33. Properties of multivariate derivative

Here we proceed to prove the standard properties of the derivative of multivariate functions of multivariate variables and then demonstrate its use.

33.1 Rules for multivariate derivatives.

We start with the simple fact:

Lemma 33.1 (Sum rule) *Let $f,g\colon \mathbb{R}^n \to \mathbb{R}^m$ be differentiable at some $x \in \operatorname{int}(\operatorname{Dom}(f)) \cap \operatorname{int}(\operatorname{Dom}(g))$. Then so is $f+g$ with*

$$D(f+g)(x) = Df(x) + Dg(x) \tag{33.1}$$

Similarly, for each $\lambda \in \mathbb{R}$ and $x \in \operatorname{int}(\operatorname{Dom}(f))$ such that f is differentiable at x, so is λf with

$$D(\lambda f)(x) = \lambda Df(x) \tag{33.2}$$

Proof. Differentiability of $f+g$ along with the formula (33.1) follows from

$$\begin{aligned}\big\|(f+g)(z)-&(f+g)(x) - Df(x)(z-x) - Dg(x)(z-x)\big\| \\ &\leqslant \big\|f(z)-f(x)-Df(x)(z-x)\big\| + \big\|g(z)-g(x)-Dg(x)(z-x)\big\|\end{aligned} \tag{33.3}$$

and the definition of the total derivative in (32.2). The proof for λf is a consequence of the homogeneity of the norm. □

A consequence of the lemma is that the map $f \mapsto Df$ is linear in the set of functions that are defined and differentiable in a given open set $U \subseteq \mathbb{R}^n$.

Next we address the product rule which we state in the simplest form of a product that takes two vector-valued functions f and g of possibly different dimensions and outputs a scalar. Since the product is *ex definitio* also required to be linear in each coordinate, it must take the form

$$x \mapsto f(x) \cdot A(x)g(x) \tag{33.4}$$

for some matrix function $x \mapsto A(x)$ of the corresponding dimensions. We will for simplicity only treat the case when $A(x)$ is a constant matrix.

Lemma 33.2 (Product rule) *Given $k,m,n \geqslant 1$, let $f\colon \mathbb{R}^n \to \mathbb{R}^m$ and $g\colon \mathbb{R}^n \to \mathbb{R}^k$ be differentiable at some $x \in \operatorname{int}(\operatorname{Dom}(f)) \cap \operatorname{int}(\operatorname{Dom}(g))$. Let $A \in \mathcal{M}_{m,k}$. Then the $\mathbb{R}$-valued function $z \mapsto f(z)\cdot Ag(z)$ is differentiable at x and*

$$D\big(f\cdot Ag\big)(x) = g(x)\cdot A^{\mathrm{T}}(Df)(x) + f(x)\cdot A(Dg)(x) \tag{33.5}$$

Proof. Proceeding as in the scalar case, we write the relevant expression as

$$\begin{aligned}\big(f\cdot Ag\big)(z) - \big(f\cdot Ag\big)(x) &- \big[g(x)\cdot A^{\mathrm{T}}(Df)(x) + f(x)\cdot A(Dg)(x)\big](z-x) \\ &= f(z)\cdot A\big[g(z)-g(x)-(Dg)(x)(z-x)\big] \\ &\quad + g(x)\cdot A^{\mathrm{T}}\big[f(z)-f(x)-Df(x)(z-x)\big]\end{aligned} \tag{33.6}$$

where we used that $f(z)\cdot Ag(x) = g(x)\cdot A^{\mathrm{T}}f(z)$. Since f is continuous at x, the function $z \mapsto \|f(z)\|$ is bounded by some $M > 0$ in a neighborhood of x. It follows the absolute value of the right-hand side is at most

$$M\|A\|\big\|g(z)-g(x)-(Dg)(x)(z-x)\big\| + \|g(x)\|\|A^{\mathrm{T}}\|\big\|f(z)-f(x)-Df(x)(z-x)\big\| \tag{33.7}$$

