Analysis II: Derivatives in Several Variables

Jesse Ratzkin

September 21, 2009

These notes introduce the notion of derivatives for functions of several variables.

It is worthwhile to first recall the derivative of a real-valued function of a single variable. Let $f : \mathbb{R} \to \mathbb{R}$, and fixed $x_0 \in \mathbb{R}$. The classical definition of f having a derivative $f'(x_0)$ at x_0 is that the limit

$$f'(x_0) = \lim_{x \to x_0} \left(\frac{f(x) - f(x_0)}{x - x_0} \right)$$

exists. Denoting the limit as $a = f'(x_0)$, we can rewrite this as

$$\lim_{x \to x_0} \left(\frac{f(x) - f(x_0) - a(x - x_0)}{x - x_0} \right) = 0.$$

Notice that any linear map from $\mathbb{R}$ to itself is just multiplication by some constant a. This is the important conceptual leap we must make: the derivative of a function at a point is not a number, or a vector. It is a **linear transformation**. The derivative of a function f at the point x_0 is the linear transformation which best approximates f near x_0. Indeed, you're already familiar with this concept from Taylor's theorem, but might not have thought about it using the language above.

Now we naturally arrive at the proper definition of a derivative:

Definition 1. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $x_0 \in \mathbb{R}^n$. We say f has a derivative at x_0 if there is some linear map $A : \mathbb{R}^n \to \mathbb{R}^m$ such that*

$$\lim_{x \to x_0} \left(\frac{f(x) - f(x_0) - A(x - x_0)}{|x - x_0|} \right) = 0.$$

If this limit exists for the linear map A, we write $Df|_{x_0} = A$ as the (total) derivative of f at the point x_0.

We can rewrite this expression as, near x_0,

$$f(x) = f(x_0) + A(x - x_0) + r(x),$$

where

$$A : \mathbb{R}^n \to \mathbb{R}^m \text{ is linear, and } \lim_{x \to x_0} \left(\frac{|r(x - x_0)|}{|x - x_0|} \right) = 0.$$

We call $r(x - x_0)$ the sublinear remainder of f at x_0. It will be convenient later to define the auxiliary remainder function $\bar{r} : [0, \infty) \to [0, \infty)$ by

$$\bar{r}(\rho) = \sup_{|x-x_0| \leq \rho} \frac{|f(x) - f(x_0) - Df|_{x_0}(x - x_0)|}{|x - x_0|}.$$

Notice that $-\bar{r}(|x - x_0|) \leq r(x - x_0) \leq \bar{r}(|x - x_0|)$, and we still have the estimate

$$\lim_{x \to x_0} \frac{\bar{r}(|x - x_0|)}{|x - x_0|} = 0.$$

Also, not that $\bar{r}(\rho)$ is a monotone function of ρ, so that for $\rho^* > \rho > 0$ we have $\bar{r}(\rho^*) \geq \bar{r}(\rho)$.

We emphasize here that it is important to remember that the derivative of f at the point x_0 is not a number or a vector, it is the **linear transformation** which best approximates f near x_0.

Proposition 1. *If f is differentiable at x_0 then it is also continuous.*

Proof. Write $x = x_0 + v$, and use the notation above. Then

$$\begin{aligned} \lim_{x \to x_0} |f(x) - f(x_0)| &= \lim_{v \to 0} |f(x_0 + v) - f(x_0)| \\ &= \lim_{v \to 0} |f(x_0) + A(v) + r(v) - f(x_0)| \\ &\leq \lim_{v \to 0} (\|A\||v| + |r(v)|) = 0 \end{aligned}$$

We conclude $\lim_{x \to x_0}(f(x)) = f(x_0)$. □

At this point we state some basic rules of differentiation.

Theorem 2. *Let f and g be differentiable functions.*

1. *$D(f + cg) = Df + cDg$ for all real numbers c.*
2. *$D(g \circ f) = (Dg) \circ (Df)$*
3. *$D(f \cdot g) = Df \cdot g + f \cdot Dg$ (Notice the order of operations!)*

The second two statements require some clarification. In the second statement, we consider $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}^k$. Then the composition maps $\mathbb{R}^n$ to $\mathbb{R}^k$, and so does it's derivative. The theorem states that the derivative of a composition is the composition of derivatives. (This statement is certainly too elegant to not be true!) In the third statement, the product in question is a bilinear operation we will explain inside the proof.

