2 Basics of 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.

You should know the notion of:	*Forgot? Then see:*
$f(x) = o(\|x\|^p)$ as $x \to 0$ [1] [2]	[Sh:Def.4.2.1 on p.132]
$f(x) = O(\|x\|^p)$ as $x \to 0$	[Sh:Def.4.2.1 on p.132]
Derivative of a mapping at a point	[Sh:Def.4.3.2 on p.141]
Linearity of derivative	[Sh:Prop.4.4.2 on p.146]
Chain rule	[Sh:Th.4.4.3 on p.147]
Derivative along vector	[Shifrin:p.83]
Partial derivative	[Sh:Def.4.5.1 on p.153]
Continuously differentiable; C^1	[Shifrin:p.93]
Gradient; directional derivative	[Sh:Def.4.8.1, Th.4.8.2 on p.179–180]
Higher order derivatives	[Sh:Sect.4.6]

2a A mapping near a point

2a1 Exercise. If a linear operator $T : \mathbb{R}^n \to \mathbb{R}^m$ satisfies $Tx = o(|x|)$ then $T = 0$. Prove it. [Sh:Prop.4.2.5]

2a2 Exercise. If a polynomial $P : \mathbb{R}^n \to \mathbb{R}$ of degree $\le k$ satisfies $P(\cdot) = o(|\cdot|^k)$ then $P = 0$. Prove it.[3]

[1] This is about $|x| \to 0$; a similar notation is used for $|x| \to \infty$, but we restrict ourselves to $x \to 0$ and $p \in [0, \infty)$.

[2] "Bachmann-Landau notation".

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

2a3 Exercise. (a) Endow the set of all germs at 0 of functions $\mathbb{R}^n \to \mathbb{R}$ (or mappings $\mathbb{R}^n \to \mathbb{R}^m$) with the structure of vector space;

(b) prove that $o(|\cdot|^p)$ may be thought of as a vector subspace of germs, and $o(|\cdot|^{p+\varepsilon}) \subsetneqq o(|\cdot|^p)$;

(c) prove that the vector space of germs is infinite-dimensional.[1],[2]

2a4 Exercise. (a) Prove that $f(\cdot) = o(1)$ if and only if $f(0) = 0$ and f is continuous at 0.

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

2a5 Exercise. (a) Find an example of a function $f : \mathbb{R}^2 \to \mathbb{R}$ such that $f(x, \cdot) = o(|\cdot|)$ for every $x \in \mathbb{R}$, and $f(\cdot, y) = o(|\cdot|)$ for every $y \in \mathbb{R}$, but $f(x, y)$ is not $o(\sqrt{x^2 + y^2})$.[3]

(b) Can it happen that a continuous $f : \mathbb{R}^2 \to \mathbb{R}$ violates $f(x, y) = o(\sqrt{x^2 + y^2})$ and nevertheless the function $f(x, \cdot)$ vanishes near 0 for every $x \in \mathbb{R}$, and the function $f(\cdot, y)$ vanishes near 0 for every $y \in \mathbb{R}$?

2a6 Exercise. (a) Find an example of a function $f : \mathbb{R}^n \to \mathbb{R}$ that violates $f(\cdot) = o(|\cdot|)$ and nevertheless the function $\mathbb{R} \ni t \mapsto f(th)$ is $o(|t|)$ for every $h \in \mathbb{R}^n$.[4]

(b) Can it happen that a continuous $f : \mathbb{R}^n \to \mathbb{R}$ violates $f(\cdot) = o(|\cdot|)$ and nevertheless the function $\mathbb{R} \ni t \mapsto f(th)$ vanishes near 0 for every $h \in \mathbb{R}^n$?

2a7 Exercise. If $f(\cdot) = o(|\cdot|^p)$ and $g(\cdot) = \mathcal{O}(|\cdot|^q)$ then $f(\cdot)g(\cdot) = o(|\cdot|^{p+q})$. 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 $g : \mathbb{R}^m \to \mathbb{R}^l$.

2a8 Exercise. If $f(\cdot) = o(|\cdot|^p)$ and $g(\cdot) = \mathcal{O}(|\cdot|^q)$ then $g(f(\cdot)) = o(|\cdot|^{pq})$. Prove it. [Sh:Prop.4.2.7]

2a9 Exercise. It can happen that $f(\cdot) = \mathcal{O}(1)$ and $g(\cdot) = o(|\cdot|^q)$ but $g(f(\cdot))$ is not $o(1)$. Find a counterexample.[5] Can it happen that $g(f(\cdot))$ is not $\mathcal{O}(1)$?

