2 Differentiation

The general notion of a continuous mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$ embraces continuous functions of several variables, linear operators $\mathbb{R}^n \to \mathbb{R}^m$, paths in $\mathbb{R}^m$, etc. Many nonlinear mappings are approximately linear near a point, be it dimension one or more; but in higher dimensions linear operators are not just coefficients.

2a A mapping near a point

We define a set of mappings $\mathbb{R}^n \to \mathbb{R}^m$, denoted $o_{n,m}(|\cdot|^p)$ or just $o(|\cdot|^p)$, for a given $p \ge 0$, as follows.[1] [2] [Sh:Def.4.2.1]

2a1 Definition. A mapping $f : \mathbb{R}^n \to \mathbb{R}^m$ belongs to $o(|\cdot|^p)$ if

$$\frac{|f(x)|}{|x|^p} \to 0 \text{ as } x \to 0 , \quad \text{and } f(0) = 0 .$$

Naturally, we write $o(1)$ instead of $o(|\cdot|^0)$, and $o(|\cdot|)$ instead of $o(|\cdot|^1)$.

One traditionally writes $f(x) = o(|x|^p)$ rather than $f \in o(|\cdot|^p)$.[3]

An equivalent formulation: for every $\varepsilon > 0$ the inequality

$$|f(x)| \le \varepsilon |x|^p$$

holds on some neighborhood of 0.

Here "neighborhood of 0" may be treated in many ways, for example,

[1] This is about $|x| \to 0$; a similar notation is used for $|x| \to \infty$, but we do not need it now.

[2] "Bachmann-Landau notation".

[3] Shurman [Sh:Sect.4.2] writes "$o(h)$" rather than "$o(|\cdot|)$" and "f is $o(h)$" rather than "$f \in o(|\cdot|)$", but does not hesitate writing "$o(h) \subset o(1)$".
Quote: How many use the symbolism $\mathcal{O}(1)$ without realizing that there is a tacit convention? It is true that $\sin x = \mathcal{O}(1)$; but it is not true that $\mathcal{O}(1) = \sin x$. (J.E. Littlewood, "A mathematician's miscellany", 1953, p.42. pdf).

* (a) an open ball $\{x : |x| < \delta\}$ ($\delta > 0$);
* (b) a closed ball $\{x : |x| \le \delta\}$ ($\delta > 0$);
* (c) an open set containing 0;
* (d) a set whose interior contains 0.

These are different classes of sets; however, every set of class (d) contains some set of class (a) (and the other way round); and the same holds for all other pairs of these classes.[1] This is why (a), (b), (c), (d) are interchangeable in the phrase "on some neighborhood of 0". This phrase may be abbreviated to "near 0".

From now on,
"neighborhood of x" stands for "a set whose interior contains x";
"open neighborhood of x" stands for "an open set containing x".

The intersection of two (or finitely many) neighborhoods of x is a neighborhood of x; for infinitely many neighborhoods this is not the case.

2a2 Exercise. If a linear operator $T : \mathbb{R}^n \to \mathbb{R}^m$ belongs to $o(|\cdot|)$ then $T = 0$. Prove it. [Sh:Prop.4.2.5]

2a3 Exercise. If a polynomial $P : \mathbb{R}^n \to \mathbb{R}$ of degree $\le k$ belongs to $o(|\cdot|^k)$ then $P = 0$. Prove it.[2]

Two mappings $f, g : \mathbb{R}^n \to \mathbb{R}^m$ equal near x are called *germ equivalent* at x. The same applies to mappings defined on (generally different) neighborhoods of x.[3] This is an equivalence relation. Its equivalence classes are called *germs.* The germ of f is denoted by $[f]_x$.

In contrast to a function, a germ has no values at points (except for x); for a given $y \ne x$,

$$[f_1]_x = [f_2]_x \quad \text{does not imply} \quad f_1(y) = f_2(y)\,.$$

Similarly to functions, germs may be multiplied by real numbers; for a given $c \in \mathbb{R}$,

$$[f_1]_x = [f_2]_x \quad \Longrightarrow \quad [cf_1]_x = [cf_2]_x\,,$$

and added:

$$\left.\begin{matrix} [f_1]_x = [f_2]_x \\ [g_1]_x = [g_2]_x \end{matrix}\right\} \quad \Longrightarrow \quad [f_1 + g_1]_x = [f_2 + g_2]_x$$

[1] In fact, the class (d) is a filter, while classes (a), (b), (c) are its bases.

[2] Hint: first, prove it for $n = 1$; then consider the polynomial $\mathbb{R} \ni t \mapsto P(tx)$.

[3] More generally, one may consider mappings f defined on $A \cap U$ where $A \subset \mathbb{R}^n$ is a given set (not dependent on f) and U is a neighborhood of x (dependent on f). Still more generally, A could also depend on f; but this situation is better served by mappings $R^n \to \mathbb{R}^m \cup \{\infty\}$, or $A \to \mathbb{R}^m \cup \{\infty\}$. Of course, x should be a limit point of A (otherwise the theory is trivial).

(even if the functions have different domains). Germs are a vector space.

2a4 Exercise. Prove that the vector space of germs is infinite-dimensional.[1],[2]

Many properties of mappings apply readily to germs, according to the pattern
$[f]_x$ is called $\boxed{}$ when f is $\boxed{}$ near x;
here $\boxed{}$ may be "linear", "bounded", "continuous", "one-to-one" etc.

If $[f_1]_x = [f_2]_x$ then $\lim_{y\to x} f_1(y) = \lim_{y\to x} f_2(y)$ in the following sense: either both limits exist and coincide, or neither limit exists. This way the notion of limit applies to germs; it is a local notion.

If $[f_1]_0 = [f_2]_0$ then $f_1 \in o(|\cdot|^p) \iff f_2 \in o(|\cdot|^p)$; thus, $o(|\cdot|^p)$ may be treated as a set of germs (rather than mappings); also a local notion.

When locality is evident, I do not hesitate writing "let $f : \mathbb{R}^n \to \mathbb{R}^m$" rather than "let $f : U \to \mathbb{R}^m$ where $U \subset \mathbb{R}^n$ is a neighborhood of x_0".

2a5 Exercise. (a) Prove that $f \in o(1)$ if and only if $f(0) = 0$ and f is continuous at 0.

(b) Prove or disprove: if $f \in o(1)$ then f is continuous near 0.