Proof. We begin with property 1. Choosing a basepoint x_0, we see

$$\begin{aligned}
\lim_{x\to x_0} & \left[\frac{|(f(x)+cg(x))-(f(x_0)+cg(x_0)+(Df|_{x_0}+c\,Dg|_{x_0})(x-x_0)|}{|x-x_0|}\right] \\
&= \lim_{x\to x_0}\left[\frac{|(f(x)-Df|_{x_0}(x-x_0)+c(g(x)-Dg|_{x_0}(x-x_0))|}{|x-x_0|}\right] \\
&\le \lim_{x\to x_0}\left[\frac{|f(x)-f(x_0)-Df|_{x_0}(x-x_0)|}{|x-x_0|}\right]+|c|\lim_{x\to x_0}\left[\frac{|g(x)-g(x_0)-Dg|_{x_0}(x-x_0)|}{|x-x_0|}\right] \\
&= 0.
\end{aligned}$$

This proves the linearity of the derivative.

Next we prove the chain rule, property 2. For simplicity of notation, we write f and g in Taylor series expansions:

$$f(x)=f(x_0)+A(x-x_0)+r_f(x-x_0), \qquad g(x)=g(x_0)+B(x-x_0)+r_g(x-x_0).$$

Here A and B are the linear transformations which are the derivatives $Df|_{x_0}$ and $Dg|_{x_0}$ respectively, and the remainder terms for f and g satisfy

$$\lim_{x\to x_0}\frac{|r_f(x-x_0)|}{|x-x_0|}=0=\lim_{x\to x_0}\frac{|r_g(x-x_0)|}{|x-x_0|}.$$

Letting

$$y=f(x), \quad y_0=f(x_0), \quad v=x-x_0, \quad w=A(v)+r_f(v),$$

we can write

$$\begin{aligned}
g\circ f(x) &= g\circ f(x)=g(y) \\
&= g(y_0+A(v)+r_f(v)) \\
&= g(y_0)+B\circ A(v)+B(r_f(v))+r_g(w).
\end{aligned}$$

It remains to show that the two remainder terms $B(r_f(v))$ and $r_g(w)$ are smaller than linear. First,

$$\lim_{v\to 0}\frac{|B(r_f(v))|}{|v|}\le\lim_{v\to 0}\frac{\|B\||r_f(v)|}{|v|}\le\lim_{v\to 0}\frac{\|B\|\bar{r}_f(|v|)}{|v|}=0.$$

Next, we decompose $w=A(v)+r_f(v)$, so that

$$\begin{aligned}
\lim_{v\to 0}\frac{|r_g(w)|}{|v|} &= \lim_{v\to 0}\frac{|r_g(A(v)+r_f(v))|}{|v|}\le\lim_{v\to 0}\frac{\bar{r}_g(|A(v)+r_f(v)|)}{|v|} \\
&\le \lim_{v\to 0}\frac{\bar{r}_g(\|A\||v|+\bar{r}_f(|v|))}{|v|}=0.
\end{aligned}$$

Putting this all together, we have

$$\lim_{x\to x_0}\frac{|g\circ f(x)-g\circ f(x_0)-B\circ A(x-x_0)|}{|x-x_0|}\le\lim_{v\to 0}\frac{|B(r_f(v))+r_g(w)|}{|v|}=0,$$

and so $B \circ A$ is the linear transformation which best approximates $g \circ f$ at x_0 as claimed.

To prove 3, we first explain the product. on the real line, this is the usual product of numbers, but on a higher dimensional vector space there are many ways to interpret a product. In this case we mean a bilinear map.