2a10 Exercise. If $f_1(\cdot) = f_2(\cdot)$ near 0, both are $o(1)$, and $g_1(\cdot) = g_2(\cdot)$ near 0, then $g_1(f_1(\cdot)) = g_2(f_2(\cdot))$ near 0. Prove it. What about $\mathcal{O}(1)$ instead of $o(1)$?

---

[1] Hint: either use (b), or use polynomials.

[2] This space is not endowed with any useful topology. By the way, the set of *all* germs (at all points) is endowed with a useful non-metrizable topology (but is not a vector space); we do not need it.

[3] Hint: recall 1a2.

[4] Hint: recall 1a3.

[5] Hint: try a constant f.

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

2a11 Exercise. Formulate and prove the componentwise nature of $o(\dots)$ and $\mathcal{O}(\dots)$.[1] [Sh:Ex.4.2.2]

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 finite-dimensional vector spaces.

2a12 Exercise. Let V_1, V_2 be finite-dimensional vector spaces, $f : V_1 \to V_2$.

(a) If $f(\cdot) = o(|\cdot|^p)$ then $f|_V(\cdot) = o(|\cdot|^p)$ 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 $f|_V(\cdot) = o(1)$ for every one-dimensional vector subspace $V \subset \mathbb{R}^2$ then $f(\cdot) = o(1)$.[2]

2a13 Exercise. Let S_1, S_2 be finite-dimensional affine 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 $f(x_0 + \cdot) - f(x_0) = o(1)$.

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

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

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

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 2a1. [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

[1] Similarly to 1a4, not 1a5!

[2] Hint: recall 1a3.

$\mathbb{R}^n$. 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 $X \subset \mathbb{R}^n$,[1] 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

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

exists; in this case[2]

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

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. Formulate and prove the componentwise nature of differentiability and derivative.[3] [Sh:Ex.4.3.3]

Here is a generalization of Def. 2b1.

2b5 Definition. Let S_1, S_2 be finite-dimensional affine 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

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

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

---

[1] It means, f (not just $f|_X$) is differentiable at every point of X.

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

[3] Recall 2a11.

2b6 Proposition. (Linearity of derivative)

Let S be a finite-dimensional affine space, V a finite-dimensional vector 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

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

2b7 Exercise. Prove this proposition.[1]

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

2b8 Proposition. (Product rule)

Let S be a finite-dimensional affine 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

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

2b9 Exercise. Prove this proposition.[2]

2b10 Exercise. Generalize the product rule[3]

(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 $g : S \to \mathbb{R}^m$.

2b11 Proposition. (Chain rule)

Let S_1, S_2, S_3 be finite-dimensional affine 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

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

2b12 Exercise. Prove this proposition.[4]

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 $m = n$?

---

[1] Hint. You may generalize the proof for $\mathbb{R}^n$ known from Analysis 2. Alternatively, you can upgrade S and V to Cartesian spaces and apply the result from Analysis 2.

[2] See the hint to 2b7.

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

[4] See the hint to 2b7.

(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 have $(Df)_x = 0$ for all x. And for a linear mapping $f = A$ we have $Df = f$.[1]

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 $(Df)_x = \vec{f}$ for all x, where $\vec{f} : \vec{S_1} \to \vec{S_2}$ satisfies $\vec{f}(x_1 - x_2) = f(x_1) - f(x_2)$ 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 $f_1, \dots, f_n : \mathbb{R} \to \mathbb{R}^m$.

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 $g : \mathbb{R} \to \mathbb{R}$.[2] [3]

2b19 Exercise. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a non-constant differentiable homogeneous 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]

2b20 Exercise. Consider a function $f : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$ defined by $f(x) = g(|x|)$ for a given differentiable $g : (0, \infty) \to \mathbb{R}$. Prove that f is differentiable on $\mathbb{R}^n \setminus \{0\}$, and for all $x \in \mathbb{R}^n \setminus \{0\}$, $h \in \mathbb{R}^n$ holds

$$(Df)_x(h) = g'(|x|) \Big\langle h, \frac{x}{|x|} \Big\rangle .$$

[1] Does it disturb you?

[2] Hint: first, do it for $g(\theta) = \cos\theta$ and $g(\theta) = \sin\theta$; then use arccos and arcsin. You really need both!

[3] 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 this point in Analysis 4.

