32. Differentiation of multivariate functions

We will now move to discuss calculus of multivariate functions of multiple variables focusing on aspects related to derivatives.

32.1 Definitions and basic facts.

Multivariate calculus is concerned with m-tuples of functions

$$f_1(x_1,\dots,x_n),\dots,f_m(x_1,\dots,x_n) \tag{32.1}$$

of n-tuples of variables $x_1,\dots,x_n$. We will typically write this as a map $f\colon \mathbb{R}^n \to \mathbb{R}^m$ which makes the expression in specific coordinates a matter of choice.

The linear structure of $\mathbb{R}^n$ and $\mathbb{R}^m$ allows us to add coordinates as well as functions, provided the corresponding dimensions match. We can also multiply coordinates and function by scalars and thus take linear combinations thereof. We will think of $\mathbb{R}^n$ and $\mathbb{R}^m$ as metric spaces with the metric induced by the Euclidean norm that we denote by $\|\cdot\|$ regardless of the underlying dimension. In particular, the symbol $z \to x$ means that $\|z-x\| \to 0$ while $f(z) \to f(x)$ means $\|f(z)-f(x)\| \to 0$.

In single variable calculus, we defined the derivative as a limit (of the slope of the secant lines) and then observed that the existence of the derivative is equivalent to the existence of a linear approximation. As we will see, this equivalence becomes problematic in the multivariable case and so we proceed directly via the latter property:

Definition 32.1 *Let $f\colon \mathbb{R}^n \to \mathbb{R}^m$ and $x \in \operatorname{int}(\operatorname{Dom}(f))$. We say that f* is differentiable *at x if there exists a $m \times n$-matrix A such that*

$$\lim_{z\to x} \frac{\|f(z)-f(x)-A(z-x)\|}{\|z-x\|} = 0 \tag{32.2}$$

If such an A exists, it is unique and we denote it $Df(x)$. We call $Df(x)$ the total derivative, or total differential of f at x.

Note that, in light of equivalence of all norms on Euclidean spaces, (32.2) does not depend on the choice of the norm in $\mathbb{R}^n$ or $\mathbb{R}^m$. The reader has likely encountered differentiation of multivariate functions in the following form:

Definition 32.2 *Let $f\colon \mathbb{R}^n \to \mathbb{R}^m$ and $x \in \operatorname{int}(\operatorname{Dom}(f))$. For each $i = 1,\dots,m$ and $j = 1,\dots,n$, the* partial derivative *of f_i with respect to x_j at x is the limit*

$$\frac{\partial f_i}{\partial x_j}(x) := \lim_{h\to 0} \frac{f_i(x+he_j)-f_i(x)}{h} \tag{32.3}$$

where e_j is the j-th coordinate vector in $\mathbb{R}^n$. Similarly, given $v \in \mathbb{R}^n$, the directional derivative *of f_i in direction v is the limit*

$$\frac{\partial f_i}{\partial v}(x) := \lim_{h\to 0^+} \frac{f_i(x+hv)-f_i(x)}{h} \tag{32.4}$$

where the limit is only through positive h.

The i-th partial derivative, if it exists, coincides with the directional derivative in direction e_i as well as the negative of the directional derivative in direction $-e_i$ (which thus both exist). As the example of $f\colon \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x,y) := \sqrt{x^2+y^2} \tag{32.5}$$

shows, one may have directional derivatives without having partial derivatives. (Indeed, here $\frac{\partial f_i}{\partial v}(0,0) = 1$ which does not reverse sign when v does.) This is no surprise; after all, the directional derivative is the counterpart of the right and left derivative for functions of one variable.