2a6 Exercise. (a) Find an example of a function $f : \mathbb{R}^2 \to \mathbb{R}$ such that $f \notin o(|\cdot|)$ and nevertheless the function $\mathbb{R} \ni y \mapsto f(x, y)$ belongs to $o(|\cdot|)$ for every $x \in \mathbb{R}$, and the function $\mathbb{R} \ni x \mapsto f(x, y)$ belongs to $o(|\cdot|)$ for every $y \in \mathbb{R}$.[3] [4]

(b) Can it happen that a continuous $f : \mathbb{R}^2 \to \mathbb{R}$ is not $o(|\cdot|)$ and nevertheless the function $\mathbb{R} \ni y \mapsto f(x, y)$ vanishes near 0 for every $x \in \mathbb{R}$, and the function $\mathbb{R} \ni x \mapsto f(x, y)$ vanishes near 0 for every $y \in \mathbb{R}$?

2a7 Exercise. (a) Find an example of a function $f : \mathbb{R}^n \to \mathbb{R}$ such that $f \notin o(|\cdot|)$ and nevertheless the function $\mathbb{R} \ni t \mapsto f(th)$ belongs to $o(|\cdot|)$ for every $h \in \mathbb{R}^n$.[5]

(b) Can it happen that a continuous $f : \mathbb{R}^n \to \mathbb{R}$ is not $o(|\cdot|)$ and nevertheless the function $\mathbb{R} \ni t \mapsto f(th)$ vanishes near 0 for every $h \in \mathbb{R}^n$?

Similarly to $o(\dots)$ we define $\mathcal{O}(\dots)$ as follows.

---

[1] Hint: think about polynomials.

[2] This space is not endowed with any useful topology.

[3] Hint: recall 1b2.

[4] See also 2e2.

[5] Hint: recall 1b3.

2a8 Definition. A mapping $f : \mathbb{R}^n \to \mathbb{R}^m$ belongs to $\mathcal{O}(|\cdot|^p)$ if for some $C < \infty$ the inequality

$$|f(x)| \le C|x|^p$$

holds near 0.

Also $\mathcal{O}(|\cdot|^p)$ may be treated as a set of germs.

Clearly,

$$f(x) = o(|x|^p) \quad \Longrightarrow \quad f(x) = \mathcal{O}(|x|^p)\,,$$

that is, $o(|\cdot|^p) \subset \mathcal{O}(|\cdot|^p)$. Also,

$$f(x) = \mathcal{O}(|x|^{p+\varepsilon}) \quad \Longrightarrow \quad f(x) = o(|x|^p)$$

whenever $\varepsilon > 0$.

2a9 Exercise. Both $o(|\cdot|^p)$ and $\mathcal{O}(|\cdot|^p)$ are vector spaces (subspaces of the vector space of germs). Prove it. [Sh:Prop.4.2.4]

2a10 Exercise. Prove that

$$\bigcap_{\varepsilon\in(0,p)} o(|\cdot|^{p-\varepsilon}) \supsetneqq \mathcal{O}(|\cdot|^p)\,, \qquad \bigcup_{\varepsilon>0} \mathcal{O}(|\cdot|^{p+\varepsilon}) \subsetneqq o(|\cdot|^p)\,.$$

2a11 Exercise. If $f(x) = o(|x|^p)$ and $g(x) = \mathcal{O}(|x|^q)$ then $f(x)g(x) = o(|x|^{p+q})$. Prove it. [Sh:Prop.4.2.6]

We turn to the composition $g \circ f : \mathbb{R}^n \to \mathbb{R}^l$ of two mappings $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}^l$.

2a12 Exercise. If $f \in o(|\cdot|^p)$ and $g \in \mathcal{O}(|\cdot|^q)$ then $g \circ f \in o(|\cdot|^{pq})$. Prove it. [Sh:Prop.4.2.7]

2a13 Exercise. It can happen that $f \in \mathcal{O}(1)$ and $g \in o(|\cdot|^q)$ but $g \circ f \notin o(1)$. Find a counterexample.[1] Can it happen that $g \circ f \notin \mathcal{O}(1)$?

2a14 Exercise. If $[f_1]_0 = [f_2]_0 \in o(1)$ and $[g_1]_0 = [g_2]_0$ then $[g_1 \circ f_1]_0 = [g_2 \circ f_2]_0$. Prove it. What about $\mathcal{O}(1)$ instead of $o(1)$?

Thus, composition of germs is well-defined under an appropriate condition.

2a15 Exercise. Formulate and prove the componentwise nature of $o(\dots)$ and $\mathcal{O}(\dots)$.[2] [Sh:Ex.4.2.2]

---

[1] Hint: try a constant f.

[2] Similarly to 1b4, not 1b5!

An arbitrary norm $\|\cdot\|$ on $\mathbb{R}^n$ being equivalent to the Euclidean norm $|\cdot|$ ($c|\cdot| \le \|\cdot\| \le C|\cdot|$, recall Sect. 1e), we may replace $|\cdot|$ with $\|\cdot\|$ in the definitions of $o(\dots)$ and $\mathcal{O}(\dots)$. Thus, $o(\dots)$ and $\mathcal{O}(\dots)$ are well-defined for mappings between arbitrary vector ^{fd}spaces.

2a16 Exercise. Let V_1, V_2 be vector ^{fd}spaces, $f : V_1 \to V_2$.

(a) If $f \in o(|\cdot|^p)$ then $f|_V \in o(|\cdot|^p)$ for every subspace $V \subset V_1$ (here $f|_V$ is the restriction); prove it. The same for $\mathcal{O}(\dots)$.

(b) Prove or disprove: if $f : \mathbb{R}^2 \to \mathbb{R}$ satisfies $f|_V \in o(1)$ for every one-dimensional vector subspace $V \subset \mathbb{R}^2$ then $f \in o(1)$.[1]

2a17 Exercise. Let S_1, S_2 be affine ^{fd}spaces. Prove that a mapping $f : S_1 \to S_2$ is continuous at a given point $x_0 \in S_1$ if and only if $f(x_0+\cdot) - f(x_0) \in o(1)$.

Less formally: $f(x_0 + h) - f(x_0) = o(1)$ as $h \to 0$.

Note that $f(x_0 + \cdot) - f(x_0) : \vec{S_1} \to \vec{S_2}$ (the difference spaces of S_1, S_2).