2b21 Exercise. Consider a mapping $f : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}^n$ defined by $f(x) = g(|x|)\frac{x}{|x|}$ for a given differentiable $g : (0, \infty) \to \mathbb{R}$. Prove that

(a) for all $x \in \mathbb{R}^n \setminus \{0\}$ and $h \in \mathbb{R}^n$,

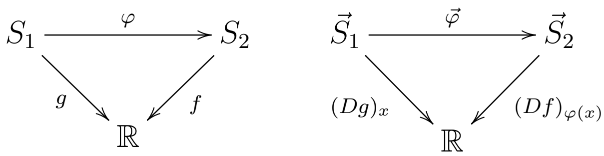
$$(Df)_x(h) = g'(|x|)h \qquad \text{for } h \shortparallel x\,,$$
$$(Df)_x(h) = \frac{g(|x|)}{|x|}h \qquad \text{for } h \perp x\,.$$

(Here $h \shortparallel x$ means $\exists \lambda\ h = \lambda x$, and $h \perp x$ means $\langle h, x \rangle = 0$.)

(b) $\|(Df)_x\| = \max\bigl(|g'(|x|)|, \frac{|g(|x|)|}{|x|}\bigr)$;

(c) $\max\limits_{|x|=r,|h|\le\varepsilon} |f(x+h) - f(x)| \le \varepsilon \max\Bigl(|g'(r)|, \dfrac{|g(r)|}{r}\Bigr) + o(\varepsilon)$ as $\varepsilon \to 0+$.[1]

2b22 Exercise. (a) Let S_1, S_2 be affine spaces, $\varphi : S_1 \to S_2$ an isomorphism, $f : S_2 \to \mathbb{R}$ a differentiable function, and $g(\cdot) = f(\varphi(\cdot)) : S_1 \to \mathbb{R}$. Then[2] $(Dg)_x = (Df)_{\varphi(x)} \circ \vec{\varphi}$ for all $x \in S_1$.



Prove it twice: first, via the isomorphism argument[3] and second, via the chain rule 2b11 (and 2b15).

(b) Consider the special case $S_1 = S_2$; reformulate (a) for this case.

From now on, $S_1 = S_2 = \mathbb{R}^2$.

(c) Consider the (more) special case $f = g$; reformulate (b) for this case.

From now on, $f = g$.

(d) Let φ be the rotation (of $\mathbb{R}^2$) by $\pi/2$; reformulate (c) for this case; make the formulation as simple as possible.

(e) A bit more general than (d): $f(r\cos\theta, r\sin\theta) = h(r, k\theta)$ for a given $k \in \{2, 3, \dots\}$ and h such that $h(r, \theta + 2\pi) = h(r, \theta)$.

(f) What about $f(ax) = f(x)$ for a given $a \in (1, \infty)$?

(g) What about $f(ax) = bf(x)$ for given $a \in (1, \infty)$ and $b \in (-\infty, 0) \cup (0, \infty)$?

(h) What about $f(r\cos\theta, r\sin\theta) = r^a h(\theta - b\ln r)$ for given $a, b \in (0, \infty)$?

(i) And what about $S_1 = S_2 = \mathbb{R}^3$ and $f(r\cos\theta, r\sin\theta, z) = \mathrm{e}^{az} h(\theta - z)$?

---

[1] Hint: $\max_{|h|\le\varepsilon} |f(x+h) - f(x)|$ depends on x via $|x|$ only.

[2] Looking at the diagram we see that $(Df)_{\varphi(x)} \circ \vec{\varphi}$ is the only plausible formula for $(Dg)_x$; do not bother to memorize it. A privilege of generality!

[3] Hint. Given an affine space S, consider triples (x, f, α) satisfying $(Df)_x = \alpha$.

2b23 Exercise. (a) Let S_1, S_2, R_1, R_2 be affine spaces, $\varphi : S_1 \to S_2$ and $\psi : R_1 \to R_2$ isomorphisms, $f : S_2 \to R_2$ a differentiable mapping, and $g = \psi^{-1} \circ f \circ \varphi : S_1 \to R_1$. Then[1] $(Dg)_x = (\vec{\psi})^{-1} \circ (Df)_{\varphi(x)} \circ \vec{\varphi}$ for all $x \in S_1$.

$$\begin{array}{ccc} S_1 & \xrightarrow{\varphi} & S_2 \\ {\scriptstyle g}\downarrow & & \downarrow{\scriptstyle f} \\ R_1 & \xrightarrow{\psi} & R_2 \end{array} \qquad\qquad \begin{array}{ccc} \vec{S}_1 & \xrightarrow{\vec{\varphi}} & \vec{S}_2 \\ {\scriptstyle (Dg)_x}\downarrow & & \downarrow{\scriptstyle (Df)_{\varphi(x)}} \\ \vec{R}_1 & \xrightarrow{\vec{\psi}} & \vec{R}_2 \end{array}$$

Prove it.[2]

(b) Consider the special case $R_1 = R_2 = S_1 = S_2$, $\varphi = \psi$; reformulate (a) for this case.

