The so-called Jacobian matrix is relegated to a section on the change of variables theorem, which, as usual, is presented as a magic formula. Overall, one gets the impression that multivariable differentiation is far more complicated than single variable differentiation, and that the usual formulas do not apply.

This is all wrong. With the right definitions, the same old theorems (and even proofs) from single variable calculus still apply, with very few changes. I recently learned about this material from Lang's *Real and Functional Analysis*, which has an elegantly written chapter on differentiation. I will be writing 5 posts covering the basics of differentiation, starting with this one.

Basic properties

Note: if you do not know what a Banach space or a Banach algebra is, replace "Banach space" with "something like $\mathbb{R}^n$ or $Mat_{n,m}(R)$ but might be infinite-dimensional", and replace "Banach algebra" with "Banach space where you can multiply vectors".

Here, E, F and G will denote real Banach spaces (although the theory remains the same for complex Banach spaces). We write $L(E, F)$ for the space of continuous linear maps from E to F. All Banach algebras are assumed to be unital. As usual, if $f: E \to F$ is a linear map then we often write fx instead of $f(x)$. Although we treat the general case here, I recommend visualizing E, F as $\mathbb{R}^n$ for concreteness.

Definition 1. Let $A \subseteq E$ be an open set and let $f: A \to F$. A continuous linear map $\lambda: E \to F$ is said to be the **Fréchet derivative** or just the **derivative** of f at a point $x \in A$ if

$$\lim_{h \to 0} \frac{f(x + h) - f(x) - \lambda h}{|h|} = 0,$$

and we write $f'(x) = \lambda$ or $Df(x) = \lambda$. If f has a derivative at a point x, we say that f is **differentiable at x**. If f is differentiable at all $x \in A$, then we say that f is **differentiable on A** or simply **differentiable**.

Obviously there is the question of uniqueness.

Theorem 2. *Let $A \subseteq E$ be an open set. The derivative of a function $f: A \to F$ at a point $x \in A$, if it exists, is unique.*

Proof. Suppose λ_1 and λ_2 are both derivatives of f at x. Subtracting gives

$$\lim_{h \to 0} \frac{(\lambda_2 - \lambda_1)h}{|h|} = 0.$$

For a fixed nonzero $u \in E$ we have

$$\lim_{t \to 0^+} \frac{(\lambda_2 - \lambda_1)tu}{|tu|} = 0,$$

and since the left hand side is independent of t we have $(\lambda_2 - \lambda_1)u = 0$ for all $u \in E$. Therefore $\lambda_1 = \lambda_2$. ∎

Theorem 3. *Let $A \subseteq E$ be an open set. If $f: A \to F$ is differentiable at $x \in A$ then f is continuous at x.*

Proof. For small h we can write

$$f(x + h) - f(x) = |h| \left(\frac{f(x + h) - f(x) - \lambda h}{|h|} \right) + \lambda h.$$

The right hand side tends to 0 as $h \to 0$, so f is continuous at x. ∎

Suppose $f: A \to F$ is differentiable on A. We have a map $f' : A \to L(E, F)$ that sends each point in A to the derivative of f at that point. If f' is continuous then we say that f is **continuously differentiable** or **of class C^1**. We will have more to say about this in a later post.

Theorem 4 (Chain rule). *Let $A \subseteq E$ and $B \subseteq F$ be open sets. Let $f: A \to F$ and $g: B \to G$ with $f(A) \subseteq B$. If f is differentiable at x and g is differentiable at $f(x)$, then $g \circ f$ is differentiable at x and*

$$(g \circ f)'(x) = g'(f(x)) \circ f'(x).$$

Proof. To save space, let $y = f(x)$ and define

$$\phi(s) = f(x + s) - f(x) - f'(x)s,$$
$$\psi(t) = g(y + t) - g(y) - g'(y)t,$$
$$\rho(h) = g(f(x + h)) - g(y) - g'(y)f'(x)h$$