Dividing by $\|z-x\|$ and taking $z \to x$, both terms tend to zero. $\square$

It is interesting to note how the product rule works in coordinates. Denoting the product by $h(x) := f(x) \cdot Ag(x)$, we have

$$h(x) = \sum_{i=1}^{m}\sum_{j=1}^{k} f_i(x)A_{ij}g_j(x) \tag{33.8}$$

and, for all $r = 1,\dots,n$,

$$\frac{\partial h}{\partial x_r}(x) = \sum_{i=1}^{m}\sum_{j=1}^{k}\left[\frac{\partial f_i}{\partial x_r}(x)A_{ij}g_j(x) + f_i(x)A_{ij}\frac{\partial g_j}{\partial x_r}(x)\right] \tag{33.9}$$

This is now readily check to wrap into the form in (33.5).

Finally, we will address compositions of functions:

Lemma 33.3 (Chain rule) *Given $k,m,n \geqslant 1$, let $f\colon \mathbb{R}^n \to \mathbb{R}^m$ and $g\colon \mathbb{R}^m \to \mathbb{R}^k$ be such that, for some $x \in \operatorname{int}(\operatorname{Dom}(f))$ with $f(x) \in \operatorname{int}(\operatorname{Dom}(g))$, the function f is differentiable at x and g is differentiable at $f(x)$. Then $g \circ f$ is differentiable at x and*

$$D(g\circ f)(x) = Dg\bigl(f(x)\bigr)Df(x) \tag{33.10}$$

Proof. We proceed very much like in the single variable case. The assumed differentiability of g gives us a function $u_{f(x)}\colon \mathbb{R}^m \to \mathcal{M}_{k,m}$ such that

$$g(y) - g\bigl(f(x)\bigr) = \bigl[Dg(f(x)) + u_{f(x)}(y)\bigr]\bigl(y - f(x)\bigr) \tag{33.11}$$

for all $y \in \operatorname{Dom}(g)$ with $\|u_{f(x)}(y)\| \to 0$ as $y \to f(x)$. The assumed differentiability of f in turn gives us a function $\tilde{u}_x\colon \mathbb{R}^n \to \mathcal{M}_{m,n}$ such that

$$f(z) - f(x) = \bigl[Df(x) + \tilde{u}_x(z)\bigr](z-x) \tag{33.12}$$

for all $z \in \operatorname{Dom}(f)$ with $\|\tilde{u}_x(z)\| \to 0$ as $z \to x$. Plugging $y = f(z)$ into (33.11) and using (33.12) then yields

$$\begin{aligned} g\bigl(f(z)\bigr) - g\bigl(f(x)\bigr) &= \bigl[Dg(f(x)) + u_{f(x)}(f(z))\bigr]\bigl[Df(x) + \tilde{u}_x(z)\bigr](z-x) \\ &= \bigl[Dg(f(x))Df(x) + \hat{u}_x(z)\bigr](z-x) \end{aligned} \tag{33.13}$$

where

$$\hat{u}_x(z) := \bigl[Dg(f(x)) + u_{f(x)}(f(z))\bigr]\tilde{u}_x(z) + u_{f(x)}\bigl(f(z)\bigr)Df(x) \tag{33.14}$$

The rules for working with norms give

$$\|\hat{u}_x(z)\| \leqslant \Bigl(\|Df(x)\| + \|u_{f(x)}(f(z))\|\Bigr)\|\tilde{u}_x(z)\| + \|u_{f(x)}(f(z))\|\|Df(x)\| \tag{33.15}$$

By Lemma 32.4, differentiability of f at x implies its continuity and so $f(z) \to f(x)$ as $z \to x$. The fact that $\|u_{f(x)}(y)\| \to 0$ as $y \to f(x)$ then implies that $z \mapsto \|u_{f(x)}(f(z))\|$ is bounded on a neighborhood of x and tends to zero as $z \to x$. Since also $\|\tilde{u}_x(z)\| \to 0$ as $z \to x$, we have that $\|\hat{u}_x(z)\| \to 0$ as $z \to x$. This proves the claim. $\square$

Writing this using functions $x \mapsto f(x)$ and $y \mapsto g(y)$, we can cast (33.10) as

$$\frac{\partial (g\circ f)_i}{\partial x_j}(x) = \sum_{\ell=1}^{m}\frac{\partial g_i}{\partial y_\ell}\bigl(f(x)\bigr)\frac{\partial f_\ell}{\partial x_j}(x) \tag{33.16}$$

for all $i = 1, \dots, k$ and all $j = 1, \dots, n$.