From now on, $S_1 = S_2 = R_1 = R_2 = \mathbb{R}^2$ and $\varphi = \psi$.

(c) Consider the (more) special case $f = g$; reformulate (b) for this case. From now on, $f = g$.

(d) Let φ be the rotation (of $\mathbb{R}^2$) by $\pi/2$; reformulate (c) for this case; give a simple description of the 2×2 matrix of $(Dg)_x$ in terms of the 2×2 matrix of $(Df)_{\varphi(x)}$.

(e) What about $f(r\cos\theta, r\sin\theta) = (h(r)\cos(\theta+\alpha(r)), h(r)\sin(\theta+\alpha(r)))$?

(f) What about $f(ax) = bf(x)$ for given $a \in (1,\infty)$ and $b \in (-\infty, 0) \cup (0,\infty)$?

2c Derivative along vector

Let V_1, V_2 be finite-dimensional vector 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 2a12(a) (and 2b5). 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,

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

[1]Again, the only plausible formula...

[2]Hint: either via the isomorphisms argument, or via the chain rule.

(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 2b11 (and 2b2): $\tilde f = f \circ \gamma$ where $\gamma(t) = th$; $\gamma'(0) = h$; $\tilde f'(0) = (D\tilde f)_0(1) = (Df)_0 \circ (D\gamma)_0(1) = (Df)_0(h)$.

The same holds for affine spaces S_1, S_2:

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

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

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

by the chain rule again.

Note that D_h is linear in h, that is,

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

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]

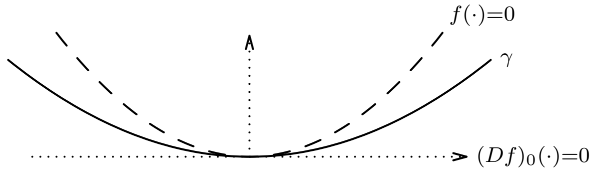
> "The multivariate derivative is truly a pan-dimensional construct,
> not just an amalgamation of cross sectional data." [Sh:p.156]

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 $\frac{\mathrm{d}}{\mathrm{d}t}\big|_{t=0} f\bigl(\gamma(t)\bigr) > 0$ and therefore $\exists \varepsilon > 0 \; \forall t \in (0, \varepsilon) \;\; f\bigl(\gamma(t)\bigr) > f(0)$ by 2b3(a). If h belongs to the hyperplane then $\frac{\mathrm{d}}{\mathrm{d}t}\big|_{t=0} f\bigl(\gamma(t)\bigr) = 0$. In this case it can happen that $\gamma(t)$ belongs to the

[1] Hint: try $(x, y) \mapsto f(x, y)\sqrt{x^2 + y^2}$ for f as in 1a2.

[2] Hint: try $(x, y) \mapsto f(x, y)\sqrt{x^2 + y^2}$ for f as in 1a3.

"positive" halfspace for 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 a finite-dimensional affine space. Therefore:

(2c3) If $f : S \to \mathbb{R}$ has a local extremum at x_0 then $(Df)_{x_0} = 0$.

2c4 Exercise. *(Multidimensional Rolle theorem)* Let S be a finite-dimensional affine space, $U \subset S$ a bounded open set, $f : \overline{U} \to \mathbb{R}$ a continuous function differentiable on U and constant on the boundary of U.

(a) Prove existence of $x \in U$ such that $(Df)_x = 0$. [Sh:Ex.4.8.4]

(b) Does it still holds for unbounded U?

2d Average velocity and maximal speed

A *path* in $\mathbb{R}^n$ is, by definition, a continuous mapping $\gamma : [t_0, t_1] \to \mathbb{R}^n$ given for some $t_0 < t_1$. Assuming that γ is differentiable on (t_0, t_1) we may treat $\gamma'(t)$ (recall 2b2) as the (instant) velocity at $t \in (t_0, t_1)$, $|\gamma'(t)|$ as the (instant) speed at t, and $\frac{1}{t_1-t_0}\big(\gamma(t_1) - \gamma(t_0)\big)$ as the average velocity. It is easy to guess that the norm of the average velocity cannot exceed the maximal speed,

$$\frac{|\gamma(t_1) - \gamma(t_0)|}{t_1 - t_0} \le \sup_{t \in (t_0, t_1)} |\gamma'(t)| \,. \tag{2d1}$$

What about a proof?

