1 Preliminaries

1a Conventions, notation, terminology etc.

$\mathbb{R}$ . . . . the real line
$\mathbb{R}^n$ . . . . $\{(x_1, \dots, x_n) : x_1, \dots, x_n \in \mathbb{R}\}$
Thus, $\mathbb{R}^{m+n} = \mathbb{R}^m \times \mathbb{R}^n$ up to canonical isomorphism.[1]
$A \subset B$ . . . . $\forall x \; (x \in A \implies x \in B)$
Thus, $(A \subset B) \wedge (B \subset A) \iff (A = B)$.[2]
$A \uplus B$ . . . . just $A \cup B$ when $A \cap B = \emptyset$, otherwise undefined.
$(1, \dots, n)$ or $(x_1, \dots, x_n)$ . . . . finite sequence
$(1, 2, \dots)$ or $(x_1, x_2, \dots)$ . . . . infinite sequence
$f : A \to B$ . . . . $f \subset A \times B$ and $\forall x \in A \; \exists!\, y \in B \;\; (x, y) \in f$.[3]
Tx . . . . the same as $T(x)$ when a mapping T is linear.
$|x|$ (for $x \in \mathbb{R}^n$) . . . . $\sqrt{x_1^2 + \dots + x_n^2}$ Euclidean norm
$\langle x, y \rangle$ (for $x, y \in \mathbb{R}^n$) . . . . $x_1 y_1 + \dots + x_n y_n$ scalar product
A°, $\overline{A}$ (for $A \subset \mathbb{R}^n$) . . . . the interior and the closure
near a point . . . . in some neighborhood of the point

Index of terminology and notation is often available at the end of a section.

[1] 'a rule of thumb: there is a canonical isomorphism between X and Y if and only if you would feel comfortable writing "X = Y"' — Reid Barton, see Mathoverflow, What is the definition of "canonical"?

[2] Why "$\subset$" and "$\subsetneqq$" rather than "$\subseteq$" and "$\subset$"? First, our textbooks do so; second, I need "$\subset$" several times a day, while "$\subsetneqq$" hardly once a month.

[3] Here B is the codomain, generally not the image of f.

1b Linear algebra

Vector space (=linear space) (usually, over $\mathbb{R}$)
Linear operator (=mapping=function) between vector spaces
Isomorphism of vector spaces: a linear bijection.
Basis of a vector space
Dimension of a finite-dimensional vector space: the number of vectors in every basis.
Two finite-dimensional vector spaces are isomorphic if and only if their dimensions are equal.
Subspace of a vector space.
Inner product on a vector space: $\langle x, y\rangle$

A basis of a subspace, being a linearly independent system, can be extended to a basis of the whole finite-dimensional vector space.

1c Topology

A sequence of points of $\mathbb{R}^n$; its convergence, limit
Mapping $\mathbb{R}^n \to \mathbb{R}^m$; continuity (at a point; on a set)
Cauchy criterion of convergence
Subsequence; Bolzano-Weierstrass theorem
Subset of $\mathbb{R}^n$, its limit points; closed set; bounded set
Compact set
Open set
Closure, boundary, interior
Open cover; Heine-Borel theorem
Open ball, closed ball, sphere

1c1 Exercise. Prove or disprove: a mapping $f : \mathbb{R}^2 \to \mathbb{R}$ is continuous if and only if it is continuous in each coordinate separately; that is, $f(x, \cdot) : \mathbb{R} \to \mathbb{R}$ is continuous for every x, and $f(\cdot, y) : \mathbb{R} \to \mathbb{R}$ is continuous for every y.

1c2 Exercise. (a) Prove that finite union of closed sets is closed, but union of countably many closed sets need not be closed; moreover, every open set in $\mathbb{R}^n$ is such union. However, intersection of closed sets is always closed.

(b) Formulate and prove the dual statement (take the complement).

1c3 Exercise. Prove that a set $K \subset \mathbb{R}^n$ is compact if and only if *every* continuous function $f : K \to \mathbb{R}$ is bounded.

1c4 Exercise. Prove that a continuous image of a compact set is compact, but a continuous image of a bounded set need not be bounded, and a continuous image of a closed set need not be closed; moreover, every open set in $\mathbb{R}^n$ is a continuous image of a closed set.[1]

1c5 Exercise. Prove that every decreasing sequence of nonempty compact sets has a nonempty intersection. Does it hold for closed sets? for open sets?

1c6 Exercise. Let $X \subset \mathbb{R}^n$ be a closed set, $f : X \to \mathbb{R}^m$ a continuous mapping. Prove that its *graph* $\Gamma_f = \{(x, f(x)) : x \in X\}$ is a closed subset of $\mathbb{R}^{n+m}$. Is the converse true?