33.2 Inverse function rule.

The last remaining differentiation "rule" concerns invertible functions. Here we focus on functions between Euclidean spaces of the same dimension because (by Invariance of Dimension theorems) there are, in fact, no continuous injective maps with a continuous inverse between open subsets of Euclidean spaces of different dimension.

Lemma 33.4 (Inverse function rule) *Given $n \geqslant 1$, let $f\colon \mathbb{R}^n \to \mathbb{R}^n$ and $x \in \operatorname{int}(\operatorname{Dom}(f))$ be such that $f(x) \in \operatorname{int}(\operatorname{Ran}(f))$. Assume that f is injective on its domain and differentiable at x with the matrix $Df(x)$ is invertible. If also f^{-1} is continuous at $f(x)$, then f^{-1} is differentiable at $f(x)$ and*

$$D(f^{-1})\big(f(x)\big) = Df(x)^{-1} \tag{33.17}$$

A key point of the proof of Lemma 33.4 is to prove differentiability of f^{-1} at $f(x)$. The proof we give does this along with proving the formula (33.17) which can be independently checked by the Chain Rule. The proof requires the following facts about the matrix norm of inverse matrices which we will prove first:

Lemma 33.5 *Given $n \geqslant 1$, let $A, B \in \mathcal{M}_{n,n}$ be such that A^{-1} exists. Then*

$$\|B\|\|A^{-1}\| < 1 \;\Rightarrow\; (A+B)^{-1} \text{ exists } \wedge\; \big\|(A+B)^{-1}\big\| \leqslant \frac{\|A^{-1}\|}{1-\|B\|\|A^{-1}\|} \tag{33.18}$$

Under the condition on the left, we then also get

$$\big\|(A+B)^{-1} - A^{-1}\big\| \leqslant \frac{\|A^{-1}\|^2}{1-\|B\|\|A^{-1}\|}\|B\| \tag{33.19}$$

In particular, the set of invertible matrices in $\mathcal{M}_{n,n}$ is open and the map $A \mapsto A^{-1}$ is continuous (in the topology induces by the matrix norm) on this set.

Proof. Writing $A + B = A(1 + A^{-1}B)$, for invertibility of $A + B$ it suffices to prove invertibility of $1 + A^{-1}B$. Here we observe that, for any $x \in \mathbb{R}^n$,

$$\begin{aligned} \big\|(1 + A^{-1}B)x\big\| &= \|x + A^{-1}Bx\| \\ &\geqslant \|x\| - \|A^{-1}Bx\| \geqslant \|x\|\big(1 - \|B\|\|A^{-1}\|\big) \end{aligned} \tag{33.20}$$

Under the condition $\|B\|\|A^{-1}\| < 1$ we get that $(1 + A^{-1}B)x = 0$ implies $x = 0$, proving that $1 + A^{-1}B$ is invertible. Using facts from linear algebra which we do not go into here, this means that the map $x \mapsto (1 + A^{-1}B)x$ is onto. Hence, for each $z \in \mathbb{R}^n$ there exists a unique $x \in \mathbb{R}^n$ such that $z = (1 + A^{-1}B)x$. Using this in (33.20) shows

$$\big\|(1 + A^{-1}B)^{-1}z\big\| \leqslant \frac{1}{1-\|B\|\|A^{-1}\|}\|z\| \tag{33.21}$$