2b Derivative

2b1 Definition. A linear operator $T : \mathbb{R}^n \to \mathbb{R}^m$ is the *derivative* (or differential) at $x_0 \in \mathbb{R}^n$ of a mapping $f : \mathbb{R}^n \to \mathbb{R}^m$ if [Sh:Def.4.3.2]

$$f(x_0 + h) - f(x_0) - T(h) = o(|h|)\,.$$

Less formally, $f(x_0 + h) = f(x_0) + T(h) + o(|h|)$, that is, $f(x) = f(x_0) + T(x - x_0) + o(|x - x_0|)$.

More formally, $f(x_0 + \cdot) - f(x_0) - T(\cdot) \in o(|\cdot|)$.

Why "the derivative" rather than "a derivative"? Since such T (if exists) is unique. Indeed, the difference between two such operators must be 0 by 2a2 (and 2a9). [Sh:Prop.4.3.3]

If the derivative exists then f is called *differentiable* at x_0, and the derivative is denoted by $(Df)_{x_0}$, or Df_{x_0}, $Df(x_0)$, $\mathrm{d}f(x_0)$ etc. (And sometimes by $f'(x_0)$; but see 2b2 below.) Being a linear operator, it may be thought of as a matrix $m \times n$. If $(Df)_x$ exists for all $x \in \mathbb{R}^n$, we say that f is differentiable on $\mathbb{R}^n$. In this case Df is a (generally nonlinear) mapping from $\mathbb{R}^n$ to $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ (or $M_{m,n}(\mathbb{R})$). Similarly, f may be differentiable on a set $X \subset \mathbb{R}^n$, and then $Df : X \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$.

2b2 Exercise. A mapping $f : \mathbb{R} \to \mathbb{R}^m$ is differentiable at $x_0 \in \mathbb{R}$ if and only if the limit

$$\frac{\mathrm{d}}{\mathrm{d}x}\Big|_{x=x_0} f(x) = f'(x_0) = \lim_{x \to x_0} \frac{1}{x - x_0}\big(f(x) - f(x_0)\big)$$

[1] Hint: recall 1b3.

exists; in this case[1]

$$(Df)_{x_0} : h \mapsto hf'(x_0) ; \qquad f'(x_0) = (Df)_{x_0}(1) .$$

Prove it.

2b3 Exercise. Let $f : \mathbb{R} \to \mathbb{R}$ be differentiable at x_0, and $(Df)_{x_0} > 0$. Prove or disprove:

(a) $\exists \varepsilon > 0 \, \forall x \in (x_0, x_0 + \varepsilon) \;\; f(x) > f(x_0)$;

(b) f is increasing near x_0.

2b4 Exercise. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $f(x, y) = (x^2 - y^2, 2xy)$. Then f is differentiable on $\mathbb{R}^2$, and $(Df)_{(x_0,y_0)} : (h, k) \mapsto (2x_0 h - 2y_0 k, 2y_0 h + 2x_0 k)$, that is, $(Df)_{(x,y)} = \begin{pmatrix} 2x & -2y \\ 2y & 2x \end{pmatrix}$. Prove it (using only 2b1 and Sect. 2a). Do you guess why this Df is a *linear* map $\mathbb{R}^2 \to M_{2,2}(\mathbb{R})$? [Sh:Exer.4.3.4]

2b5 Exercise. (a) Prove that continuity at x_0 is necessary for differentiability at x_0. [Sh:Prop.4.3.4]

(b) Is it sufficient?

2b6 Exercise. (a) Prove that the relation $f(x_0 + \cdot) - f(x_0) \in \mathcal{O}(|\cdot|)$ is necessary for differentiability at x_0. [Sh:proof of Prop.4.3.4]

(b) Is it sufficient?

2b7 Exercise. Formulate and prove the componentwise nature of differentiability and derivative.[2] [Sh:Ex.4.3.3]

Here is a generalization of Def. 2b1.

2b8 Definition. Let S_1, S_2 be affine ^{fd}spaces. A linear operator $T : \vec{S}_1 \to \vec{S}_2$ is the *derivative* (or differential) at $x_0 \in S_1$ of a mapping $f : S_1 \to S_2$ if

$$f(x_0 + h) - f(x_0) - T(h) = o(|h|) .$$

Note that $f(x_0 + \cdot) - f(x_0) - T : \vec{S}_1 \to \vec{S}_2$.

We may upgrade S_1, S_2 to vector spaces, taking $x_0 = 0$ and $f(x_0) = 0$. Then the relation $f(x_0 + h) - f(x_0) - T(h) = o(|h|)$ becomes just

$$f - T \in o(|\cdot|) .$$

Locality of $o(|\cdot|)$ implies locality of the derivative (and differentiability) at a point.

[1] When $m = 1$ it is more convenient to write $f'(x_0)h$ rather than $hf'(x_0)$.

[2] Recall 2a15.

2b9 Proposition. (Linearity of derivative) [Sh:Prop.4.4.2]

Let S be an affine ^{fd}space, V a vector ^{fd}space, $f, g : S \to V$, $a, b \in \mathbb{R}$, and $x_0 \in S$. If f, g are differentiable at x_0 then also $af + bg$ is, and

$$\big(D(af + bg)\big)_{x_0} = a(Df)_{x_0} + b(Dg)_{x_0} .$$

Proof. We upgrade S to vector space taking $x_0 = 0$. We get $f - (Df)_0 \in o(|\cdot|)$, $g - (Dg)_0 \in o(|\cdot|)$. Thus, $(af + bg) - \big(a(Df)_0 + b(Dg)_0\big) = a\big(f - (Df)_0\big) + b\big(g - (Dg)_0\big) \in o(|\cdot|)$ by 2a9. $\square$

For $f, g : S_1 \to S_2$ we cannot take arbitrary linear combinations $af + bg$, but still can take affine combinations $af + bg$ with $a + b = 1$; and still, $\big(D(af + bg)\big)_{x_0} = a(Df)_{x_0} + b(Dg)_{x_0}$.