If γ is *continuously* differentiable on $[t_0, t_1]$, that is, γ' is continuous on (t_0, t_1) and extends by continuity to $[t_0, t_1]$, then

$$\gamma(t_1) - \gamma(t_0) = \int_{t_0}^{t_1} \gamma'(t)\,\mathrm{d}t \,, \quad \text{therefore}$$

$$|\gamma(t_1) - \gamma(t_0)| \le \int_{t_0}^{t_1} |\gamma'(t)|\,\mathrm{d}t \,,$$

that is, the norm of the average velocity cannot exceed the average speed. Practically, this argument solves the problem. But theoretically,

(a) it would be nicer to avoid integration, and

(b) the derivative need not be continuous.
For example, given $\alpha > 1$ and $\beta > 0$, a function

$$f(t) = \begin{cases} t^\alpha \sin t^{-\beta} & \text{for } t \ne 0, \\ 0 & \text{for } t = 0 \end{cases} \tag{2d2}$$

is differentiable on $\mathbb{R}$,

$$f'(t) = \begin{cases} \alpha t^{\alpha-1} \sin t^{-\beta} - \beta t^{\alpha-\beta-1} \cos t^{-\beta} & \text{for } t \ne 0, \\ 0 & \text{for } t = 0\,; \end{cases}$$

if $\alpha - \beta - 1 > 0$ then f' is continuous; if $\alpha - \beta - 1 < 0$ then f' is unbounded near the origin; and if $\alpha - \beta - 1 = 0$ then f' is bounded, but discontinuous at the origin.[1]

We'll consider several approaches to (2d1); the result is of little practical value (since such functions as (2d2) are rather exotic), but the ways are instructive, they demonstrate useful techniques able to meet the challenge.

The case $n = 1$ is simple; $\gamma : [t_0, t_1] \to \mathbb{R}$; by the one-dimensional mean value theorem,

$$\exists t \in (t_0, t_1) \quad \frac{1}{t_1 - t_0} \big(\gamma(t_1) - \gamma(t_0) \big) = \gamma'(t)\,; \tag{2d3}$$

(2d1) follows.

For $n > 1$ (2d3) fails; a counterexample: $\gamma(t) = (\cos t, \sin t) \in \mathbb{R}^2$ for $t \in [0, 2\pi]$ (or $t \in [0, \pi]$).[2]

Applying the one-dimensional result to each coordinate γ_k of $\gamma : t \mapsto \gamma(t) = \big(\gamma_1(t), \dots, \gamma_n(t)\big) \in \mathbb{R}^n$ we get

$$\begin{aligned} \frac{|\gamma(t_1) - \gamma(t_0)|}{t_1 - t_0} &= \frac{1}{t_1 - t_0} \sqrt{\sum_k (\gamma_k(t_1) - \gamma_k(t_0))^2} \le \\ \le \sqrt{n} \max_k \frac{|\gamma_k(t_1) - \gamma_k(t_0)|}{t_1 - t_0} &\le \sqrt{n} \max_k \sup_t |\gamma_k'(t)| = \sqrt{n} \sup_t \max_k |\gamma_k'(t)| \le \\ &\le \sqrt{n} \sup_t |\gamma'(t)|\,. \end{aligned}$$

This is weaker than (2d1) but still useful.[3]

---

[1]By the way, a differentiable function on $(0, 1)$ can be nowhere monotone. Did you know?! See for example Sect. 9c of my advanced course "Measure and category".

[2]In fact, $\frac{1}{t_1 - t_0}(\gamma(t_1) - \gamma(t_0))$ is some convex combination of $\gamma'(s_1), \dots, \gamma'(s_{n+1})$ for some $s_1, \dots, s_{n+1} \in (t_0, t_1)$ (and if γ' is continuous, $n + 1$ may be replaced with n).

[3]See Fleming, page 133, proof of Prop. 4.5; note the requirement $|d\tilde{g}^i(u)| \le \varepsilon/n$ rather than $|d\tilde{g}(u)| \le \varepsilon$.

It is better to apply the one-dimensional result to a linear combination of the coordinates of γ:

$$\frac{1}{t_1 - t_0} |\langle \gamma(t_1) - \gamma(t_0), a \rangle| \le \sup_t |\langle \gamma'(t), a \rangle| \le \sup_t |\gamma'(t)| \tag{2d4}$$

for arbitrary $a \in \mathbb{R}^n$, $|a| = 1$. Taking the optimal a, satisfying $\gamma(t_1) - \gamma(t_0) = |\gamma(t_1) - \gamma(t_0)| a$, we get (2d1).