With $1 + A^{-1}B$ invertible we have $(A + B)^{-1} = (1 + A^{-1}B)^{-1}A^{-1}$ and using (32.26) we then get (33.18). For the second part of the claim we note that

$$(A+B)^{-1} - A^{-1} = -(A+B)^{-1}BA^{-1} \tag{33.22}$$

Using the "product rule" (32.26) for the matrix norm, we get (33.19). □

Proof of Lemma 33.4. Let $x \in \operatorname{int}(\operatorname{Dom}(f))$ be such that $f(x) \in \operatorname{int}(\operatorname{Ran}(f))$. Differentiability of f at x implies

$$f(z) - f(x) = \big[Df(x) + u_x(z)\big](z - x) \tag{33.23}$$

with $\|u_x(z)\| \to 0$ as $z \to x$. Since $Df(x)$ is invertible, it follows that there exists $\delta > 0$ such that $B(x, \delta) \subseteq \operatorname{Dom}(f)$ and

$$\forall z \in B(x,\delta)\colon\ \ \|u_x(z)\|\big\|Df(x)^{-1}\big\| < 1/2 \tag{33.24}$$

The continuity of f^{-1} at $f(x)$ in turn implies that there exists $\epsilon > 0$ such that

$$B\big(f(x), \epsilon\big) \subseteq f\big(B(x,\delta)\big) \tag{33.25}$$

For all $y \in B(f(x), \epsilon)$ we thus have $z := f^{-1}(y) \in B(x, \delta)$ and, by Lemma 33.5 and (33.24), the matrix $Df(x) + u_x(z)$ is invertible. Plugging $z := f^{-1}(y)$ into (33.23) then gives

$$\begin{aligned} f^{-1}(y) - f^{-1}\big(f(x)\big) &= \big[Df(x) + u_x(f^{-1}(y))\big]^{-1}\big(y - f(x)\big) \\ &= \big[Df(x)^{-1} + \hat{u}_x(y)\big]\big(y - f(x)\big) \end{aligned} \tag{33.26}$$

where

$$\hat{u}_x(y) := \big[Df(x) + u_x(f^{-1}(y))\big]^{-1} - Df(x)^{-1} \tag{33.27}$$

The estimate (33.19) enabled by (33.24) now gives

$$\big\|\hat{u}_x(y)\big\| \leqslant 2\big\|Df(x)^{-1}\big\|^2\big\|u_x(f^{-1}(y))\big\| \tag{33.28}$$

Since the continuity of f^{-1} at $f(x)$ implies that $f^{-1}(y) \to x$ as $y \to f(x)$, we get that $\|u_x(f^{-1}(y))\| \to 0$ as $y \to f(x)$. But this means that $\|\hat{u}_x(y)\| \to 0$ as $y \to f(x)$, which proves the the claim via Lemma 32.8. □

To give an example, let $f\colon (0,\infty)^2 \to (0,\infty)^2$ be defined by

$$f(x,y) := \begin{pmatrix} xy \\ y/x \end{pmatrix} \tag{33.29}$$

Then f is differentiable with

$$Df(x,y) = \begin{pmatrix} y & x \\ -y/x^2 & 1/x \end{pmatrix} \tag{33.30}$$

For this matrix we get

$$\det(Df)(x,y) = 2y/x \tag{33.31}$$

which is non-zero, and so $Df(x,y)$ is invertible, for $x, y > 0$. The inverse of f can be computed by inverting the substitution

$$\begin{aligned} u &:= xy \\ v &:= y/x \end{aligned} \tag{33.32}$$

as

$$\begin{aligned} x &= \sqrt{u/v} \\ y &= \sqrt{uv} \end{aligned} \tag{33.33}$$

which means that f^{-1} exists and

$$f^{-1}(u,v) = \begin{pmatrix} \sqrt{u/v} \\ \sqrt{uv} \end{pmatrix} \tag{33.34}$$

This inverse function is defined on all of $(0,\infty)^2$ and is continuous there so the assumptions of Lemma 33.4 are satisfied. A calculation shows that

$$(Df)^{-1}(x,y) = \frac{1}{2y/x}\begin{pmatrix} 1/x & -x \\ y/x^2 & y \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1/y & x^2/y \\ 1/x & x \end{pmatrix} \tag{33.35}$$

An explicit differentiation in turn gives

$$D(f^{-1})(u,v) = \frac{1}{2}\begin{pmatrix} 1/\sqrt{uv} & -\sqrt{u/v^3} \\ \sqrt{v/u} & \sqrt{u/v} \end{pmatrix} \tag{33.36}$$

which is easily checked to coincide with (33.35) under the substitution (33.32), proving the formula (33.17) by explicit calculation.