Various alternative notations exist; e.g., $\partial_j f_i$ for $\frac{\partial f_i}{\partial x_j}$ and $D_v f$ for $\frac{\partial f_i}{\partial v}$. The connection between the three differentiation concepts comes in:

Lemma 32.3 *Suppose $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $x \in \operatorname{int}(\operatorname{Dom}(f))$. Then the partial derivatives at x exist and we have*

$$Df(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \dots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \dots & \frac{\partial f_m}{\partial x_n}(x) \end{pmatrix} \tag{32.6}$$

Moreover, for any $v \in \mathbb{R}^n$,

$$\frac{\partial f_i}{\partial v}(x) = e_i \cdot Df(x)v \tag{32.7}$$

Proof. For (32.7), note that for $z := x + hv$ with $h > 0$ we have

$$\left|\frac{f_i(x+hv) - f_i(x)}{h} - e_i \cdot Df(x_0)v\right| \leqslant \|v\| \frac{\|f(z) - f(x) - D(f_0)(z-x)\|}{\|z-x\|} \tag{32.8}$$

where we used $\|z-x\| = h\|v\|$ and also noted that for all $a = (a_1,\dots,a_m) \in \mathbb{R}^m$ and all $i = 1,\dots,m$ we have $|a_i| \leqslant \|a\|$. For (34.32) we specialize to the case $v := e_j$. □

We call the matrix on the right of (34.32) the *Jacobian matrix* associated with f at x. The row-vectors can be interpreted using the notion of the *gradient* which assigns to a function $f\colon \mathbb{R}^n \to \mathbb{R}$ the vector

$$\nabla f(x) := \Bigl(\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x)\Bigr) \tag{32.9}$$

Both notions play an important role in multivariable calculus and differential geometry.

As in the single-variable case, we have:

Lemma 32.4 *Suppose that $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $x \in \operatorname{int}(\operatorname{Dom}(f))$. Then f is continuous at x.*

Proof. Let $\delta > 0$ be such that the ratio on (32.2) is less than 1. Then $\|f(z) - f(x)\| \leqslant \|A(z-x)\| + \|z-x\|$ and the right hand side tends to zero as $z \to x$. □

We remark that the existence of partial derivatives or even directional derivatives is not strong enough to ensure continuity. For the former case, this is seen by considering the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x,y) := \begin{cases} \frac{xy}{x^2+y^2} & \text{if } (x,y) \neq 0 \\ 0 & \text{if } (x,y) = 0. \end{cases} \tag{32.10}$$

For any $t \in \mathbb{R}$, we have $\lim_{x\to 0} f(x,tx) = \frac{t}{1+t^2}$ and so f is NOT continuous at $(0,0)$. Yet both partial derivatives of f at $(0,0)$ exist and are equal to zero. We point out an example with all directional derivatives an no continuity after (32.14).

32.2 Sufficient conditions for differentiability.

In spite of the undue emphasis on partial derivatives in Calculus, the sheer existence of partial derivatives is not sufficient for a function to be differentiable. A trivial counterexample is provided by the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x,y) := \begin{cases} x+y & \text{if } x=0 \vee y=0 \\ x^2+y^2 & \text{else} \end{cases} \tag{32.11}$$

whose partial derivatives at $(x,y) := (0,0)$ equal 1 yet the directional derivative in all non-coordinate directions equals zero (which, if f were differentiable at zero, would force $Df(0,0) = 0$ and thus make all partial derivatives vanish). A little less singular-looking example is

$$f(x,y) := \begin{cases} \frac{x^3}{x^2+y^2} & \text{if } x^2+y^2>0 \\ 0 & \text{else} \end{cases} \tag{32.12}$$

where the second line ensures that the function is continuous at zero. Here the partial derivatives at $(0,0)$ equal 1 and 0, respectively, yet $\frac{\partial f_i}{\partial v}(0,0) = \frac{v_1^3}{v_1^2+v_2^2}$ which is not linear in v and thus cannot be of the form (32.7).