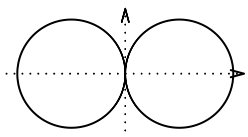
1c7 Exercise. Formulate accurately and prove: composition of two continuous mappings is continuous.

1c8 Exercise. Prove existence of a bijection f from the open unit ball $\{x : |x| < 1\} \subset \mathbb{R}^n$ onto the whole $\mathbb{R}^n$ such that f and f^{-1} are continuous. (Such mappings are called *homeomorphisms*). What about the closed ball?

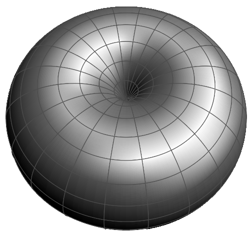
1c9 Exercise. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous bijection. Prove that $f^{-1} : \mathbb{R} \to \mathbb{R}$ is continuous.

1c10 Exercise. Give an example of a continuous bijection $f : [0, 1) \to S^1 = \{(x, y) : x^2 + y^2 = 1\} \subset \mathbb{R}^2$ such that $f^{-1} : S^1 \to [0, 1)$ fails to be continuous. The same for $f : [0, \infty) \to S^1$.

1c11 Exercise. Give an example of a continuous bijection $f : \mathbb{R} \to A = \{(x, y) : (|x| - 1)^2 + y^2 = 1\} \subset \mathbb{R}^2$ such that $f^{-1} : A \to \mathbb{R}$ fails to be continuous.



1c12 Exercise. Give an example of a continuous bijection $f : \mathbb{R}^2 \to B = \{(x, y, z) : (\sqrt{x^2 + y^2} - 1)^2 + z^2 = 1\} \subset \mathbb{R}^3$ such that $f^{-1} : B \to \mathbb{R}^2$ fails to be continuous.[2]



---

[1] Hint: the closed set need not be connected.

[2] What about a continuous bijection $f : \mathbb{R}^n \to \mathbb{R}^n$? In fact, f^{-1} is continuous, which can be proved using powerful means of topology (the Brouwer invariance of domain theorem); we'll return to this point later.

1d Differentiation

$$f : \mathbb{R}^n \to \mathbb{R}^m \,; \qquad \begin{aligned} f(x) &= f(x_0) + A(x - x_0) + o(|x - x_0|) \,, \quad \text{or} \\ f(x+h) &= f(x) + Ah + o(|h|) \,; \end{aligned}$$

A a matrix, or a linear mapping $\mathbb{R}^n \to \mathbb{R}^m$

$A = (Df)_x = Df(x) = df(x) = \text{(etc)} : \mathbb{R}^n \to \mathbb{R}^m$ derivative, or differential

$Ah = A(h) = (D_h f)_x = (Df)_x h = Df(x)h = df(x,h) = \text{(etc)} \in \mathbb{R}^m$ derivative along vector

$D_k f = D_{e_k} f \in \mathbb{R}^m, \quad e_k = (0, \dots, 0, 1, 0, \dots, 0)$ partial derivative

$(Df)_x h = h_1 (D_1 f)_x + \dots + h_n (D_n f)_x \in \mathbb{R}^m$ since $h = h_1 e_1 + \dots + h_n e_n$

$(Df)_x = \big((D_1 f)_x, \dots, (D_n f)_x \big)$ (columns of matrix)

$f(x) = \begin{pmatrix} f_1(x) \\ \dots \\ f_m(x) \end{pmatrix}; \qquad (Df)_x = \begin{pmatrix} (Df_1)_x \\ \dots \\ (Df_m)_x \end{pmatrix}$ (rows of matrix)

$(Df)_x = \big((D_j f_i)_x \big)_{i=1,\dots,m, j=1,\dots,n}$ (elements of matrix)

$\big(D(f+g) \big)_x = (Df)_x + (Dg)_x, \quad \big(D(cf) \big)_x = c(Df)_x$ linearity of D

$\big(D(g \circ f) \big)_x = (Dg)_{f(x)} (Df)_x$ chain rule

For $m = 1$ only: $(D_h f)_x = \langle \nabla f(x), h \rangle; \quad \nabla f(x) \in \mathbb{R}^n$ gradient

$\nabla \big(f(x) g(x) \big) = f(x) \nabla g(x) + g(x) \nabla f(x)$ product rule

For $n = 1$ only: $(Df)_x h = h f'(x), \quad f'(x) \in \mathbb{R}^m, \quad h \in \mathbb{R}.$

If $D_1 f, \dots, D_n f$ exist and are continuous, then Df exists (and is continuous).[1]

If $D_i D_j f$ and $D_j D_i f$ exist and are continuous, then $D_i D_j f = D_j D_i f$.[2]

1d1 Exercise. Generalize the product rule[3]