Now (2d1) is proved; however, first, it would be nice to avoid the one-dimensional case; and second, why only the Euclidean norm? A more general claim

$$\frac{\|\gamma(t_1) - \gamma(t_0)\|}{t_1 - t_0} \le \sup_{t \in (t_0, t_1)} \|\gamma'(t)\| \tag{2d5}$$

should hold in every n-dimensional normed space.

It is possible to generalize the argument of (2d4).[1] But what about a more straightforward proof?

2d6 Exercise. Given $C \in (0, \infty)$, consider a set

$$A = \{ t \in [t_0, t_1] : \forall s \in [t_0, t] \;\; \|\gamma(s) - \gamma(t_0)\| \le C(s - t_0) \} .$$

(a) Prove that A is closed.

(b) Assuming that $C > \sup_t \|\gamma'(t)\|$ prove that A is relatively open in $[t_0, t_1]$.[2]

Thus, A is a relatively clopen subset, containing t_0, of the connected space $[t_0, t_1]$. Therefore A contains t_1. We see that $\|\gamma(t_1) - \gamma(t_0)\| \le C(t_1 - t_0)$ for all $C > \sup_t \|\gamma'(t)\|$; (2d5) follows.

Here is an alternative way.[3] Assume that $C > \sup_t \|\gamma'(t)\|$ and nevertheless $\frac{\|\gamma(t_1)-\gamma(t_0)\|}{t_1-t_0} \ge C$. We take $t = (t_0 + t_1)/2$ and note that $\frac{\|\gamma(t_1)-\gamma(t)\|}{t_1-t} \ge C$ or $\frac{\|\gamma(t)-\gamma(t_0)\|}{t-t_0} \ge C$. Repeating this argument we get a sequence of nested intervals $[t_0, t_1] \supset [r_1, s_1] \supset [r_2, s_2] \supset \dots$ whose intersection is a single point $t \in [t_0, t_1]$, and such that $\forall k \;\; \frac{\|\gamma(s_k)-\gamma(r_k)\|}{s_k-r_k} \ge C$.

2d7 Exercise. Finalize this proof.[4]

---

[1] In fact, for every x such that $\|x\| = 1$ there exists a such that $\langle x, a \rangle = 1$ and $\forall y \;\; |\langle y, a \rangle| \le \|y\|$.

[2] Hint: $\|\gamma(t + \varepsilon) - \gamma(t_0)\| \le \|\gamma(t + \varepsilon) - \gamma(t)\| + \|\gamma(t) - \gamma(t_0)\|$.

[3] Zorich, Sect. 10.4.1.

[4] Hint: $\|\gamma(s) - \gamma(t)\| < C|s - t|$ for all s near t.

Still another way. Every $t \in [t_0, t_1]$ has a relative neighborhood $U_t = (t-\varepsilon_t, t+\varepsilon_t) \cap [t_0, t_1]$ in $[t_0, t_1]$ such that $\forall s \in U_t \; |\gamma(s) - \gamma(t)| \le C|s-t|$. By compactness of $[t_0, t_1]$ there exists a finite subcovering. It means, $s_0, \dots, s_m$ such that $t_0 = s_0 < s_1 < \dots < s_m = t_1$ and $r_1, \dots, r_m$ such that $s_0 < r_1 < s_1 < r_2 < \dots < r_m < s_m$ and $r_1 \in U_{s_0} \cap U_{s_1}$, $r_2 \in U_{s_1} \cap U_{s_2}$, ..., $r_m \in U_{s_{m-1}} \cap U_{s_m}$.

Now, at last, we stop (re)proving (2d1), (2d5) and start using them.

Let a mapping $f : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable (on $\mathbb{R}^n$). Then every differentiable path $\gamma : [t_0, t_1] \to \mathbb{R}^n$ leads to another differentiable path $f \circ \gamma : [t_0, t_1] \to \mathbb{R}^m$, and

$$(f \circ \gamma)'(t) = (Df)_{\gamma(t)}(\gamma'(t)) = \big(D_{\gamma'(t)} f\big)_{\gamma(t)}$$

by the chain rule 2b11 (and 2b2).