Since we are at the counterexamples, let us note that not even the existence and linearity of the directional derivative is sufficient for differentiability. This is seen from the example of $f\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x,y) := \begin{cases} 0 & \text{if } x>0 \wedge x^2<y<2x^2 \\ x+y & \text{else} \end{cases} \tag{32.13}$$

Here the directional derivative at $(x,y) = (0,0)$ equals v_1+v_2 in all directions $v = (v_1,v_2)$ because each half line emanating from $(0,0)$ has an initial segment in the region where $f(x,y) = x+y$. Yet for $z \in \{(x,y) \in \mathbb{R}^2 \colon x>0 \wedge x^2<y<2x^2\}$ we have

$$f(z) - f(0) - (1,1)\cdot(z-0) = -(1,1)\cdot z = -(z_1+z_2) \tag{32.14}$$

and the fact that $|z_1+z_2|/\|z\| \geqslant \frac{1}{\sqrt{2}}$ once $0 \leqslant z_2 \leqslant z_1$ then shows that (32.2) fails, disproving differentiability of f at $(0,0)$. Another way to violate differentiability would be to change f to take value 1 in the first line in (32.13). This will not affect directional derivatives but would make f discontinuous at $(x,y) = (0,0)$.

A natural question in light of the above counterexamples is what conditions on the partial derivatives guarantee differentiability. We address this in:

Lemma 32.5 (A sufficient condition for differentiability) *Let $f\colon \mathbb{R}^n \to \mathbb{R}^m$ and let $x \in$* $\operatorname{Dom}(f)$ *be such that, for some $\delta > 0$, we have $B(x,\delta) \subseteq \operatorname{Dom}(f)$ and the following holds:*

(1) *the partial derivatives exists at all points of $B(x,\delta)$, and*
(2) *they are continuous at point x.*

Then f is differentiable at x.

Proof. Given $\epsilon > 0$, the continuity of the partial derivatives implies existence of some $\delta' \in (0, \delta)$ such that

$$\forall y, z \in B(x, \delta') \,\forall i = 1, \dots, m \,\forall j = 1, \dots, n \colon \left| \frac{\partial f_i}{\partial x_j}(y) - \frac{\partial f_i}{\partial x_j}(x) \right| < \epsilon \tag{32.15}$$

Now pick $z \in B(x, \delta')$ and, for each $k = 1, \dots, n$, denote

$$z^{(k)} := x + \sum_{j=1}^{k} (z_j - x_j) e_j \tag{32.16}$$

Then $z^{(n)} = z$ and $z^{(0)} = x$ and so

$$f_i(z) - f_i(x) = \sum_{k=1}^{n} \big[f_i(z^{(k)}) - f_i(z^{(k-1)}) \big] \tag{32.17}$$

Noting that the linear segment connecting $z^{(k-1)}$ to $z^{(k)}$ lies entirely in $B(x, \delta)$, the fact that $z^{(k)} - z^{(k-1)} = (z_j - x_j) e_j$ along with the Mean-Value Theorem ensure the existence of $y^{(k)}$ on this segment such that

$$f_i(z^{(k)}) - f_i(z^{(k-1)}) = \frac{\partial f_i}{\partial x_k}(y^{(k)})(z_k - x_k) \tag{32.18}$$

Subtracting $\frac{\partial f_i}{\partial x_k}(x)(z_k - x_k)$ on both sides and taking absolute value we then get

$$\left| f_i(z) - f_i(x) - \sum_{k=1}^{n} \frac{\partial f_i}{\partial x_k}(x)(z_k - x_k) \right| \leqslant \sum_{k=1}^{n} \left| \frac{\partial f_i}{\partial x_k}(x) - \frac{\partial f_i}{\partial x_k}(y^{(k)}) \right| |z_k - x_k| \tag{32.19}$$

Bounding the first absolute value on the right using (32.15) and invoking the Cauchy-Schwarz inequality, the right-hand side is at most $\epsilon \sqrt{n} \|z - x\|$. Squaring both sides, suming over $i = 1, \dots, m$ and taking square roots then yields

$$\| f(z) - f(x) - A(z - x) \| \leqslant \epsilon \sqrt{mn} \|z - x\| \tag{32.20}$$

for all $z \in B(x, \delta')$, where A is the matrix on the right of (34.32). Since ϵ was arbitrary, this implies (32.2) with $Df(x) = A$. □

We note, however, that differentiability at x does not tell us anything about differentiability, or even continuity away from x. This is seen from the example

$$f(x, y) := (x^2 + y^2) 1_{\mathbb{Q}}(x) 1_{\mathbb{Q}}(y) \tag{32.21}$$

which is continuous and differentiable at $(x, y) = (0, 0)$, yet is neither differentiable nor even continuous at all $(x, y) \neq (0, 0)$.