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

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

Some clarifications

For Df to be defined *at* x it is *necessary* that f is defined *near* x. If f is defined on a set with empty interior, we have no Df. For example, consider the mapping from the cylinder $C = \{(x,y,z) : x^2 + y^2 = 1, -1 < z < 1\}$ to the sphere $S = \{(x,y,z) : x^2 + y^2 + x^2 = 1\}$, defined by $f(x,y,z) = \big(x\sqrt{1-z^2}, y\sqrt{1-z^2}, z \big)$. As you'll see in Analysis-4, in this case $(Df)_x$ for

[1] Moreover, if $D_1 f, \dots, D_n f$ exist *near* x_0 and are continuous *at* x_0, then Df exists *at* x_0. (Zorich, Sect. 8.4.2, Th. 2.)

[2] Moreover, if $D_i D_j f$ exists *near* x_0 and is continuous *at* x_0, then $D_j D_i f$ exists *at* x_0, and $(D_i D_j f)_{x_0} = (D_j D_i f)_{x_0}$. (Courant, Sect. 1.4d.)

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

$x \in C$ is a linear operator[1] from the tangent plane $T_x C$ to C to the tangent plane $T_{f(x)}S$ to S. But it is not a 3×3 matrix, and is beyond Analysis-3. Never mind (until you reach Analysis-4). But note that linear operators will be more useful than matrices.[2]

For $f : \mathbb{R}^n \to \mathbb{R}^m$ we have $Df : \mathbb{R}^n \to \mathbb{R}^n \to \mathbb{R}^m$ ("currying"), in the sense that $Df : x \mapsto \big(h \mapsto (Df)_x h \big)$. Sometimes we treat it as a function of x, sometimes as a function of h. For example, it is usual to say that "if f is linear then $Df = f$". Really?! For $m = n = 1$ we know that $(\mathrm{e}^x)' = \mathrm{e}^x$, while $x' \ne x$, $x' = 1$ (a constant). What happens?

For $f(x) = \mathrm{e}^x$ we have $(Df)_x = f'(x) = \mathrm{e}^x$, but this e^x is treated as a 1×1 matrix (e^x), thus, the linear mapping $h \mapsto \mathrm{e}^x h$;

$$D\big(x \mapsto \mathrm{e}^x \big) : x \mapsto \big(h \mapsto \mathrm{e}^x h \big) \, .$$

For $g(x) = x$ we have $(Dg)_x = g'(x) = 1 : h \mapsto 1 \cdot h$;

$$D(\underbrace{x \mapsto x}_{\text{id}}) : \underbrace{x \mapsto (\underbrace{h \mapsto h}_{\text{id}})}_{\text{const}} \, .$$

In some sense this is id, and in another sense this is const.

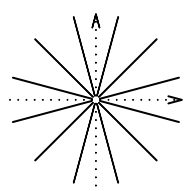
It is also usual to say that "the differential of the composition is the composition of differentials". Really?! For $m = n = 1$ we know that $(\mathrm{e}^{\sin x})' = \mathrm{e}^{\sin x} \cos x \ne \mathrm{e}^{\cos x}$. Yes, but one means that, given $f(x) = y$ and $g(y) = z$, we have $\big(D(g \circ f) \big)_x = (Dg)_y \circ (Df)_x$ (the chain rule); and it is usual to write AB rather than $A \circ B$ when A, B are linear operators.

1d2 Exercise. Formulate accurately and prove the following two claims about a differentiable mapping $f : \mathbb{R}^n \to \mathbb{R}^m$:

(a) f is linear if and only if $Df = f$;

(b) f is linear if and only if $f(0) = 0$ and Df is constant.

1d3 Exercise.



Consider functions $f : \mathbb{R}^2 \setminus \{(0,0)\} \to \mathbb{R}$ constant on all rays from the origin; that is, $f(r \cos \varphi, r \sin \varphi) = h(\varphi)$ for some $h : \mathbb{R} \to \mathbb{R}$, $h(\varphi + 2\pi) = h(\varphi)$. Assume that h is continuous.

(a) Prove that the iterated limits

$$\lim_{x \to 0+} \lim_{y \to 0+} f(x,y) \quad \text{and} \quad \lim_{y \to 0+} \lim_{x \to 0+} f(x,y)$$

---

[1]Not isometric, but preserves the area.

[2]Zorich requires f to be defined near x in Sect. 8.2.2 and later, but not in Sect. 8.2.1 (thus, Df need not be unique in 8.2.1).

exist and are equal to $h(0)$ and $h(\pi/2)$ respectively.

(b) prove that the "full" limit

$$\lim_{(x,y)\to(0,0),x>0,y>0} f(x,y)$$

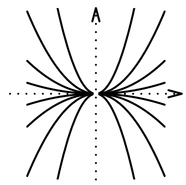
exists if and only if h is constant on $[0,\pi/2]$.