2b10 Proposition. (Product rule)

Let S be an affine ^{fd}space, $f, g : S \to \mathbb{R}$, and $x_0 \in S$. If f, g are differentiable at x_0 then also fg (the pointwise product) is, and

$$\big(D(fg)\big)_{x_0} = f(x_0)(Dg)_{x_0} + g(x_0)(Df)_{x_0} .$$

Proof. We upgrade S to vector space taking $x_0 = 0$. We get

$$\underbrace{f - f(0) - (Df)_0}_{\tilde f} \in o(|\cdot|) , \quad \underbrace{g - g(0) - (Dg)_0}_{\tilde g} \in o(|\cdot|) .$$

Thus,

$$\begin{aligned} fg - (fg)(0) &= \big(f(0) + (Df)_0 + \tilde f\big)\big(g(0) + (Dg)_0 + \tilde g\big) - f(0)g(0) = \\ &= f(0)(Dg)_0 + g(0)(Df)_0 + \underbrace{f(0)\tilde g + g(0)\tilde f + \big((Df)_0 + \tilde f\big)\big((Dg)_0 + \tilde g\big)}_{\in o(|\cdot|)} . \end{aligned}$$

$\square$

2b11 Exercise. Generalize the product rule[1]

(a) for the inner product $\langle f(\cdot), g(\cdot)\rangle$ where $f, g : S \to \mathbb{R}^m$;

(b) for the pointwise product fg where $f : S \to \mathbb{R}$ and $g : S \to \mathbb{R}^m$.

2b12 Proposition. (Chain rule) [Sh:Th.4.4.3]

Let S_1, S_2, S_3 be affine ^{fd}spaces, $f : S_1 \to S_2$, $g : S_2 \to S_3$, and $x_0 \in S_1$. If f is differentiable at x_0 and g is differentiable at $f(x_0)$ then $g \circ f$ is differentiable at x_0, and

$$\big(D(g \circ f)\big)_{x_0} = (Dg)_{f(x_0)} \circ (Df)_{x_0} .$$

[1] More generally: [Sh:Ex.4.4.8,4.4.9].

Proof. We upgrade the three affine spaces to vector spaces taking $x_0 = 0$, $f(x_0) = 0$ and $g(f(x_0)) = 0$. We get

$$f - \underbrace{(Df)_0}_{S} \in o(|\cdot|)\,, \quad g - \underbrace{(Dg)_0}_{T} \in o(|\cdot|)\,.$$

Thus,[1] [2]

$$g \circ f - T \circ S = T \circ (f - S) + (g - T) \circ f \in o(|\cdot|)\,.$$

□

2b13 Exercise. Assume that mappings $f : \mathbb{R}^n \to \mathbb{R}^m$, $g : \mathbb{R}^m \to \mathbb{R}^n$ satisfy
$f(0_n) = 0_m$, $g(0_m) = 0_n$;
$g(f(x)) = x$ for all x near 0_n;
$f(g(y)) = y$ for all y near 0_m.

(a) Does it follow that $m = n$?

(b) Let f be differentiable at 0_n and g differentiable at 0_m. Prove that $m = n$.

(c) Let f be continuous at 0_n and g continuous at 0_m. Does it follow that $m = n$?

2b14 Exercise. An affine mapping $f : \mathbb{R}^n \to \mathbb{R}^m$, $f(x) = Ax + b$, is differentiable everywhere, and $(Df)_x = A$ for all x; thus Df is a constant mapping $\mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$.

In particular, for a constant mapping $f(x) = b$ we get $(Df)_x = 0$ for all x.

Prove it. [Sh:Prop.4.4.1]

2b15 Exercise. Prove that an affine mapping $f : S_1 \to S_2$ is differentiable on S_1, and $(Df)_x = T$ for all x, where $T : \vec{S}_1 \to \vec{S}_2$ satisfies $T(x_1 - x_2) = f(x_1) - f(x_2)$ for all $x_1, x_2 \in S_1$.

Below, by "differentiate" I mean: (1) find the derivative at every point of differentiability, and (2) prove non-differentiability at every other point.

2b16 Exercise. Differentiate a mapping $\mathbb{R}^n \ni (x_1, \dots, x_n) \mapsto f_1(x_1) + \dots + f_n(x_n) \in \mathbb{R}^m$ for given differentiable $f_1, \dots, f_n : \mathbb{R} \to \mathbb{R}^m$.

---

[1] We move from $g \circ f$ to $T \circ S$ in two steps, changing one mapping at a time. Recall a similar argument from Analysis 1: if $x_n \to x$ and $y_n \to y$ then $x_n y_n \to xy$ since $x_n y_n - xy = x(y_n - y) + (x_n - x)y_n \to 0$. (Or, equally well, $x_n(y_n - y) + (x_n - x)y$.)

[2] Do you think that $g \circ f - T \circ S = g \circ (f - S) + (g - T) \circ S$ as well?

2b17 Exercise. (a) Differentiate a function $\mathbb{R}^n \ni x \mapsto |x| \in \mathbb{R}$.

(b) Differentiate a function $\mathbb{R}^n \ni x \mapsto |x|^2 \in \mathbb{R}$.

(c) Differentiate a function $\mathbb{R}^n \ni x \mapsto |x-a|^p \in \mathbb{R}$ for given $a \in \mathbb{R}^n$ and $p > 0$.

2b18 Exercise. (a) Differentiate a mapping $\mathbb{R}^n \setminus \{0\} \ni x \mapsto \frac{1}{|x|}x \in \mathbb{R}^n$.

(b) Differentiate a mapping $\mathbb{R}^2 \setminus \{0\} \to \mathbb{R}$ defined by $f(r\cos\theta, r\sin\theta) = g(\theta)$ (where $r > 0$) for a given 2π-periodic differentiable $g : \mathbb{R} \to \mathbb{R}$.[1] [2]

2b19 Exercise. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a non-constant homogeneous differentiable function; that is, $f(tx) = t^k f(x)$ for all $x \in \mathbb{R}^n$, $t \ge 0$. Prove that (a) $k \ge 1$; (b) $(Df)_x(x) = kf(x)$ (Euler's identity). [Sh:Ex.4.5.9]

2c Derivative along vector

Let V_1, V_2 be vector ^{fd}spaces, $V \subset V_1$ a vector subspace, and $f : V_1 \to V_2$. If f is differentiable at 0 then also its restriction $f|_V$ is, and

$$\big(D(f|_V)\big)_0 = (Df)_0|_V \,,$$

which follows readily from 2a16(a) (and 2b8). In particular it holds for one-dimensional subspaces

$$V_h = \{th : t \in \mathbb{R}\}\,, \qquad h \in V_1\,, \quad h \ne 0\,;$$

here $f|_{V_h}$ is basically a function of one variable t, $f(th) = \tilde f(t)$, and we have

$$\tilde f'(0) = \lim_{t\to 0} \frac{1}{t}\big(\tilde f(t) - \tilde f(0)\big) = \lim_{t\to 0}\frac{1}{t}\big(f(th) - f(0)\big) = (Df)_0(h)\,;$$

this is called the derivative (of f at 0) *along* h and denoted by $(D_h f)_0$ or $\nabla_h f(0)$. Thus,

$$(D_h f)_0 = \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} f(th) = (Df)_0(h)\,.$$

(The case $h = 0$ is harmless: just $(D_0 f)_0 = 0$.) We may also treat it as a special case of the chain rule: $\tilde f = f \circ \gamma$ where $\gamma(t) = th$; $(D\gamma)_0 = \gamma$ by 2b14, thus $(D\tilde f)_0 = (Df)_0 \circ (D\gamma)_0 = (Df)_0 \circ \gamma$ by 2b12, and $\tilde f'(0) = (D\tilde f)_0(1) = (Df)_0(h)$.