32.3 Matrix norm.

In order to start working with derivatives, and also their own analytic properties, we need to be able to quantify closeness of two matrices. For this we introduce:

Definition 32.6 (Matrix norm) *Given $m, n \geqslant 1$, denote by $\mathcal{M}_{m,n}$ the set of $m \times n$-matrices with real entries. Then for all $A \in \mathcal{M}_{m,n}$,*

$$\|A\| := \sup_{x \in \mathbb{R}^n \smallsetminus \{0\}} \frac{\|Ax\|}{\|x\|} \tag{32.22}$$

is the matrix norm of A. *If we think of A as a linear operator $x \mapsto Ax$ taking $\mathbb{R}^n$ to $\mathbb{R}^m$, we call $\|A\|$ the* operator norm of A.

Note that $\mathcal{M}_{m,n}$ is a vector space, in which two matrices are added component by component and the multiplication by a scalar is also done component by component. To justify the use of the term norm, we now observe:

Lemma 32.7 *For each $A \in \mathcal{M}_{m,n}$, we have*

$$\max_{i=1,\dots,m} \max_{j=1,\dots,n} |A_{ij}| \leqslant \|A\| \leqslant \left(\sum_{i=1}^m \sum_{j=1}^n A_{i,j}^2 \right)^{1/2} \tag{32.23}$$

and so $0 \leqslant \|A\| < \infty$ with $\|A\| = 0$ implying $A = 0$. Moreover, for all $A, B \in \mathcal{M}_{m,n}$ all $\lambda \in \mathbb{R}$,

$$\|A + B\| \leqslant \|A\| + \|B\| \tag{32.24}$$

and

$$\|\lambda A\| = |\lambda| \|A\| \tag{32.25}$$

Finally, for all $A \in \mathcal{M}_{m,n}$ and all $B \in \mathcal{M}_{k,m}$ we also have

$$\|BA\| \leqslant \|B\| \|A\| \tag{32.26}$$

In short, $A \mapsto \|A\|$ is a norm on $\mathcal{M}_{m,n}$ under which the matrix product is continuous.

Proof. By the Cauchy-Schwarz inequality,

$$|(Ax)_i|^2 = \left(\sum_{j=1}^n A_{ij} x_j \right)^2 \leqslant \left(\sum_{j=1}^n A_{i,j}^2 \right) \|x\|^2 \tag{32.27}$$

Summing over $i = 1, \dots, m$ and taking the square root we get $\|A\| \leqslant (\sum_{i=1}^m \sum_{j=1}^n A_{i,j}^2)^{1/2}$. Similarly, $\|Ae_j\| \geqslant |(Ae_j)_i| = |A_{ij}|$ implies $\|A\| \geqslant |A_{ij}|$ for all i and j, proving (32.23).

Noting that $\|(A + B)x\| \leqslant \|Ax\| + \|Bx\|$ we then get (32.24). The relation (32.25) then follows from $\|\lambda Ax\| = |\lambda| \|Ax\|$. Hence, $A \mapsto \|A\|$ is a norm on $\mathcal{M}_{m,n}$. Concerning the product, we note that the definition of the matrix norm implies

$$\forall x \in \mathbb{R}^n \colon \ \|Ax\| \leqslant \|A\| \|x\| \tag{32.28}$$

with $\|A\|$ being the smallest constant for which this inequality holds. (The case $x = 0$ is checked directly.) Hence $\|BAx\| \leqslant \|B\| \|Ax\| \leqslant \|B\| \|A\| \|x\|$, proving (32.26) as well. □

Note that the bound (32.23) readily implies

$$\frac{1}{\sqrt{mn}} \left(\sum_{i=1}^m \sum_{j=1}^n A_{i,j}^2 \right)^{1/2} \leqslant \|A\| \leqslant \left(\sum_{i=1}^m \sum_{j=1}^n A_{i,j}^2 \right)^{1/2} \tag{32.29}$$

Since $\mathcal{M}_{m,n}$ is naturally identified with $\mathbb{R}^{mn}$, the operator norm is comparable to the Euclidean norm on $\mathbb{R}^{mn}$. (It can be checked that operator norm-squared is equal to the maximal eigenvalue of the positive semi-definite matrix $A^{\mathrm{T}}A$ while $\sum_{i=1}^m \sum_{j=1}^n A_{i,j}^2$ equals the trace of $A^{\mathrm{T}}A$ which is the sum of its eigenvalues.)