Let V, W, Z all be vector spaces. A bilinear map $\cdot : V \times W \to Z$ is any map which is linear in each factor. More precisely, if we fix $v \in V$ then the map $w \mapsto v \cdot w : W \to Z$ is linear. Similarly, if we fix $w \in W$ then the map $v \mapsto v \cdot w : V \to Z$ is also linear. Again, pay attention to the order of operations; a bilinear map doesn't need to be symmetric. Examples of bilinear maps include the dot product and matrix multiplication. For the proper statement of 3, we write the bilinear form as $\beta(v, w) = v \cdots w$. Then we have

$$D\beta(f,g)|_x(v) = \beta(Df|_x(v), g(x)) + \beta(f(x), Dg|_x(v)).$$

To prove this statement we again write the first order Taylor expansion of f and g, with $Df|_x = A$ and $Dg|_x = B$. Then

$$\begin{aligned} \beta(f(x+v), g(x+v)) &= \beta(f(x) + A(v) + r_f(v), g(x) + B(v) + r_g(v)) \\ &= \beta(f(x), g(x)) + \beta(A(v), g(x)) + \beta(f(x), B(v)) \\ &\quad + \beta(f(x), r_g(v)) + \beta(A(v), B(v) + r_g(v)) \\ &\quad + \beta(r_f(v), g(x) + B(v) + r_g(v)). \end{aligned} \tag{1}$$

We must verify that the last three terms are sublinear in v. In order to do this, we need to think of a good way to bound a bilinear map. In this case, if $\beta : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^k$, we can also think of β as a linear map $T_\beta : \mathbb{R}^n \to \mathbb{R}^{mk} = L(\mathbb{R}^m, \mathbb{R}^k)$. That is, T_β is a linear map from $\mathbb{R}^n$ into the *space of linear maps from* $\mathbb{R}^m$ *to* $\mathbb{R}^k$. The map T_β is given by $(T_\beta(v))(w) = \beta(v, w)$. Now, T_β is a linear operator between finite-dimensional vector spaces, so it has a finite operator norm

$$\|T_\beta\| = \sup\{\|T_\beta(v)\| : |v| = 1\} = \sup\{|(T_\beta(v))(w)| : |v| = 1, |w| = 1\},$$

and so

$$|\beta(v, w)| \leq \|T_\beta\||v||w|$$

Now we can estimate the last three terms in (2). First of all,

$$\lim_{v \to 0} \frac{|\beta(f(x), r_g(x))}{|v|} \leq \lim_{v \to 0} \frac{\|T_\beta\| |f(x)| \bar{r}_g(|v|)}{|v|} = 0.$$

The other two estimates are similar.

□

As mentioned above, we often call the linear map Df the total derivative of f. From it we can recover other things like directional derivatives and partial derivatives.

Definition 2. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $v \in \mathbb{R}^n$. If $x_0 \in \mathbb{R}^n$ the directional derivative of f at x_0 in the direction v is*

$$Df|_{x_0}(v) = \lim_{t\to 0}\left(\frac{f(x_0+tv)-f(x_0)}{t}\right),$$

provided the limit exists. (Notice that the limit is a vector in $\mathbb{R}^m$.) We will typically take $v \in \mathbf{S}^{n-1}$ to be a unit vector. In the case $v = e_j$, one of the standard basis vectors of $\mathbb{R}^n$, we obtain the partial derivative

$$\frac{\partial f_i}{\partial x_j} = D(f_i)|_{x_0}(e_j),$$

where f_i is the ith component of the function f.

Proposition 3. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $v \in \mathbb{R}^n$. If the total derivative $Df|_{x_0}$ exists then so does the directional derivative, and in this case $Df|_{x_0}(v)$ is the linear transformation $Df|_{x_0}$ applied to the vector v.*

Proof. Suppose $Df|_{x_0}$ exists, and denote it as A. Then

$$\begin{aligned}\lim_{t\to 0}\left(\frac{|f(x_0+tv)-f(x_0)|}{|t|}\right) - A(v) &= \lim_{t\to 0}\left(\frac{|f(x_0+tv)-f(x_0)-A(tv)|}{t}\right)\\ &= \lim_{t\to 0}\left(\frac{|r(tv)|}{|t|}\right) = 0.\end{aligned}$$