[1]Hint: first, do it for $g(\theta) = \cos\theta$ and $g(\theta) = \sin\theta$; then use arccos and arcsin. You really need both! You also need two cases: constant and non-constant g.

[2]It is tempting to say that $g(\theta)$ is a differentiable function on the circle (and so, (b) follows from (a) via the chain rule). However, the circle is not an open set! We'll return to the point in Analysis 4.

The same holds for affine spaces S_1, S_2:

$$\big(D(f|_S)\big)_{x_0} = (Df)_{x_0}|_{\vec{S}}$$

for $f : S_1 \to S_2$ differentiable at x_0, and affine subspace $S \ni x_0$. For a one-dimensional S we have $S = \{x_0 + th : t \in \mathbb{R}\}$, $h \in \vec{S_1}$, $(Df)_{x_0}(h) = (D_h f)_{x_0} = \frac{\mathrm{d}}{\mathrm{d}t}\big|_{t=0} f(x_0 + th)$.

Nonlinear paths may be used, too. Let $\gamma : \mathbb{R} \to S$ be differentiable at 0, $\gamma(0) = x_0$, $\gamma'(0) = h$, then

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0} f\big(\gamma(t)\big) = (D_h f)_{x_0}$$

by the chain rule: $\big(D(f \circ \gamma)\big)_0 = (Df)_{\gamma(0)} \circ (D\gamma)_0$; $(f \circ \gamma)'(0) = \big(D(f \circ \gamma)\big)_0(1) = (Df)_{\gamma(0)}\big((D\gamma)_0(1)\big) = (Df)_{x_0}(h)$.

Note that D_h is linear in h, that is,

$$(D_{c_1h_1+c_1h_2} f)_{x_0} = c_1(D_{h_1} f)_{x_0} + c_2(D_{h_2} f)_{x_0}$$

due to differentiability of f at x_0.

2c1 Exercise. It can happen that $\frac{\mathrm{d}}{\mathrm{d}t}\big|_{t=0} f(x_0 + th)$ exists for all h but is not linear in h. (Of course, such f cannot be differentiable at x_0.) Give an example.[1] [Sh:Ex.4.8.10]

2c2 Exercise. It can happen that $\frac{\mathrm{d}}{\mathrm{d}t}\big|_{t=0} f(x_0 + th)$ exists for all h and is linear in h and nevertheless f is not differentiable at x_0. Give an example.[2] [Sh:Ex.4.8.11]

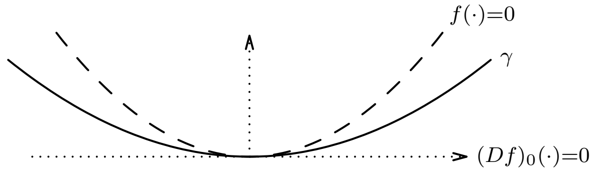
> "The multivariate derivative is truly a pan-dimensional construct, not just an amalgamation of cross sectional data." [Sh:p.156]

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at 0, and $(Df)_0 \ne 0$. Consider the hyperplane $\{h \in \mathbb{R}^n : (Df)_0(h) = 0\}$, the "positive" halfspace $\{h \in \mathbb{R}^n : (Df)_0(h) > 0\}$ and the "negative" halfspace $\{h \in \mathbb{R}^n : (Df)_0(h) < 0\}$. Let $\gamma : \mathbb{R} \to \mathbb{R}^n$ be differentiable at 0, $\gamma(0) = 0$, $\gamma'(0) = h \ne 0$. If h belongs to the "positive" halfspace then $\frac{\mathrm{d}}{\mathrm{d}t}\big|_{t=0} f\big(\gamma(t)\big) > 0$ and therefore $\exists \varepsilon > 0 \; \forall t \in (0, \varepsilon) \;\; f\big(\gamma(t)\big) > f(0)$. If h belongs to the hyperplane then $\frac{\mathrm{d}}{\mathrm{d}t}\big|_{t=0} f\big(\gamma(t)\big) = 0$. In this case it can happen that $\gamma(t)$ belongs to the "positive" halfspace for

[1] Hint: try $(x, y) \mapsto f(x, y)\sqrt{x^2 + y^2}$ for f as in 1b2.

[2] Hint: try $(x, y) \mapsto f(x, y)\sqrt{x^2 + y^2}$ for f as in 1b3.

all $t > 0$ and nevertheless $f\big(\gamma(t)\big) < f(0)$ for all $t > 0$. An example: $n = 2$; $f(x, y) = y - 2x^2$; $\gamma(t) = (t, t^2)$.



The same holds for $f : S \to \mathbb{R}$ where S is an affine *fd* space.

Thus, if $f : S \to \mathbb{R}$ has a local extremum at x_0 then $(Df)_{x_0} = 0$.

2c3 Exercise. (Rolle theorem in $\mathbb{R}^n$) Let $U \subset \mathbb{R}^n$ be a bounded open set, $f : \overline{U} \to \mathbb{R}$ a continuous function differentiable on U and vanishing on the boundary of U. Prove existence of $x \in U$ such that $(Df)_x = 0$. [Sh:Ex.4.8.4]

2d Mean value theorem

2d1 Proposition. Assume that $x_0, h \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $x_0 + th$ for all $t \in (0, 1)$, and continuous at x_0 and $x_0 + h$. Then there exists $t \in (0, 1)$ such that

$$f(x_0 + h) - f(x_0) = (D_h f)_{x_0+th}\,.$$

Proof. We introduce $\varphi : [0, 1] \to \mathbb{R}$ by $\varphi(t) = f(x_0 + th)$, note that $\varphi'(t) = (D_h f)(x_0+th)$ for $0 < t < 1$, and apply to φ the one-dimensional mean value theorem. $\square$

The set $\{x_0 + th : 0 < t < 1\}$ is the straight interval, sometimes denoted by $(x_0, x_0 + h)$. In terms of $a = x_0$ and $b = x_0 + h$ we get

$$f(b) - f(a) = (D_{b-a} f)_\xi = (Df)_\xi(b - a)$$

for some $\xi \in (a, b)$, assuming that f is continuous on $[a, b]$ and differentiable on (a, b); f must be defined at least at a, b and on some open set containing (a, b).