Using the matrix norm, we rephrase the notion of differentiability in terms similar to those in the proof of Lemma 29.5:

Lemma 32.8 *Suppose $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $x \in \operatorname{int}(\operatorname{Dom}(f))$. Then there exists $u_x\colon \mathbb{R}^n \to \mathcal{M}_{m,n}$ with $\operatorname{Dom}(u_x) = \operatorname{Dom}(f)$ such that*

$$\forall z \in \operatorname{Dom}(f)\colon\;\; f(z) = f(x) + \big[Df(x) + u_x(z)\big](z-x) \tag{32.30}$$

and

$$\lim_{z\to x} \|u_x(z)\| = 0 \tag{32.31}$$

Conversely, if there exists a matrix $Df(x) \in \mathcal{M}_{m,n}$ and a function $u_x\colon \mathbb{R}^n \to \mathcal{M}_{m,n}$ with $\operatorname{Dom}(u_x) = \operatorname{Dom}(f)$ such that (32.30–32.31) hold, then f is differentiable at x.

Proof. Recall from Lemma 32.3 that differentiability implies existence of partial derivatives and that $(Df(x))_{ij} = \frac{\partial f_i}{\partial x_j}(x)$. For $z \neq x$ define

$$\big(u_x(z)\big)_{ij} := \left[f_i(z) - f_i(x) - \sum_{k=1}^n \frac{\partial f_i}{\partial x_k}(x)(z_k - x_k)\right]\frac{z_j - x_j}{\|z-x\|^2} \tag{32.32}$$

and set $u_x(x) := 0$. Then

$$\sum_{j=1}^n \big(u_x(z)\big)_{ij}(z_j - x_j) = f_i(z) - f_i(x) - \sum_{k=1}^n \frac{\partial f_i}{\partial x_k}(x)(z_k - x_k) \tag{32.33}$$

which rewrites as $u_x(z)(z-x) = f(z) - f(x) - Df(x)(z-x)$ and proves (32.30).

For (32.31) we square (37.34) and sum over i and j to get

$$\sum_{i=1}^m \sum_{j=1}^m \left|\big(u_x(z)\big)_{ij}\right|^2 = \left(\sum_{i=1}^m \left| f_i(z) - f_i(x) - \sum_{k=1}^n \frac{\partial f_i}{\partial x_k}(x)(z_k - x_k)\right|^2\right) \sum_{j=1}^n \frac{|z_j - x_j|^2}{\|z-x\|^4} \tag{32.34}$$

The second sum equals $\|x-z\|^{-2}$ while the first sum equals the squared-norm of $f(z) - f(x) - Df(x)(z-x)$. Using the inequality on the right of (32.23) along with the bound in (32.28) we thus have

$$\|u_x(z)\| \leqslant \frac{\|f(z) - f(x) - Df(x)(z-x)\|}{\|z-x\|} \tag{32.35}$$

The definition (32.2) of $Df(x)$ then implies (32.31).

For the converse, if (32.30) is true, then (32.35) holds with the inequality reversed (proving, somewhat remarkably, that equality holds there). The convergence (32.31) then implies (32.2). □

The concept of the matrix norm allows us to introduce:

Definition 32.9 *Let $f\colon \mathbb{R}^n \to \mathbb{R}^m$ be differentiable at every point of a non-empty open set $U \subseteq \operatorname{int}(\operatorname{Dom}(f))$. We say that Df is continuous at $x \in U$ if*

$$\lim_{z\to x} \|Df(z) - Df(x)\| = 0 \tag{32.36}$$

If this holds at all $x \in U$, we say that Df is continuously differentiable *in U.*