□

In this case we have a nice expression for $Df|_{x_0}(v)$ in terms of the component of v and the partial derivatives of f. Write $v = \sum_{j=1}^n v_j e_j$, then by linearity

$$Df|_{x_0}(v) = \sum_{j=1}^n v_j\, Df|_{x_0}(e_j) = \sum_{j=1}^n v_j \frac{\partial f}{\partial x_j}.$$

Here we consider $\frac{\partial f}{\partial x_j}$ as a vector in $\mathbb{R}^m$, whose ith component is $\frac{\partial f_i}{\partial x_j}$.

The reverse implication does not hold. In contrast to what happens with functions of one variable, it is possible for a function of several variables to have partial derivatives, or even directional derivatives in all directions, at a point without even being continuous. Consider the real valued function of two variables

$$f(x,y) = \begin{cases} \frac{x^3}{x^2+y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0). \end{cases}$$

Away from the origin $(0,0)$ this is a perfectly smooth function. Also, for any unit vector $u \in \mathbf{S}^{n-1}$, the directional derivative exists. In fact, writing $u = (\cos\theta, sin\theta)$ for some angle θ, we have

$$Df|_{(0,0)} = \lim_{t\to 0}\left(\frac{t^3\cos^3\theta}{t(t^2\cos^2\theta + t^2\sin^2\theta)}\right) = \cos^3\theta.$$

So we see that the directional derivatives not only exist, they are uniformly bounded between -1 and 1. It's not too much more work to see that restricting f to any smooth curve through the origin produces a smooth function of one variable. Despite all this, f is **not** differentiable at $(0,0)$. First observe that the partial derivatives of f at $(0,0)$ are

$$\left.\frac{\partial f}{\partial x}\right|_{(0,0)} = 1, \qquad \left.\frac{\partial f}{\partial y}\right|_{(0,0)} = 0.$$

However, if $\theta = \pi/4$ then $u = (\cos\theta, \sin\theta) = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ and

$$Df|_{(0,0)}(u) = \frac{1}{\sqrt{8}} \neq \frac{1}{\sqrt{2}} = \frac{\partial f}{\partial x}\cos\theta + \frac{\partial f}{\partial y}\sin\theta.$$

One should keep the example directly above in mind to remember the difference between directional derivatives and the total derivative for for functions of several variables. Fortunately, these two coincide for functions of one variable, whether the target is one-dimensional or not. This is because, in $\mathbb{R}$, one can only take a limit in one direction, rather than several in two (or more) dimensions. Thus we see that all the rules of calculus we're used to for scalar-valued functions $f : \mathbb{R} \to \mathbb{R}$ still hold for vector-valued functions $f : \mathbb{R} \to \mathbb{R}^m$. In practice, one can write $f(t) = (f_1(t), \ldots, f_m(t))$, where $t \in \mathbb{R}$ is the independent parameter, and apply one-variable rules of calculus to each component $f_i(t)$ separately.

We can refine the notion of a differentiable function $f : \mathbb{R}^n \to \mathbb{R}^m$ by asking for the derivative to be continuous. We give the space of linear transformations $\mathbb{R}^n \to \mathbb{R}^m$ a topology from the operator norm, and ask that the derivative $Df|_{x_0}$ vary contiuously (as a function of the base point x_0) in this topology. Equivalently, we can ask that the coefficients $\frac{\partial f_i}{\partial x_j}$ of the matrix associated to the linear transformation $Df|_{x_0}$ are continuous functions. (It is worthwhile to check that these two conditions are equivalent.)

Definition 3. *We say that $f : \mathbb{R}^n \to \mathbb{R}^m$ is of class C^1 if the derivative $Df|_{x_0}$ varies continuously with respect to the base point x_0.*

We will return to the subject on C^1 functions later, and prove that $f : \mathbb{R}^n \to \mathbb{R}^m$ is C^1 if and only if its partial derivatives $\frac{\partial f_i}{\partial x_j}$ exist and are continuous. First, though, we prove some mean-value estimates which are very useful. In this discussion, we let D be a *domain*, that is an open, connected set in $\mathbb{R}^n$, and consider $f : D \to \mathbb{R}^m$. If $p, q \in D$, denote the line segment joining p to q by $[p, q]$. Recall that a domain $D \subset \mathbb{R}^n$ is convex if for every pair $p, q \in D$ the line segment $[p, q]$ is also in D.