The same holds for an arbitrary vector (as well as affine) *fd* space in place of $\mathbb{R}^n$.

The mean value theorem fails for $f : \mathbb{R}^n \to \mathbb{R}^m$, $m > 1$ (even if $n = 1$; try $[0, 2\pi] \ni t \mapsto (\cos t, \sin t) \in \mathbb{R}^2$). Nevertheless. . .

2d2 Exercise. (a) Let U be a connected open subset of an affine *fd* space, and $f : U \to \mathbb{R}^m$ a differentiable mapping satisfying $Df = 0$ on U. Prove that f is constant on U.

(b) Does the same hold for a disconnected U?

(c) Generalize it to $Df = T$ on U (the same T for all points of U).

2d3 Exercise. (a) Let $U \subset \mathbb{R}^2$ be a *convex* open set, $f : U \to \mathbb{R}^m$, and $D_1 f = 0$ on U. Prove that $f(x, y)$ does not depend on y; that is, $f(x, y_1) = f(x, y_2)$ whenever $(x, y_1), (x, y_2) \in U$.

(b) Does the same hold if U is not convex, but still connected?

2e Partial derivative

For a mapping $f : \mathbb{R}^n \to \mathbb{R}^m$ the k-th *partial derivative* [Sh:Def.4.5.1]

$$
\begin{aligned}
(D_k f)_x = \partial_k f(x) = \frac{\partial}{\partial x_k} f(x_1, \dots, x_n) = \\
= \lim_{t \to 0} \frac{1}{t} \big(f(x_1, \dots, x_{k-1}, x_k + t, x_{k+1}, \dots, x_n) - f(x_1, \dots, x_n) \big)
\end{aligned}
$$

is just $(D_h f)_x$ where h is the k-th basis vector, $h = e_k = (0, \dots, 0, 1, 0, \dots, 0)$. That is, D_k is rather a shortcut for D_{e_k}, provided that f is differentiable at x_0; for now we assume that it is.

In terms of the $m \times n$ matrix A of the linear map $(Df)_{x_0}$ the vector $(D_k f)_{x_0} \in \mathbb{R}^m$ is the k-th column Ae_k of A.

In terms of the coordinate functions $f_1, \dots, f_m : \mathbb{R}^n \to \mathbb{R}$ satisfying $f(x) = (f_1(x), \dots, f_m(x))$ we have $A = \big((D_k f_l)_{x_0} \big)_{k=1,\dots,n; l=1,\dots,m}$. [Sh:Th.4.5.2] One often writes

$$
Df = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \dots & \frac{\partial y_1}{\partial x_n} \\ \dots & \dots & \dots \\ \frac{\partial y_m}{\partial x_1} & \dots & \frac{\partial y_m}{\partial x_n} \end{pmatrix} \quad \text{where} \quad \begin{cases} y_1 = f_1(x_1, \dots, x_n), \\ \dots \\ y_m = f_m(x_1, \dots, x_n). \end{cases}
$$

The chain rule leads to matrix multiplication: [Sh:Th.4.5.4]

$$
\frac{\partial z_k}{\partial x_i} = \sum_j \frac{\partial z_k}{\partial y_j} \frac{\partial y_j}{\partial x_i} \quad \text{where} \quad z = g(y) \, , \quad y = f(x) \, .
$$

Now, what happens if f is not assumed to be differentiable at x_0? By 2c1, existence of $(D_k f)_{x_0}$ for all $k = 1, \dots, n$ does not imply existence of $(Df)_{x_0}$.

2e1 Proposition. Assume that all partial derivatives of a mapping $f : \mathbb{R}^n \to \mathbb{R}^m$ exist *near* x_0 and are continuous *at* x_0. Then f is differentiable at x_0. [Sh:Th.4.5.3]

2e2 Lemma. Assume that $f : \mathbb{R}^n \to \mathbb{R}$, $D_n f$ exists near 0, is continuous at 0, and $(D_n f)_0 = 0$. Then the function $g : \mathbb{R}^n \to \mathbb{R}$ defined by

$$
g(x_1, \dots, x_n) = f(x_1, \dots, x_n) - f(x_1, \dots, x_{n-1}, 0)
$$

belongs to $o(|\cdot|)$.[1]

[1] See also 2a6(a).

Proof. The mean value theorem gives

$$g(x_1, \dots, x_n) = \underbrace{(D_n f)_{(x_1,\dots,x_{n-1},\xi(x_1,\dots,x_n))}}_{o(1)} \underbrace{x_n}_{O(|\cdot|)}$$

for some $\xi(x_1, \dots, x_n)$ between 0 and x_n. $\square$

Proof of Prop. 2e1. [1] By the componentwise nature of continuity (1b4), derivative (2b7) and partial derivative we take $m = 1$. As before, we take $x_0 = 0$ and $f(0) = 0$. Now we use induction in n. The case $n = 1$ is trivial. Let $n > 1$. It is sufficient to prove that the linear operator $T : \mathbb{R}^n \to \mathbb{R}$ defined by $T(x_1, \dots, x_n) = (D_1 f)_0 x_1 + \dots + (D_n f)_0 x_n$ satisfies $f - T \in o(|\cdot|)$. On one hand, the induction hypothesis gives

$$f(x_1, \dots, x_{n-1}, 0) - T(x_1, \dots, x_{n-1}, 0) = o(|\cdot|)\,.$$

On the other hand, Lemma 2e2 applied to $f - T$ gives

$$(f - T)(x_1, \dots, x_n) - (f - T)(x_1, \dots, x_{n-1}, 0) = o(|\cdot|)\,.$$

Thus, $f - T \in o(|\cdot|)$. $\square$

2e3 Definition. Let $U \subset \mathbb{R}^n$ be an open set. A differentiable mapping $f : U \to \mathbb{R}^m$ is *continuously differentiable* if the mapping Df is continuous (from U to $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$). The set of all continuously differentiable mappings $U \to \mathbb{R}^m$ is denoted by $C^1(U \to \mathbb{R}^m)$. In particular, $C^1(U) = C^1(U \to \mathbb{R})$.