(c) It can happen that the two iterated limits exist and are equal, but the "full" limit does not exist. Give an example.

(d) The same as (c) and in addition, f is a rational function (that is, the ratio of two polynomials).[1]

(e) Generalize all that to arbitrary (not just positive) x, y.

1d4 Exercise.
Consider functions $g : \mathbb{R}^2 \setminus \{(0,0)\} \to \mathbb{R}$ of the form $g(x,y) = f(x^2, y)$ where f is as in 1d3.



(a) Prove that the limit

$$\lim_{t\to0+} g(ta, tb)$$

exists for every $(a,b) \ne (0,0)$; calculate the limit in terms of the function h of 1d3.

(b) It can happen that the "full" limit

$$\lim_{(x,y)\to(0,0)} g(x,y)$$

does not exist. Give an example.

1d5 Exercise. [2] 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.[3]

1d6 Exercise. [4] 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.[5]

> "The multivariate derivative is truly a pan-dimensional construct, not just an amalgamation of cross sectional data."
>
> (Shurman, p.156)

[1]Hint: try $x^2 + y^2$ in the denominator.

[2]Shurman, Ex.4.8.10.

[3]Hint: try $(x,y) \mapsto f(x,y)\sqrt{x^2+y^2}$ for f as in 1d3.

[4]Shurman, Ex.4.8.11.

[5]Hint: try $(x,y) \mapsto f(x,y)\sqrt{x^2+y^2}$ for f as in 1d4.

1e Textbooks to 1b, 1c, 1d

* R. Courant, F. John "Introduction to calculus and analysis" vol. 2, Springer 1989.
* W. Fleming "Functions of several variables" Springer 1977.
* J. Hubbard, B. Hubbard "Vector calculus, linear algebra, and differential forms" Prentice-Hall 2002.
* S. Lang "Undergraduate analysis" Springer 1997.
* T. Shifrin "Multivariable mathematics" Wiley 2005.
* J. Shurman "Multivariable calculus" (online only).
* V. Zorich "Mathematical analysis I" Springer 2004.

Textbook	linear algebra	topology	differentiation
Courant	2.1–2.3	1.1–1.3; A.1–A.3	1.4–1.7
Fleming	1.2–1.3	1.4; 2.1–2.5; ≈ 2.8, 2.11	3.1–3.3; 4.1–4.4
Hubbard	1.4	1.5–1.6	1.7–1.9
Lang	6.1–6.3	6.4–7.2; 8	15.1–15.2; ≈ 17
Shifrin	1; 4.3; 5.1–5.3	2; 5.1	3
Shurman	2.1–2.2; 3.1–3.2; 3.5–3.7	2.3–2.4	4.1–4.7.1; 4.8
Zorich	8.1	7	8.2–8.4.4

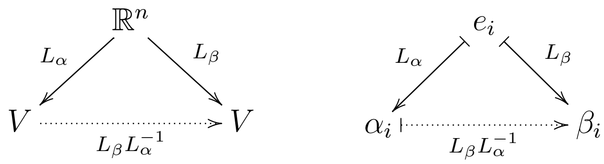
1f Change of basis

LINEAR ALGEBRA

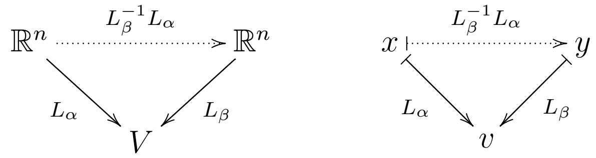
Let V be an n-dimensional vector space, and $(\alpha_1, \dots, \alpha_n)$ a basis of V. Then each $v \in V$ is $x_1\alpha_1 + \dots + x_n\alpha_n$ for some $x_1, \dots, x_n \in \mathbb{R}$, uniquely determined by v, and the mapping $L_\alpha : \mathbb{R}^n \to V$ defined by $L_\alpha(x_1, \dots, x_n) = x_1\alpha_1 + \dots + x_n\alpha_n$, is an isomorphism (of vector spaces). One says that these $x_1, \dots, x_n$ are the coordinates of v w.r.t. this basis, and $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ is the coordinate vector of v relative to this basis.

In particular, if $V = \mathbb{R}^n$ and $(\alpha_1, \dots, \alpha_n)$ is the standard basis $(e_1, \dots, e_n)$ of $\mathbb{R}^n$, then $L_\alpha = \mathrm{id}$, that is, $L_\alpha(x_1, \dots, x_n) = (x_1, \dots, x_n)$. In general, $L_\alpha(e_i) = \alpha_i$ for $i = 1, \dots, n$.