Theorem 4. *Let $p, q \in D$ such that $[p, q] \subset D$. Then there is a linear transformation $A : \mathbb{R}^n \to \mathbb{R}^m$ such that*

$$f(q) - f(p) = A(q - p).$$

In fact,

$$A = A_{q,p} = \int_0^1 Df|_{tq+(1-t)p}\, dt.$$

Moreover, this linear transformation $A_{q,p}$ depends continuously on the endpoints p and q.

Proof. Define the vector-valued function of one variable $g(t) = f(tq + (1-t)p)$. By the fundamental theorem of calculus for functions of one variable and the chain rule,

$$f(q) - f(p) = g(1) - g(0) = \int_0^1 g'(t)dt = \int_0^1 Df|_{tq+(1-t)p}\,(q-p)dt.$$

It remains to see that $A_{q,p}$ is continuous. Take $\epsilon > 0$. The segment $[p, q]$ is compact, so Df is uniformly continuous on it. So there is $\delta > 0$ such that

$$|tq + (1-t)p - u| < \delta \Rightarrow \| Df|_{tq+(1-t)p} - Df|_u \| \leq \frac{\epsilon}{2}$$

for all $t \in [0, 1]$. Choosing x and y so that $|x - p| < \delta$ and $|y - q| < \delta$ we see

$$\| Df|_{tq+(1-t)p} - Df|_{ty+(1-t)x} \| \leq \epsilon.$$

We conclude

$$\|A_{q,p} - A_{y,x}\| \leq \int_0^1 \| Df|_{tq+(1-t)p} - Df|_{ty+(1-t)x} \| dt \leq \epsilon.$$

□

Theorem 5. *Choose p, q so that $[p, q] \subset D$ and let $M = \sup_{0\leq t\leq 1} \| Df|_{tq+(1-t)p} \|$. Then*

$$|f(q) - f(p)| \leq M|q - p|.$$

Proof. Without loss of generality, we can assume M is finite. It suffices to show $|f(p) - f(q)| \leq (M + \epsilon)|p - q|$ for any $\epsilon > 0$. Consider the set

$$X = \{x \in [p, q] : |f(y) - f(p)| \leq (M + \epsilon)|y - p| \quad \forall y \in [p, x]\}.$$

Observe both sides of the defining inequality for X are zero if $x = p$, so $p \in X$. Also, because f is continuous, X is a closed subset of the interval $[p, q]$. If we can show X is also an open subset of $[p, q]$, then we must have $X = [p, q]$ because the interval is connected. We show this by examining the first order Taylor expansion of f at any $x_0 \in X$. Indeed,

$$|f(x) - f(x_0)| \leq | Df|_{x_0}\,(x - x_0)| + \bar{r}_f(|x - x_0|) \leq (M + \epsilon)|x - x_0|.$$

There are three possible configurations for the triple x, x_0, y in $[p, q]$ if $y \in [p, x]$: either x_0 is between x and y, x is between x_0 and y, or y is between x and x_0. (It might help to draw a picture here.) In the first two cases, $x_0 \in X$ and y comes before x_0, so $|f(y) - f(p)| \leq (M + \epsilon)|y - p|$. In the last case,

$$|f(y)-f(p)| \leq |f(y)-f(x_0)|+|f(x_0)-f(p)| \leq (M+\epsilon)(|y-x_0|+|x_0-p|) \leq (M+\epsilon)|y-p|.$$

In any case, we conclude that if $x \in [p, q]$ is near $x_0 \in X$ then $x \in X$, and so X is an open subset of the interval $[p, q]$. Because $[p, q]$ is connected, its only nonempty subset which is both open and closed is itself, and so $|f(p) - f(q)| \leq (M + \epsilon)|p - q|$ as we claimed.