In particular, for a straight path from a to $b = a + h$ (for given $a, h \in \mathbb{R}^n$),

$$\gamma(t) = a + th \quad \text{for } t \in [0,1] \,,$$

we have $\gamma'(t) = h$ for all $t \in (0,1)$, thus,

$$(f \circ \gamma)'(t) = \big(D_h f\big)_{a+th} \,.$$

Now, (2d3) gives for $m = 1$

$$\text{(2d8)} \qquad \exists t \in (0,1) \;\; f(a+h) - f(a) = (D_h f)_{a+th} \,,$$

while (2d1) gives for arbitrary m

$$\text{(2d9)} \qquad |f(a+h) - f(a)| \le \sup_{0<t<1} \big|\big(D_h f\big)_{a+th}\big| \,.$$

Taking into account that $|(D_{\gamma'(t)} f)_{\gamma(t)}| = |(Df)_{\gamma(t)}(\gamma'(t))| \le \|(Df)_{\gamma(t)}\| \cdot |\gamma'(t)|$ we get

$$\text{(2d10)} \qquad |f(b) - f(a)| \le |b-a| \sup_{0<t<1} \|(Df)_{a+th}\| \,.$$

The same holds for $f : G \to \mathbb{R}^m$ provided that an open set G in $\mathbb{R}^n$ contains $[a,b] = \{a + th : 0 \le t \le 1\}$. Moreover, it is enough to require that G contains $(a,b) = \{a + th : 0 < t < 1\}$, and f extends by continuity to a, b. However, if G is not convex, it may happen that G contains a and b but not $[a,b]$. In this case we have

$$\text{(2d11)} \qquad |f(b) - f(a)| \le d_G(a,b) \sup_G \|Df\|$$

where d_G is the "shortest path metric" (recall Sect. 1c).

The same holds for finite-dimensional Euclidean vector or affine spaces, as well as normed vector spaces and the corresponding affine spaces.

2d12 Exercise. (a) Let S_1, S_2 be finite-dimensional affine spaces, $U \subset S_1$ a connected open set, and $f : U \to S_2$ 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).

2e Partial derivative

For a mapping $f : \mathbb{R}^n \to \mathbb{R}^m$ the k-th *partial derivative*

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

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.

2e1 Exercise. (a) Let $U \subset \mathbb{R}^2$ be a *convex* open set, $f : U \to \mathbb{R}^m$ differentiable, and $D_2 f = 0$ 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?

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]

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

For a mapping $f : X \to Y$ partial derivatives are well-defined only if X is a Cartesian space; but Y may be just an affine space (and then $D_k f : X \to \vec{Y}$).

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 $(Df)_{x_0}$.

2e2 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]

2e3 Exercise. Prove 2e2.[1]

2e4 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, $C^1(U) = C^1(U \to \mathbb{R})$.

Here $\mathbb{R}^n$ and $\mathbb{R}^m$ may be replaced with finite-dimensional affine spaces (and $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ with $\mathcal{L}(\vec{S}_1, \vec{S}_2)$).

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 $f, g \in C^1(U)$.

2e5 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]

2e6 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 $U \to \mathbb{R}^m$.

Composition of C^1 mappings is a C^1 mapping, as we'll see in 2g1.

2e7 Exercise. Let $U \subset \mathbb{R}^n$ be an open set and $f \in C^1(U \to \mathbb{R}^m)$. Then for every compact $K \subset U$,

$$\sup_{x,y \in K, x \ne y} \frac{|f(x) - f(y)|}{|x - y|} < \infty \,.$$

Prove it.[3]

2e8 Exercise. (a) Differentiate[4] a mapping $\mathbb{R}^2 \ni (r, \theta) \mapsto (r \cos\theta, r \sin\theta) \in \mathbb{R}^2$.

(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 $g : \mathbb{R}^2 \to \mathbb{R}$.

---

[1] Hint. The following lemma may help:
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 $o(|\cdot|)$. (See also 2a5(a).)

[2] Hint: use 2b10.

[3] Hint: otherwise $\frac{|f(x_k) - f(y_k)|}{|x_k - y_k|} \to \infty$, $x_k \to x$, $y_k \to y$.

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

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

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

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

2e9 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.

2e10 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.

2e11 Exercise. Consider the affine space $S_n = \{f : f^{(n)}(\cdot) = n!\}$ (a special case of 1b15 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 $f : E \to \mathbb{R}$ (but not $E \to \mathbb{R}^m$) whenever E is an n-dimensional *Euclidean* affine space. In contrast to the derivative, the gradient is ill-defined if E is just an affine space, vector space, or even a normed (but not Euclidean) 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

$$|(D_h f)_{x_0}| \le |\nabla f(x_0)| \quad \text{for } |h| = 1 \, ;$$

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 (2d8), $f(b) - f(a) = \langle \nabla f(\xi), b - a \rangle$ for some $\xi \in (a, b)$; thus,

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

(compare it with (2d10)).

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 $m = 1$...

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

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.