Another basis $(\beta_1, \dots, \beta_n)$ of V leads to another isomorphism $L_\beta : \mathbb{R}^n \to V$, $L_\beta(e_i) = \beta_i$; and then we have



That is, $L_\beta L_\alpha^{-1} : V \to V$, $L_\beta L_\alpha^{-1} \alpha_i = \beta_i$. This is the so-called active transformation of V that transforms $(\alpha_1, \dots, \alpha_n)$ to $(\beta_1, \dots, \beta_n)$. On the other hand we have



$x_1\alpha_1 + \dots + x_n\alpha_n = v = y_1\beta_1 + \dots + y_n\beta_n$; $L_\beta^{-1} L_\alpha : \mathbb{R}^n \to \mathbb{R}^n$, $L_\beta^{-1} L_\alpha(x_1, \dots, x_n) = (y_1, \dots, y_n)$. This is the so-called passive transformation of $\mathbb{R}^n$ that transforms the coordinate vector (of arbitrary $v \in V$) relative to one basis into the coordinate vector (of the same v) relative to the other basis.

Let $A = (a_{i,j})_{i,j}$ be the matrix of the operator $L_\beta^{-1} L_\alpha : \mathbb{R}^n \to \mathbb{R}^n$; that is, $y_i = \sum_j a_{i,j} x_j$. Then $\sum_j x_j \alpha_j = v = \sum_i y_i \beta_i = \sum_{i,j} a_{i,j} x_j \beta_i = \sum_j x_j \sum_i a_{i,j} \beta_i$, that is,

$$\alpha_j = \sum_i a_{i,j} \beta_i \, .$$

We see that A describes both the passive transformation and the relation between the two bases.[1]

1f1 Exercise. [2] Consider the 2-dimensional vector subspace $V = \{(x, y, z) : x + y + z = 0\}$ of $\mathbb{R}^3$, and two bases:

$$\begin{aligned} \alpha_1 &= (1, -1, 0) \, , \\ \alpha_2 &= (1, 0, -1) \, , \end{aligned} \qquad \text{and} \qquad \begin{aligned} \beta_1 &= (0, 1, -1) \, , \\ \beta_2 &= (1, 1, -2) \, . \end{aligned}$$

Find the change-of-basis matrix A.

1f2 Exercise. Consider the 3-dimensional vector space V of all functions $P : \mathbb{R} \to \mathbb{R}$ such that $\forall x \; P'''(x) = 0$, and two coordinate systems on V:

$$P \mapsto \big(P(0), P'(0), P''(0) \big) \qquad \text{and} \qquad P \mapsto \big(P(-1), P(0), P(1) \big) \, .$$

Find the two bases of V (that correspond to these coordinate systems), and the change-of-basis matrix.

[1] See also: "Change of basis" and "Active and passive transformation" in Wikipedia; Hubbard Sect. 2.6.

[2] Hubbard 2.6.17. A quote therefrom:
Note that unlike $\mathbb{R}^3$, for which the "obvious" basis is the standard basis vectors, the subspace $V \subset \mathbb{R}^3$ in Example 2.6.17 does not come with a distinguished basis.

TOPOLOGY

We may transfer all topological notions from $\mathbb{R}^n$ to arbitrary n-dimensional vector space. For example, consider the space V of quadratic polynomials (Exer. 1f2); given $P, P_k \in V$, we may interpret $P_k \to P$ as $P_k(0) \to P(0)$, $P'_k(0) \to P'(0)$, $P''_k(0) \to P''(0)$. Or alternatively, as $P_k(-1) \to P(-1)$, $P_k(0) \to P(0)$, $P_k(1) \to P(1)$. Is it the same? Yes, it is, as we'll see soon.

1f3 Exercise. (a) Every linear mapping $\mathbb{R}^n \to \mathbb{R}^m$ is continuous;

(b) every invertible linear mapping $\mathbb{R}^n \to \mathbb{R}^n$ is a homeomorphism (that is, continuous invertible mapping with continuous inverse).

Prove it.

1f4 Exercise. Every homeomorphism $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ preserves topological notions; namely:

$x_k \to x \quad \Longleftrightarrow \quad \varphi(x_k) \to \varphi(x)$;

A is open $\Longleftrightarrow$ $\varphi(A)$ is open; and the same for "closed", and "compact";

$\varphi(A^\circ) = \big(\varphi(A)\big)^\circ$; $\varphi(\overline{A}) = \overline{\varphi(A)}$; and $\varphi(\partial A) = \partial\big(\varphi(A)\big)$ (the boundary, $\partial A = \overline{A} \setminus A^\circ$).

Prove it.

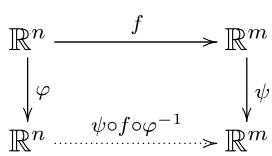
We apply this, in particular, to $\varphi = L_\beta^{-1} L_\alpha$, and conclude.