Note that $C^1(U \to \mathbb{R}^m)$ is a vector space, and $C^1(U)$ is an algebra: $fg \in C^1(U)$ for all $f, g \in C^1(U)$.

2e4 Exercise. (a) Let $f \in C^1(U)$ and $g \in C^1(U \to \mathbb{R}^m)$; prove that $fg \in C^1(U \to \mathbb{R}^m)$.

(b) Let $f, g \in C^1(U \to \mathbb{R}^m)$; prove that $\langle f(\cdot), g(\cdot) \rangle \in C^1(U)$.[2]

2e5 Exercise. Prove that $f : U \to \mathbb{R}^m$ is continuously differentiable if and only if its partial derivatives $D_1 f, \dots, D_n f$ are continuous mappings $U \to \mathbb{R}^m$.

2e6 Exercise. (a) Differentiate[3] a mapping $\mathbb{R}^2 \ni (r, \theta) \mapsto (r \cos\theta, r \sin\theta) \in \mathbb{R}^2$.

[1] The proof in [Sh:p.156–157] is incorrect. The mean value theorem is not mentioned there, and the conclusion of Lemma 2e2 is just taken for granted.

[2] Hint: use 2b11.

[3] Do not forget the phrase before 2b16.

(b) Differentiate a function $f : (0,\infty) \times \mathbb{R} \to \mathbb{R}$ defined by $f(r,\theta) = g(r\cos\theta, r\sin\theta)$ for a given differentiable $g : \mathbb{R}^2 \to \mathbb{R}$.

(c) For f, g as in (b) prove that

$$\left(\frac{\partial g}{\partial x}\right)^2 + \left(\frac{\partial g}{\partial y}\right)^2 = \left(\frac{\partial f}{\partial r}\right)^2 + \frac{1}{r^2}\left(\frac{\partial f}{\partial \theta}\right)^2$$

whenever $x = r\cos\theta$, $y = r\sin\theta$, $r > 0$.

2e7 Exercise. On the vector space $M_{n,n}(\mathbb{R})$ of all $n \times n$ matrices consider the function $f : A \mapsto \det(A)$ (determinant). Prove that

(a) f is differentiable everywhere, and Df is continuous everywhere (as a mapping from $M_{n,n}(\mathbb{R})$ to $\mathcal{L}(M_{n,n}(\mathbb{R}), \mathbb{R})$).

(b) $(Df)_I(H) = \operatorname{tr}(H)$ for all $H \in M_{n,n}(\mathbb{R})$; here I is the unit matrix.

(c) $(D\log f)_A(H) = \operatorname{tr}(A^{-1}H)$ for all $H \in M_{n,n}(\mathbb{R})$ and all invertible $A \in M_{n,n}(\mathbb{R})$.[1]

Thus,

$$\log\det(A+H) \approx \log\det A + \operatorname{tr}(A^{-1}H)$$

for small H.

2e8 Exercise. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable and symmetric in the sense that $f(x_1,\dots,x_n)$ is insensitive to any permutation of $x_1,\dots,x_n$. Prove that

(a) $(D_i f)_{(x_1,\dots,x_n)} = (D_j f)_{(x_1,\dots,x_n)}$ whenever $x_i = x_j$;

(b) the operator $(Df)_{(x_1,\dots,x_n)}$ cannot be one-to-one if some of $x_1,\dots,x_n$ are equal.

2e9 Exercise. Consider the affine space $S_n = \{f : f^{(n)}(\cdot) = n!\}$ (a special case of 1c3 for a constant function $g(\cdot) = n!$) and a mapping $\varphi : \mathbb{R}^n \to S_n$,

$$\varphi(t_1,\dots,t_n) : t \mapsto (t-t_1)\dots(t-t_n)\,.$$

Prove that

(a) the operator $(D\varphi)_{(t_1,\dots,t_n)}$ cannot be invertible if some of $t_1,\dots,t_n$ are equal;

(b) the operator $(D\varphi)_{(t_1,\dots,t_n)}$ is invertible whenever $t_1,\dots,t_n$ are pairwise distinct;

(c) $\dim(D\varphi)_{(t_1,\dots,t_n)}(\mathbb{R}^n) = |\{t_1,\dots,t_n\}|$;

that is, the dimension of the image is equal to the number of distinct coordinates.

---

[1] [Sh:Ex.4.4.9].

2f Gradient, directional derivative

Let[1] $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at $x_0 \in \mathbb{R}^n$; then $(Df)_{x_0} = T : \mathbb{R}^n \to \mathbb{R}$, $T(h_1, \dots, h_n) = t_1 h_1 + \dots + t_n h_n = \langle (t_1, \dots, t_n), (h_1, \dots, h_n) \rangle$. Denoting $(t_1, \dots, t_n)$ by[2] $\nabla f(x_0)$ we get

$$(Df)_{x_0} : h \mapsto \langle \nabla f(x_0), h \rangle \,;$$

the vector $\nabla f(x_0)$ is called the *gradient* of f at x_0.

Similarly to the derivative, the gradient is a local notion, well-defined for $f : E \to \mathbb{R}$ whenever E is a *Euclidean* affine ^{fd}space (recall 1c7). In contrast to the derivative, the gradient is ill-defined if E is just an affine (or vector) ^{fd}space.

When $E = \mathbb{R}^n$, the gradient is related to partial derivatives by

$$\nabla f(x_0) = \big((D_1 f)_{x_0}, \dots, (D_n f)_{x_0} \big) \,.$$

In Euclidean E the gradient is related to derivative along vector by

$$(D_h f)_{x_0} = \langle \nabla f(x_0), h \rangle \,,$$

both sides being $(Df)_{x_0}(h)$.

When $|h| = 1$, $(D_h f)_{x_0}$ is also called the *directional derivative* (of f at x_0 in the direction h). Note that

$$|(D_h f)_{x_0}| \le |\nabla f(x_0)| \,;$$

the equality is reached when $h = \pm \frac{1}{|\nabla f(x_0)|} \nabla f(x_0)$ (assuming $\nabla f(x_0) \ne 0$). Thus, $\nabla f(x_0)$ points in the direction of greatest increase of f at x_0, and $|\nabla f(x_0)|$ is this greatest rate. [Sh:p.180] Also, $\nabla f(x_0)$ is orthogonal to the hyperplane $\{ h \in \mathbb{R}^n : (Df)_0(h) = 0 \}$.