so that

$$\lim_{s \to 0} \frac{\phi(s)}{|s|} = \lim_{t \to 0} \frac{\psi(t)}{|t|} = 0$$

since f is differentiable at x and g is differentiable at y. We want to show that

$$\lim_{h \to 0} \frac{\rho(h)}{|h|} = 0.$$

For all sufficiently small h,

$$g(f(x + h)) - g(y) = g(y + f'(x)h + \phi(h)) - g(y)$$
$$= g'(y)(f'(x)h + \phi(h)) + \psi(f'(x)h + \phi(h))$$

and

$$\rho(h) = g'(y)\phi(h) + \psi(f'(x)h + \phi(h)).$$

Since $g'(y)$ is continuous,

$$\lim_{h \to 0} \frac{g'(y)\phi(h)}{|h|} = g'(y)\left[\lim_{h \to 0} \frac{\phi(h)}{|h|} \right] = 0,$$

and it remains to show that

$$\lim_{h \to 0} \frac{\psi(f'(x)h + \phi(h))}{|h|} = 0.$$

Let $\varepsilon > 0$. By (*) there exist $\delta_1, \delta_2, \delta_3 > 0$ such that $|\psi(t)| \le \varepsilon|t|$ whenever $|t| \le \delta_1$, $|f'(x)h + \phi(h)| \le \delta_1$ whenever $|h| \le \delta_2$, and $|\phi(s)| \le |s|$ whenever $|s| \le \delta_3$. Then for all $0 < |h| \le \min(\delta_2, \delta_3)$,

$$\frac{|\psi(f'(x)h + \phi(h))|}{|h|} \le \varepsilon \left(\frac{|f'(x)h|}{|h|} + \frac{\phi(h)}{|h|} \right)$$
$$\le \varepsilon(|f'(x)| + 1).$$

∎

Theorem 5. *Let $F_1, ..., F_m$ be Banach spaces, let $A \subseteq E$ be an open set and let $f: A \to F_1$ and $g: A \to F_2$ be differentiable at $x \in A$.*

1. *If f is constant then $f'(x) = 0$.*
2. *If $f(x) = \lambda x$ for some continuous linear map λ, then $f'(x) = \lambda$.*
3. *If $F_1 = F_2$ then $(f + g)'(x) = f'(x) + g'(x)$.*
4. *$(cf)'(x) = cf'(x)$ for all scalars c.*

5. (Product rule). Suppose there is a continuous bilinear map $\cdot : F_1 \times F_2 \to G$. Then

$$(fg)'(x) = f'(x)g(x) + f(x)g'(x),$$

where $f'(x)g(x)$ is the linear map that takes u to $f'(x)u \cdot g(x)$.

6. If $h: F_1 \times \cdots \times F_m \to G$ is a continuous multilinear map, then

$$h'(x_1, \ldots, x_m)(u_1, \ldots, u_m) = \sum_{j=1}^{m} h(x_1, \ldots, u_j, \ldots, x_m).$$

Proof. The first 4 parts are obvious, so we only prove (5). We have

$$0 = \lim_{h \to 0} \frac{[f(x+h) - f(x) - f'(x)h]g(x+h) + f(x)[g(x+h) - g(x) - g'(x)h]}{|h|}$$

$$= \lim_{h \to 0} \frac{(fg)(x+h) - (fg)(x) - [f'(x)g(x+h) + f(x)g'(x)]h}{|h|}.$$

Now

$$\frac{|f'(x)h[g(x+h) - g(x)]|}{|h|} \le |f'(x)| \, |g(x+h) - g(x)| \to 0$$

as $h \to 0$ since $\cdot$ is continuous and g is continuous at x, so

$$\lim_{h \to 0} \frac{f'(x)h[g(x+h) - g(x)]}{|h|} = 0.$$

Adding this to (*) gives

$$\lim_{h \to 0} \frac{(fg)(x+h) - (fg)(x) - [f'(x)g(x) + f(x)g'(x)]h}{|h|} = 0.$$