Topological notions in $\mathbb{R}^n$ are insensitive to a change of basis. Topological notions are well-defined in every n-dimensional vector space, and preserved by isomorphisms of these spaces.

1f5 Exercise. Every (vector) subspace of a finite-dimensional vector space is closed (topologically).

Prove it.[1,2]

A mapping $f : \mathbb{R}^n \to \mathbb{R}^m$ relates two spaces; accordingly, we introduce two homeomorphisms, $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ and $\psi : \mathbb{R}^m \to \mathbb{R}^m$,



getting a mapping $g = \psi \circ f \circ \varphi^{-1} : \mathbb{R}^n \to \mathbb{R}^m$.

[1] Hint: choose a basis.

[2] This claim fails in infinite dimension.

1f6 Exercise. (a) $\big(f \text{ is continuous} \big) \iff \big(g \text{ is continuous} \big)$;
(b) $\forall x \in \mathbb{R}^n \quad \big(f \text{ is continuous at } x \iff g \text{ is continuous at } \varphi(x) \big)$;
(c) $\forall x \in \mathbb{R}^n \quad \big(f \text{ is continuous near } x \iff g \text{ is continuous near } \varphi(x) \big)$.
Prove it.

Thus, when checking continuity of a given mapping, we may choose at will a pair of bases. This applies to any pair of finite-dimensional vector spaces. The case $m = n$ is not an exception; for $f : \mathbb{R}^n \to \mathbb{R}^n$ we still may use two different bases, thus treating f as a mapping between two copies of $\mathbb{R}^n$.

METRIC

A Euclidean metric on an n-dimensional vector space V may be defined equivalently as

* an inner product $x, y \mapsto \langle x, y \rangle$ on V;
* a norm $x \mapsto |x|$ on V that corresponds to some inner product by $|x|^2 = \langle x, x \rangle$; in this case the norm $|\cdot|$ is called Euclidean, and $\langle x, y \rangle = \frac12 \big(|x+y|^2 - |x|^2 - |y|^2 \big) = \frac14 \big(|x+y|^2 - |x-y|^2 \big)$;
* distance function $x, y \mapsto |x - y|$ that corresponds to some Euclidean norm $|\cdot|$.

On $\mathbb{R}^n$ we have the standard Euclidean metric, and the standard basis of $\mathbb{R}^n$ is orthonormal in this metric.

An arbitrary basis $(\alpha_1, \dots, \alpha_n)$ of a vector space V leads to the Euclidean metric $|x_1\alpha_1 + \dots + x_n\alpha_n| = \sqrt{x_1^2 + \dots + x_n^2}$, and is orthonormal in this (and only this) metric. On the other hand, for arbitrary Euclidean metric on V there exists an orthonormal basis (due to the orthogonalization process).

An n-dimensional vector space endowed with a Euclidean metric is called n-dimensional *Euclidean space.*

Let E be an n-dimensional Euclidean space. A basis $(\alpha_1, \dots, \alpha_n)$ of E is orthonormal if and only if the operator $L_\alpha : (x_1, \dots, x_n) \mapsto x_1\alpha_1 + \dots + x_n\alpha_n$ is *isometric,* that is, $\forall x \in \mathbb{R}^n \;\; |L_\alpha x| = |x|$. By isomorphism of Euclidean spaces we mean an isometric invertible linear operator. All n-dimensional Euclidean spaces are isomorphic (to each other, and to $\mathbb{R}^n$).

For arbitrary (not just isometric) invertible linear operator $L : E_1 \to E_2$ between Euclidean spaces there exist $a, b \in (0, \infty)$ such that

$$\forall x \in E_1 \;\; a|x| \le |Lx| \le b|x| \,. \tag{1f7}$$

Indeed, the ball $B = \{ x \in E_1 : |x| \le 1 \}$ is compact, therefore $L(B) \subset E_2$ is compact, which gives $b < \infty$. The same argument applies to $L^{-1} : E_2 \to E_1$, giving $1/a < \infty$.

It follows that two arbitrary Euclidean norms $|\cdot|_1$, $|\cdot|_2$ on a n-dimensional vector space V are equivalent:[1]

(1f8) $$\exists a, b \in (0, \infty)\ \forall x \in V\ \ a|x|_1 \le |x|_2 \le b|x|_1 \,.$$

Proof: apply 1f7 to $E_1 = (V, |\cdot|_1)$, $E_2 = (V, |\cdot|_2)$ and $L = \mathrm{id} : x \mapsto x$.

1f9 Exercise. Find an orthonormal basis in the space V of 1f1 with the standard Euclidean metric inherited from $\mathbb{R}^3$.

1f10 Exercise. Is it possible to endow V of 1f2 with a Euclidean metric such that both bases (mentioned in 1f2) are orthonormal?