By the mean value theorem, $f(b) - f(a) = \langle \nabla f(\xi), b - a \rangle$ for some $\xi \in (a, b)$; thus,

$$|f(b) - f(a)| \le |b - a| \sup_{0<t<1} \big| \nabla f\big(a + t(b-a)\big) \big| \,.$$

2f1 Exercise. Let $U \subset \mathbb{R}^n$ be an open set, $f \in C^1(U)$, and $|\nabla f| \le M$ on U. Then the relation

$$|f(b) - f(a)| \le M |b - a| \quad \text{for all } a, b \in U$$

must hold whenever U is convex (prove it), but can fail when U is connected and not convex (find an example).

---

[1]Note $m = 1 \dots$

[2]It really means $(\nabla f)(x_0)$ rather than $\nabla(f(x_0))$.

2f2 Exercise. Let $U \subset \mathbb{R}^n$ be an open set, $f \in C^1(U)$, and $K \subset U$ is compact. Prove that $f\big|_K$ is a Lipschitz function; that is, $\exists M \quad \forall x, y \in K \quad |f(x) - f(y)| \le M|x - y|$.[1]

Gradient is defined for functions $\mathbb{R}^n \to \mathbb{R}$, not mappings $\mathbb{R}^n \to \mathbb{R}^m$. However, gradients of coordinate functions are related to the derivative of a mapping.

2f3 Lemma. Let a mapping $f : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable at x_0, and $f_1, \dots, f_m : \mathbb{R}^n \to \mathbb{R}$ be the coordinate functions of f (that is, $f(x) = \big(f_1(x), \dots, f_m(x)\big)$. Then the following two conditions are equivalent:

(a) vectors $\nabla f_1(x_0), \dots, \nabla f_m(x_0)$ are linearly independent;

(b) the linear operator $(Df)_{x_0}$ maps $\mathbb{R}^n$ *onto* $\mathbb{R}^m$.

Proof. We have (recall 2b7)

$$(Df)_{x_0}(h) = \big((Df_1)_{x_0}(h), \dots, (Df_m)_{x_0}(h)\big) = \big(\langle \nabla f_1(x_0), h\rangle, \dots, \langle \nabla f_m(x_0), h\rangle\big) .$$

Violation of (a), that is, linear dependence between the gradient vectors means existence of $c = (c_1, \dots, c_m) \in \mathbb{R}^m$ such that the vector $c_1 \nabla f_1(x_0) + \dots + c_m \nabla f_m(x_0)$ vanishes. Equivalently: this vector has zero scalar product by every vector $h \in \mathbb{R}^n$. That is, $(Df)_{x_0}(h)$ is orthogonal to c for all h. Existence of such c is exactly a violation of (b). $\square$

2g Higher order derivative

Given an open set $U \subset \mathbb{R}^n$, we define a set $C^k(U \to \mathbb{R}^m)$ of mappings $U \to \mathbb{R}^m$ recursively. First, $f \in C^0(U \to \mathbb{R}^m)$ if and only if f is continuous on U; further, $f \in C^{k+1}(U \to \mathbb{R}^m)$ if and only if f is differentiable on U and $D_h f \in C^k(U \to \mathbb{R}^m)$ for all $h \in \mathbb{R}^n$.

In particular, $C^k(U) = C^k(U \to \mathbb{R})$.

The same applies to functions on a vector of affine ^{fd}space.

Clearly,

$$C^0(U \to \mathbb{R}^m) \supset C^1(U \to \mathbb{R}^m) \supset C^2(U \to \mathbb{R}^m) \supset \dots$$

Treating D_h as a linear operator $C^{k+1}(U) \to C^k(U)$ we note linearity in h:

$$D_h = h_1 D_1 + \dots + h_n D_n \quad \text{for } h = (h_1, \dots, h_n) .$$

[1]Hint: note that U need not be convex; assuming $f(x_n) - f(y_n) \ge n|x_n - y_n|$ take a convergent subsequence...

Thus, the condition "$D_h f \in C^k(U \to \mathbb{R}^m)$ for all $h \in \mathbb{R}^n$" is equivalent to "$D_1 f, \dots, D_n f \in C^k(U \to \mathbb{R}^m)$". That is, $C^k(U)$ consists of functions f such that $D_{i_1} \dots D_{i_k} f \in C^0(U)$ for all $i_1, \dots, i_k \in \{1, \dots, n\}$.

By the well-know theorem (Young, Schwarz, Clairaut),

$$D_i D_j = D_j D_i \quad \text{and therefore } D_{h_1} D_{h_2} = D_{h_2} D_{h_1} \,.$$

The *second differential* of $f \in C^2(U)$ at $x_0 \in U$ is the symmetric bilinear form

$$\mathbb{R}^n \times \mathbb{R}^n \ni (h_1, h_2) \mapsto D_{h_1} D_{h_2} f(x_0) \in \mathbb{R} \,;$$

its matrix $\big(D_i D_j f(x_0)\big)_{i,j}$, consisting of second partial derivatives, is called *Hessian matrix.* The corresponding quadratic form

$$\mathbb{R}^n \ni h \mapsto D_h D_h f(x_0) \in \mathbb{R}$$

occurs in the second order multivariate Taylor formula

$$f(x_0 + h) = f(x_0) + D_h f(x_0) + \frac{1}{2} D_h D_h f(x_0) + o(|h|^2) \,.$$

This is the same as the univariate Taylor formula for the function $\mathbb{R} \ni t \mapsto f(x_0 + th)$ at $t = 1$.[1] For a higher order the situation is similar:

$$f(x_0 + h) = f(x_0) + D_h f(x_0) + \frac{1}{2!} D_h D_h f(x_0) + \dots + \frac{1}{k!} D_h^k f(x_0) + o(|h|^k) \,;$$

the sum of homogeneous polynomials (higher differentials).[2]

[1] In order to get $o(|h|^2)$ one needs uniform (in h) continuity of the functions $D_h D_h f$, but this is not a problem: they all boil down to the finite set of functions $D_i D_j f$.

[2] Such a polynomial is unique by 2a3.

Index