∎

Theorem 6. Let E be a Banach algebra and let U be the open set of its invertible elements. Then the map $x \mapsto x^{-1}$ is differentiable on U, and its derivative at a point x is given by

$$u \mapsto -x^{-1}ux^{-1}.$$

Proof. For sufficiently small h we have $|e - (e + x^{-1}h)| = |x^{-1}h| < 1/2$, so $e + x^{-1}h$ is invertible and

$$(x+h)^{-1} - x^{-1} + x^{-1}hx^{-1} = (x(e + x^{-1}h))^{-1} - x^{-1} + x^{-1}hx^{-1}$$
$$= (e + x^{-1}h)^{-1}x^{-1} - x^{-1} + x^{-1}hx^{-1}$$
$$= [(e + x^{-1}h)^{-1} - (e - x^{-1}h)]x^{-1}.$$

Now

$$|(e + x^{-1}h)^{-1} - (e - x^{-1}h)| = \left| \sum_{k=0}^{\infty} (-x^{-1}h)^k - (e - x^{-1}h) \right|$$

$$= \left| \sum_{k=2}^{\infty} (-x^{-1}h)^k \right|$$

$$\le \frac{|x^{-1}h|^2}{1 - |x^{-1}h|}$$

$$\le 2|x^{-1}|^2|h|^2.$$

Combining this with (*) shows that

$$\frac{(x+h)^{-1} - x^{-1} + x^{-1}hx^{-1}}{|h|} \to 0$$

as $h \to 0$. ∎ **Corollary 7** (Quotient rule). Let F_1 be a Banach space, let F_2 be a Banach algebra, and let U be the open set of the invertible elements in F_2. Let $A \subseteq E$ be an open set and let $f: A \to F_1$ and $g: A \to U$ be differentiable at $x \in A$. Suppose there is a continuous bilinear map $\cdot : F_1 \times F_2 \to G$. Write fg^{-1} for the map $(fg^{-1})(x) = f(x)g(x)^{-1}$. Then $(fg^{-1})'(x)$ is given by

$$u \mapsto [f'(x)u]g(x)^{-1} - f(x)g(x)^{-1}[g'(x)u]g(x)^{-1}.$$

In particular, if F_2 is commutative then

$$(f/g)'(x) = [f'(x)g(x) - f(x)g'(x)]g(x)^{-2}.$$

Linear maps and direct sums

Before continuing with the properties of the derivative, we describe a generalization of the concept of block matrices. Suppose $E_1, \ldots, E_m$ and $F_1, \ldots, F_n$ are vector spaces, and $\lambda: E_1 \times \cdots \times E_m \to F_1 \times \cdots \times F_n$ is a linear map. We note that in the case of finitely many vector spaces, the notions of direct sum (coproduct) and direct product (product) coincide. Therefore, we have unique linear maps $\lambda_{i,j}: E_j \to F_i$ such that $\lambda_{i,j} = \pi_i \circ \lambda \circ \iota_j$, where

$\pi_i : F_1 \times \cdots \times F_n \to F_i$ is the canonical projection and $\iota_j : E_j \to E_1 \times \cdots \times E_m$ is the canonical injection. In this case we write

$$\lambda = \begin{bmatrix} \lambda_{1,1} & \cdots & \lambda_{1,m} \\ \vdots & \ddots & \vdots \\ \lambda_{n,1} & \cdots & \lambda_{n,m} \end{bmatrix}$$

and say that the matrix **represents** λ, and sometimes that the maps $\lambda_{i,j}$ are the **components** of λ. If $\tau : F_1 \times \cdots \times F_n \to G_1 \times \cdots \times G_p$ is another linear map, then it is easy to verify that $\tau \circ \lambda$ is represented by the product of the matrix representing τ with the matrix representing λ, with the usual matrix multiplication formula. On the other hand, if we start with linear maps $\lambda_{i,j}$, then there is a unique linear map λ that has the components $\lambda_{i,j}$. The situation is entirely analogous to block matrices in R or C, so we do not go any further.