SPACE OF MATRICES OR LINEAR OPERATORS

1f11 Definition. The norm $\|A\|$ of a linear operator $A : E_1 \to E_2$ between finite-dimensional Euclidean vector spaces E_1, E_2 is

$$\|A\| = \sup_{x \in E_1, x \ne 0} \frac{|Ax|}{|x|} \,.$$

Also,

$$\|A\| = \max_{|x| \le 1} |Ax|$$

(think, why); this is the maximum of a continuous function on a compact set.

The *operator norm* $\|A\|$ of a matrix $A : \mathbb{R}^n \to \mathbb{R}^m$ is, by definition, the norm of the corresponding operator.

1f12 Exercise. If a matrix $A = (a_{i,j})_{i,j}$ is diagonal then

$$\|A\| = \max_{i=1,\dots,\min(m,n)} |a_{i,i}|.$$

Prove it.

The set $\mathcal{L}(\mathbb{R}^n \to \mathbb{R}^m)$ of all matrices evidently is an mn-dimensional vector space. Does the operator norm turn it to a Euclidean space? No, it does not. Even if we restrict ourselves to $\mathcal{L}(\mathbb{R}^2 \to \mathbb{R}^2)$, and even to its 2-dimensional subspace of diagonal matrices, we get (by 1f12, up to isomorphism) $\mathbb{R}^2$ with the norm

$$\|(s, t)\| = \max(|s|, |t|) \,,$$

[1]In fact, two norms (Euclidean or not) are always equivalent in finite dimension (but not in infinite dimension).

its unit ball $\{x : \|x\| \le 1\}$ being the square $[-1,1] \times [-1,1]$. This is not the Euclidean plane! For two non-collinear vectors $a = (1,1)$ and $b = (1,-1)$ we have $\|a\| = 1$, $\|b\| = 1$ and $\|a+b\| = 2$, which never happens on the Euclidean plane. Also, the "parallelogram equality" $|a-b|^2 + |a+b|^2 = 2|a|^2 + 2|b|^2$ holds for arbitrary vectors a, b of a Euclidean space, but fails for the operator norm.

1f13 Exercise. Prove that $\|\cdot\|$ is a norm on $\mathcal{L}(\mathbb{R}^n \to \mathbb{R}^m)$, that is,

$$\begin{gathered} \|tA\| = |t| \cdot \|A\| \quad \text{for all } A \in \mathcal{L}(\mathbb{R}^n \to \mathbb{R}^m),\ t \in \mathbb{R}\,; \\ \|A+B\| \le \|A\| + \|B\| \quad \text{for all } A, B \in \mathcal{L}(\mathbb{R}^n \to \mathbb{R}^m)\,; \\ \|A\| > 0 \quad \text{whenever } A \ne 0\,. \end{gathered}$$

1f14 Exercise. Consider the composition $BA : E_1 \to E_3$ of two linear operators $A : E_1 \to E_2$ and $B : E_2 \to E_3$ between Euclidean spaces E_1, E_2, E_3; prove that $\|BA\| \le \|B\| \cdot \|A\|$.

Treating a matrix as just mn numbers, we have a Euclidean norm, the so-called *Hilbert-Schmidt norm* $\|A\|_{\text{HS}}$ of a matrix $A = (a_{i,j})_{i,j}$: $\|A\|_{\text{HS}} = \big(\sum_{i,j} a_{i,j}^2\big)^{1/2}$.

1f15 Exercise. (a) $\|A\|_{\text{HS}} = \sqrt{\operatorname{trace}(A^*A)}$;

(b) $\|A\| \le \|A\|_{\text{HS}} \le \sqrt{n}\|A\|$.[1]

Prove it.

Thus, the operator norm is equivalent to the Euclidean norm; both may be used when dealing with topological notions in $\mathcal{L}(\mathbb{R}^n \to \mathbb{R}^m)$.

1f16 Exercise. The following conditions on matrices $A, A_k \in \mathcal{L}(\mathbb{R}^n \to \mathbb{R}^m)$ are equivalent:

(a) $A_k \to A$;

(b) all elements of A_k converge to the corresponding elements of A; that is, $(A_k)_{i,j} \to A_{i,j}$ as $k \to \infty$ for all i, j.

Prove it.

1f17 Exercise. In the situation of 1f14 prove that BA is a continuous function of A, B, in two ways (via 1f14, and via 1f16).

[1] Hint to $\|A\| \le \|A\|_{\text{HS}}$: denoting the rows of A by $r_1, \dots, r_m \in \mathbb{R}^n$ we have $Ax = \begin{pmatrix} \langle r_1, x\rangle \\ \dots \\ \langle r_m, x\rangle \end{pmatrix}$. Hint to $\|A\|_{\text{HS}} \le \sqrt{n}\|A\|$: denoting the columns of A by $c_1, \dots, c_n \in \mathbb{R}^m$ we have $|c_j| \le \|A\|$ for each $j = 1, \dots, n$.