□

We leave the proofs of the following three corollaries as exercises. Recall that a function is Lipschitz continuous, with Lipschitz constant L, if

$$|f(p) - f(q)| \leq L|p - q|.$$

Corollary 6. *Let f be differentiable with $\| Df|_{x_0} \| \leq M$ for all $x_0 \in D$. Then f is Lipschitz with Lipschitz constant M.*

Corollary 7. *A continuously differentiable function is locally Lipschitz.*

(Hint: Pick a basepoint p and choose a small, convex neighborhood of p. Then use the fact that if q is near q then $Df|_p$ has to be close to $Df|_q$.)

Corollary 8. *If $Df|_p = 0$ for all $p \in D$ and D is connected then f is constant on D.*

(Hint: Use the Corollary above with the best Lipschitz constant you can think of.)

Now, using some of the mean-value estimates we have just derived, we will prove the following theorem.

Theorem 9. *The function $f : \mathbb{R}^n \to \mathbb{R}^m$ is of class C^1 if and only if the partial derivatives $\frac{\partial f_i}{\partial x_j}$ exist and are continuous.*

Proof. If f is C^1, the derivative exists, and so the partial derivatives also exist. The partial derivatives are also the composition of the derivative map and a dot product, so they are continuous.

Now suppose the partial derivatives exist and are continuous and fix $x \in \mathbb{R}^n$. We have to show that, for $v \in \mathbb{R}^n$ small,

$$r = f(x + v) - f(x) - A(v)$$

is smaller than $|v|$, where A is the linear transformation whose components ar $a_{ij} = \frac{\partial f_i}{\partial x_j}$. Choose $\epsilon > 0$. By continuity, there is a $\delta > 0$ such that if $|v| < \delta$ then

$$\left| \left. \frac{\partial f_i}{\partial x_j} \right|_{x+v} - \left. \frac{\partial f_i}{\partial x_j} \right|_{x} \right| \leq \frac{\epsilon}{nm}$$

Now join x to $x + v$ by the sequence of n line segments $\sigma_j = [p_j, q_j]$, $j = 1, \ldots, n$, where $p_1 = x$, $q_j = p_j + v_j e_j$, and $p_{j+1} = q_j$. Then, using the mean value theorem, we can write

$$f_i(x+v) - f_i(x) = \sum_j f_i(q_j) - f_i(p_j) = \sum_j \left. \frac{\partial f_i}{\partial x_j} \right|_{p_j + \theta_{ij} v_j e_j} v_j,$$

for some $\theta_{ij} \in [0, 1]$. Note that, because $|v| < \delta$, the points p_j and q_j are all in the ball $B_\delta(x)$, so we conclude

$$\begin{aligned} \left| f_i(x+v) - f_i(x) - \sum_j \left. \frac{\partial f_i}{\partial x_j} \right|_x v_j \right| &= \left| \sum_j \left(\left. \frac{\partial f_i}{\partial x_j} \right|_{p_j + \theta_{ij} v_j e_j} - \left. \frac{\partial f_i}{\partial x_j} \right|_x \right) v_j \right| \\ &\leq \frac{\epsilon}{m}. \end{aligned}$$

It follows from this estimate that the derivative exists, and $Df|_x$ is the linear transformation whose components are $a_{ij} = \left. \frac{\partial f_i}{\partial f_j} \right|_x$. Moreover, these component vary continuously with x, so f is in fact C^1.

□

The same estimate leads one to a slightly more general statement. (Details of the proof are left to the reader.)