Theorem 8. Let $A \subseteq E$ be an open set, let $F_1, ..., F_n$ be Banach spaces, let $f : A \to F_1 \times \cdots \times F_n$ and let $f_i = \pi_i \circ f$ be the component functions of f, where $\pi_i : F_1 \times \cdots \times F_n \to F_i$ is the canonical projection. Then f is differentiable at a point $x \in A$ if and only if every f_i is differentiable at x. In that case we have $f_i'(x) = \pi_i \circ f'(x)$, i.e.

$$f'(x) = \begin{bmatrix} f_1'(x) \\ \vdots \\ f_n'(x) \end{bmatrix}.$$

Proof. If λ is a linear map then the ith entry (with respect to the direct sum decomposition) of

$$T(h) = \frac{f(x+h) - f(x) - \lambda h}{|h|}$$

is simply

$$T_i(h) = \frac{f_i(x+h) - f_i(x) - \pi_i \lambda h}{|h|}.$$

Therefore $T(h)$ approaches 0 as $h \to 0$ if and only if every $T_i(h)$ approaches 0 as $h \to 0$. The second statement is clear from the above. ▢

The fundamental theorem of calculus

The definition of the derivative extends naturally to closed intervals (with more than one point) in R. If $f : [a, b] \to E$ is differentiable and $x \in [a, b]$, then f' is a linear map from R to E. For convenience, we identify $f'(x)$ with $f'(x)(1)$ and write $f'(x) = c \in E$, where $c = f'(x)(1)$. This coincides with the elementary definition of the derivative as

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

Lemma 9. Let $f : [a, b] \to E$ be differentiable. If $f'(x) = 0$ for all $x \in [a, b]$, then f is constant.

Proof. Suppose $f(t) \neq f(a)$ for some $t \in [a, b]$, and choose a linear functional λ such that $\lambda(f(t)) \neq \lambda(f(a))$, e.g. by applying the Hahn-Banach theorem. Then $\lambda \circ f$ is differentiable and $(\lambda \circ f)'(x) = 0$ for all $x \in [a, b]$, which implies that $\lambda \circ f$ is constant. This is a contradiction. ▢

Integration is essential to the study of differentiation. Here we assume the existence of an integral that can integrate Banach space valued continuous functions defined on closed intervals, e.g. the Bochner integral or the Regulated integral. However, we only use the most basic properties of the integral, such as linearity and the absolute value estimate.

Theorem 10 (Fundamental theorem of calculus). *Let $f : [a, b] \to E$ be an integrable function, and suppose that f is continuous at $x \in [a, b]$. Then the map*

$$t \mapsto \int_a^t f$$

is differentiable at x and its derivative is $f(x)$.

Proof. We have

$$\frac{1}{|h|} \left| \int_a^{x+h} f - \int_a^x f - f(x)h \right| = \frac{1}{|h|} \left| \int_x^{x+h} [f(t) - f(x)] \, dt \right|$$
$$\leq \sup_t |f(t) - f(x)|$$
$$\to 0$$

as $h \to 0$, where the sup is taken over all t between x and $x+h$ where $f(t)$ is defined. ▢

Corollary 11. Let $f : [a, b] \to E$ be continuous, let $F : [a, b] \to E$, and suppose that $F'(x) = f(x)$ for all $x \in [a, b]$. Then

$$\int_a^b f = F(b) - F(a).$$

Proof. Apply Lemma 8 to the map

$$x \mapsto F(x) - \int_a^x f.$$

◼

Corollary 12 (Integration by parts). *Let E_1, E_2, F be Banach spaces and suppose there is a continuous bilinear map $\cdot : E_1 \times E_2 \to F$. Let $f : [a, b] \to E_1$ and $g : [a, b] \to E_2$ be continuously differentiable functions. Then*

$$\int_a^b f' g + \int_a^b f g' = f(b)g(b) - f(a)g(a).$$

Proof. We have $(fg)' = f'g + fg'$ by the product rule, so integrating both sides from a to b and applying Corollary 11 produces the result. ◼