2f2 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;

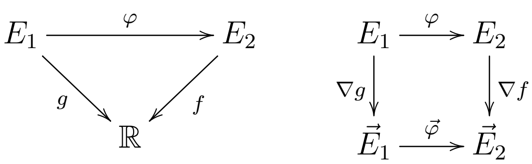
(b) the linear operator $(Df)_{x_0}$ maps $\mathbb{R}^n$ *onto* $\mathbb{R}^m$.

Proof. We have (recall 2b4)

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

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 (but $c \ne 0$). 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). □

2f3 Exercise. (a) Let E_1, E_2 be Euclidean affine spaces, $\varphi : E_1 \to E_2$ an isomorphism, $f : E_2 \to \mathbb{R}$ a differentiable function, and $g(\cdot) = f(\varphi(\cdot)) : E_1 \to \mathbb{R}$. Then $\vec{\varphi}\big(\nabla g(x)\big) = \nabla f\big(\varphi(x)\big)$ for all $x \in E_1$.



Prove it twice: first, via the isomorphism argument[1] and second, via 2b22(a).[2]

(b) Let $E = (V, |\cdot|)$ be a Euclidean vector space, $c \in (0, \infty)$, and $E_1 = (V, c|\cdot|)$ another Euclidean vector space. Then

$$\nabla^E f(x) = c^2 \nabla^{E_1} f(x)$$

for every $x \in E$ and every differentiable $f : V \to \mathbb{R}$; here ∇^E is the gradient on E, and ∇^{E_1} the gradient on E_1. Prove it.[3]

(c) The same as (a), but φ is not isometric; rather, $|\varphi(x) - \varphi(y)| = c|x - y|$ for all $x, y \in E_1$ (and a given $c \in (0, \infty)$). Then

$$\vec{\varphi}\big(\nabla g(x)\big) = c^2 \nabla f\big(\varphi(x)\big)$$

[1] Hint: given a Euclidean affine space E, consider triples (x, f, a) satisfying $\nabla f(x) = a$.

[2] Hint: $\langle a, b\rangle = \langle \vec{\varphi} a, \vec{\varphi} b\rangle$ for all $a, b \in \vec{E}$.

[3] Hint: $\langle a, b\rangle_1 = c^2 \langle a, b\rangle$ for all $a, b \in V$.

for all $x \in E_1$. Prove it.[1]
From now on, $E_1 = E_2$ and $f = g$.

(d) What about $f(ax) = bf(x)$ for given $a \in (1, \infty)$ and $b \in (-\infty, 0) \cup (0, \infty)$?

(e) Find the gradient of the function $x \mapsto |x|^p$ on $\mathbb{R}^n \setminus \{0\}$, and check (d) on this case.[2]

2g Higher order derivatives

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 finite-dimensional vector of affine space.

By induction,

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

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

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

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[3] $i_1, \dots, i_k \in \{1, \dots, n\}$.

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

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

2g1 Proposition. If $f \in C^k(\mathbb{R}^n \to \mathbb{R}^m)$ and $g \in C^k(\mathbb{R}^m \to \mathbb{R}^\ell)$ then $g \circ f \in C^k(\mathbb{R}^n \to \mathbb{R}^\ell)$.

***Proof** (sketch).* Induction in k. For $k = 0$: composition of continuous mappings is continuous. For $k > 0$: we'll prove that the derivative of $g \circ f$ (that is, the mapping $x \mapsto (D(g \circ f))_x$) belongs to $C^{k-1}(\mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n \to \mathbb{R}^\ell))$. By the chain rule 2b11, $(D(g \circ f))_x = (Dg)_{f(x)} \circ (Df)_x$; we know that $Df \in C^{k-1}(\mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n \to \mathbb{R}^m))$; it is sufficient to prove that the mapping $x \to (Dg)_{f(x)}$ belongs to $C^{k-1}(\mathbb{R}^n \to \mathcal{L}(\mathbb{R}^m \to \mathbb{R}^\ell))$, since the product

[1] Hint: use (a) and (b).
[2] Hint: 2b17(c) or 2b20.
[3] Not necessarily different.

of two matrices is a *polynomial* function of these matrices. We apply the claim for $k-1$ to the composition of $f \in C^k(\mathbb{R}^n \to \mathbb{R}^m) \subset C^{k-1}(\mathbb{R}^n \to \mathbb{R}^m)$ and the mapping $Dg \in C^{k-1}(\mathbb{R}^m \to \mathcal{L}(\mathbb{R}^m \to \mathbb{R}^\ell))$. □

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]

Index

[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 2a2.