Theorem 10. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and suppose the partial derivatives exist and are continuous near x. Then f is differentiable at x. Further, if the partial derivatives of f exist and are bounded then f is locally Lipschitz.*

We'll finish off this section of notes with some brief asides. What does it mean for a function $f : \mathbb{C} \to \mathbb{C}$ to have a derivative? Recall that the complex plane $\mathbb{C}$ is bijective to the real plane $\mathbb{R}^2$ under the map $z = x + iy$, where $i^2 = -1$. However, $\mathbb{C}$ has the additional algebraic structure of multiplication, which makes many aspects of analysis on it particularly nice. Indeed, under the bijection listed above, multiplication by the complex number $a + ib$ is the same as multiplication by the 2×2 matrix $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$. As we have written before, $f : \mathbb{C} \to \mathbb{C}$ has a derivative at z_0 if there is a linear map $T : \mathbb{C} \to \mathbb{C}$ such that

$$f(z+h) = f(z) + T(h) + r(h).$$

However, the fact that T is a linear map from $\mathbb{C}$ to itself means it must be multiplication by a complex number. Writing f in real components, $f(z) = u(z) + iv(z)$, we see the the matrix of partial derivatives take the form

$$\begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}.$$

Thus we have derived the Cauchy-Riemann equations for an analytic function:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

If you take a course in complex analysis you'll that these are very powerful equations.

Finally, recall that we previously proved that matrix inversion is a continuous function. We can improve this with the following theorem.

Theorem 11. *Recall that $Gl(n, \mathbb{R}) \subset \mathbb{R}^{n^2}$ is the set of invertible linear transformations from $\mathbb{R}^n$ to itself. The inversion map* $\mathrm{Inv} : Gl(n, \mathbb{R}) \to GL(n, \mathbb{R})$ *is of class C^1.*

In fact, the same proof shows Inv has as many derivatives as you please.

We will need the following lemma.

Lemma 12. *If $A \in Gl(n, \mathbb{R})$ and $\|A\| < 1$ then $(\mathrm{Id} - A)^{-1}$ exists, and is the limit (as $k \to \infty$) of the partial sums*

$$S_k = \sum_{j=0}^{k} A^j = \mathrm{Id} + A + \cdots + A^k.$$

Proof. The identity has conorm $\mathbf{m}(\mathrm{Id}) = 1$, so we can bound the conorm

$$\mathbf{m}(\mathrm{Id} - A) \geq \mathbf{m}(\mathrm{Id} - \|A\| > 0,$$

which shows $\mathrm{Id} - A$ is invertible. Next we show S_k is a Cauchy sequence. Choose $\epsilon > 0$, and choose $N \in \mathbb{N}$ large enough so that

$$\|A\|^N < \epsilon(1 - \|A\|).$$

Because $\|A\| < 1$, the sum $\sum_{j=0}^{\infty} \|A\|^j$ is a geometric series, and converges to $(1 - \|A\|)^{-1}$. If $k > l > N$ then we estimate

$$\begin{aligned} \|S_k - S_l\| &= \|A^k + A^{k-1} + \cdots + A^{l+1}\| \leq \|A\|^k + \cdots \|A\|^{l+1} \\ &\leq \|A\|^{l+1}(1 + \cdots + \|A\|^{k-l+1}) \leq \|A\|^{l+1} \sum_{j=0}^{\infty} \|A\|^j \\ &= \frac{\|A\|^{l+1}}{1 - \|A\|} < \epsilon. \end{aligned}$$

The space $\mathbb{R}^{n^2}$ is a finite-dimensional normed vector space, so it is complete, and thus the series $\sum_{j=0}^{\infty} A^j$ converges to some linear transformation S. We want to show $S = (\mathrm{Id} - A)^{-1}$. However, we can telescope the sum to get

$$S_k \circ (\mathrm{Id} - A) = (\mathrm{Id} - A) \circ S_k = \mathrm{Id} - A^{k+1}.$$

(Write this out if you don't see it.) Also, $\|A^{k+1}\| \leq \|A\|^{k+1} \to 0$, so, letting $k \to \infty$ we see

$$S \circ (\mathrm{Id} - A) = (\mathrm{Id} - A) \circ S = \mathrm{Id}$$

as claimed. □

Now we prove that matrix inversion is C^1.