Mean value inequalities

Let $\alpha : [a, b] \to L(E, F)$ be a continuous map into the space of linear maps from E to F. If $x \in [a, b]$ and $y \in E$ then we write $\alpha(x)y$ for the element $\alpha(x)(y) \in F$.

Lemma 13. *Let $\alpha : [a, b] \to L(E, F)$ be a continuous map and let $y \in E$. Then*

$$\int_a^b \alpha(t)y \, dt = \left(\int_a^b \alpha(t) \, dt \right) y.$$

Proof. The map $\lambda \mapsto \lambda(y)$ is a continuous linear map from $L(E, F)$ to F, so the result follows from the general fact that

$$\varphi \int_X f = \int_X \varphi \circ f$$

whenever φ is a continuous linear map between Banach spaces. ◼

Theorem 14 (Mean value theorem). *Let $A \subseteq E$ be an open set, let $f : A \to F$ be continuously differentiable, let $x \in A$, and let $v \in E$. If the line segment $x + tv$ with $0 \le t \le 1$ is contained in A, then*

$$f(x + v) - f(x) = \int_0^1 f'(x + tv)v \, dt = \left(\int_0^1 f'(x + tv) \, dt \right) v.$$

Proof. Let $g(t) = f(x + tv)$ so that $g'(t) = f'(x + tv)v$. By the fundamental theorem of calculus, we have

$$g(1) - g(0) = \int_0^1 g'.$$

Since $g(0) = f(x)$ and $g(1) = f(x + v)$, the result follows and we can apply Lemma 13. ◼

The mean value theorem shows that the change in $f(x)$ is determined by its derivative and the change in x. This is made precise in the corollary below.

Corollary 15 (Mean value inequality). *Let $A \subseteq E$ be an open set, let $f : A \to F$ be continuously differentiable, and let $x, y \in A$. If the line segment between x and y is contained in A, then*

$$|f(y) - f(x)| \le |y - x| \sup_u |f'(u)|,$$

where the sup is taken over all u in the line segment. If $z \in A$, then

$$|f(y) - f(x) - f'(z)(y - x)| \le |y - x| \sup_u |f'(u) - f'(z)|,$$

with the sup as above.

Proof. We have

$$|f(y) - f(x)| = \left| \int_0^1 f'(x + t(y - x))(y - x) \, dt \right|$$

$$\le |y - x|(1 - 0) \sup_{t \in [0,1]} |f'(x + t(y - x))|,$$

which proves the first statement. Then apply this result to the map defined by $g(v) = f(v) - f'(z)v$ to obtain the second statement. ◼

Corollary 16 (Lipschitz estimate for C^1 maps). *Let $A \subseteq E$ be a convex open set and let $f : A \to F$ be continuously differentiable. If there is a constant M such that $|f'(x)| \le M$ for all $x \in A$, then*

$$|f(x) - f(y)| \le M|x - y|$$

for all $x, y \in A$.

This allows us to generalize Lemma 9.

Corollary 17. *Let $A \subseteq E$ be a connected open set and suppose that the derivative of $f: A \to F$ is zero on A. Then f is constant.*

Proof. If $x \in A$ and $B_r(x)$ is any open ball around x contained in A then Corollary 16 shows that f is constant on $B_r(x)$. Since A is connected, f is constant on A. ▢

Next time, we will look at higher derivatives and the symmetries that arise.

wj32

View all posts by wj32 →

1 Comment

andre
April 11, 2016 at 4:09 pm

Thanks man, this is helpful.