Proof. If $A \in Gl(n, \mathbb{R})$ and $V \in \mathbb{R}^{n^2}$ is small, we want to prove the Taylor estimate

$$\mathrm{Inv}(A + V) = \mathrm{Inv}(A) + L(V) + r((V),$$

where $L : Gl(n, \mathbb{R}) \to \mathbb{R}^{n^2}$ is a linear transformation and r is sublinear. It will help to factor $(A + V)^{-1} = A^{-1}(\mathrm{Id} + VA^{-1})^{-1}$, which in particular (by the lemma) implies $A + V$ is invertible for V sufficiently small. Furthermore,

$$(A + V)^{-1} = A^{-1} \left(\sum_{j=0}^{\infty} (-VA)^j \right) = A^{-1} - A^{-1} \circ V \circ A^{-1} + r(V),$$

where in the last expression we have retained only the first two terms of the series expansion. The map $V \mapsto -A^{-1} \circ V \circ A^{-1}$ is the linear function of V we're looking for, and our only remaining task is to show the remainder $r(V) = A^{-1} \left(\sum_{j=2}^{\infty} (-VA)^j \right)$ is smaller than V. In fact,

$$\|r(V)\| \leq \|A^{-1}\| \sum_{j=2}^{\infty} (\|V\|\|A^{-1}\|)^j = \|A^{-1}\| \frac{(\|V\|\|A^{-1}\|)^2}{1 - \|V\|\|A^{-1}\|},$$

so

$$\lim_{V \to 0} \frac{\|r(V)\|}{\|V\|} \leq \lim_{V \to 0} \frac{\|V\|\|A^{-1}\|^3}{1 - \|V\|\|A^{-1}\|} = 0.$$

Finally, we need to prove the map $V \mapsto -A^{-1} \circ V \circ A^{-1}$ depends continuously on A. We already know Inv is a continuous function, so if $A_k \to A$ then $A_k^{-1} \to A^{-1}$. Thus

$$\begin{aligned} \| - A_k^{-1} \circ V \circ A_k^{-1} + A^{-1} \circ V \circ A^{-1} \| &\leq \|(A^{-1} - A_k^{-1}) \circ V \circ A^{-1}\| + \|A_k^{-1} \circ V \circ (A^{-1} - A_k^{-1})\| \\ &\leq \|A^{-1} - A_k^{-1}\|(\|A^{-1}\| + \|A_k^{-1}\|) \to 0. \end{aligned}$$

□

Some comments are in order. In the case of $n = 1$, the formula for matrix inversion reverts to the familiar

$$\frac{d}{dx}\left(\frac{1}{x}\right) = -\frac{1}{x^2}.$$

In higher dimensions the formula is more complicated, and one must pay close attention to the correct order of operations. This order of operations becomes more important for higher powers. As an exercise, you might determine whether $(A^2)^{-1} = (A^{-1})^2$ for linear transformations from $\mathbb{R}^n$ to itself (equivalently, $n \times n$ matrices). Start by thinking about 2×2 matrices.

One can use series to define many familiar functions of matrices. For instance, the exponential of an $n \times n$ matrix A is given by the familiar power series

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}.$$

It's not too hard to show this series is absolutely and uniformly convergent (in fact, one can pretty much copy the $n = 1$ proof), and so one can even differentiate the power series using the familiar formula. However, usually $e^{A+B} \neq e^A \cdot e^B$ for matrices. If you start to expand out the corresponding series, you'll find the right condition to force $e^{A+B} = e^A \cdot e^B$. Again, one must pay close attention to the order of multiplication.

Here are some more exercises. If you're doing a lot of work for any of these, stop! Each one is actually very quick if you understand what's going on.

1. Let $f : \mathbb{R}^3 \to \mathbb{R}^2$ be orthogonal projection onto the horizontal plane. What is the derivative $Df|_x$? Does it depend on the base point x?

2. Fix $v = (1, 0, 0, 1, -2, 1) \in \mathbb{R}^6$ and let $f(x) : \mathbb{R}^6 \to \mathbb{R}$ be given by $f(x) = \langle x, v \rangle$. What is the derivative Df?

3. Fix any 3×4 matrix A and consider the function $f : \mathbb{R}^{4\times 2} \to \mathbb{R}^{3\times 2}$ given by $f(B) = A \cdot B$. What is the derivative Df?