1f18 Exercise. (a) Determinant is a continuous function $A \mapsto \det A$ on $\mathcal{L}(\mathbb{R}^n \to \mathbb{R}^n)$;

(b) invertible operators are an open set;

(c) the mapping $A \mapsto A^{-1}$ is continuous on this open set.

Prove it.[1]

1f19 Exercise. If $A \in \mathcal{L}(\mathbb{R}^n \to \mathbb{R}^n)$ satisfies $\|A\| < 1$, then

(a) the series $\mathrm{id} - A + A^2 - A^3 + \dots$ converges in $\mathcal{L}(\mathbb{R}^n \to \mathbb{R}^n)$;

(b) the sum S of this series satisfies $(\mathrm{id} + A)S = \mathrm{id}$, $S(\mathrm{id} + A) = \mathrm{id}$; thus, $\mathrm{id} + A$ is invertible;

(c) $\det(\mathrm{id} + A) > 0$.

Prove it.[2]

DIFFERENTIATION

Looking at the definition of $(Df)_x$ for $f : \mathbb{R}^n \to \mathbb{R}$,

$$f(x+h) = f(x) + (Df)_x h + o(|h|) \,,$$

we observe that it does not involve any basis. True, it involves the Euclidean norm; but the notion $o(|h|)$ is insensitive to the choice of a norm due to (1f8), and we may write $o(h)$ instead of $o(|h|)$.

For $f : \mathbb{R}^n \to \mathbb{R}^m$, two norms appear:

$$\frac{|f(x+h) - f(x) - (Df)_x h|_{\mathbb{R}^m}}{|h|_{\mathbb{R}^n}} \to 0 \quad \text{as } h \to 0 \,,$$

and still, (1f8) ensures that both norms do not matter.

When differentiating a given mapping, we may choose at will a pair of bases. This applies to any pair of finite-dimensional vector spaces.

Here, by "differentiating" we mean checking differentiability and calculating the differential (interpreted as a linear operator, not matrix).

In contrast, partial derivatives (elements of the matrix of the linear operator) depend on the bases. Moreover, sometimes the partial derivative exist but the differential does not exist.

1f20 Exercise. It can happen that both partial derivatives of $f : \mathbb{R}^2 \to \mathbb{R}$ at $(0,0)$ vanish in the standard basis of $\mathbb{R}^2$, but do not vanish in another basis. Give an example.[3]

---

[1]Hint: recall the algebraic formulas for $\det A$ and A^{-1}.

[2]Hint: (c) consider $\det(\mathrm{id} + tA)$ for $t \in [0,1]$.

[3]Hint: similar to 1d5.

Looking at the definition of the gradient,

$$\langle \nabla f(x), h \rangle = (D_h f)_x \qquad \text{for } f : \mathbb{R}^n \to \mathbb{R}\,,$$

we observe that it does not involve any basis, but involves the Euclidean metric. And indeed, the gradient depends on the choice of the metric. It is well-defined for differentiable real-valued functions on a Euclidean space. Any *orthonormal* basis may be used equally well.

1f21 Exercise. On the space V of 1f2 consider the function $f : P \mapsto \int_{-1}^{1} P(t)\,\mathrm{d}t$. Find $\nabla f(0)$ twice, in the two bases mentioned in 1f2 (that is, relative to the two corresponding Euclidean metrics). Did you get two different elements of V?

1f22 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 vector spaces.

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

1f23 Exercise. For $f \in C^1(U \to \mathbb{R}^m)$ and $g \in C^1(\mathbb{R}^m \to \mathbb{R}^\ell)$ prove that $g \circ f \in C^1(U \to \mathbb{R}^\ell)$.[1]

1f24 Exercise. A mapping f is continuously differentiable if and only if all parial derivatives $D_i f_j$ exist and are continuous. (Here $f(x) = \big(f_1(x), \dots, f_m(x)\big)$.)

Prove it.

1f25 Exercise. (a) Let $f \in C^1(U)$ and $g \in C^1(U \to \mathbb{R}^m)$; prove that $fg \in C^1(U \to \mathbb{R}^m)$ (pointwise product).

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

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

1f26 Exercise. (a) Differentiate the mapping $\mathbb{R}^2 \ni (r, \theta) \mapsto (r\cos\theta, r\sin\theta) \in \mathbb{R}^2$.

(b) Differentiate the 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: chain rule, 1c7 and 1f17.

[2] Hint: use 1d1.

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

1f27 Exercise. [1] (a) Determinant is a continuously differentiable function $f : A \mapsto \det A$ on $\mathcal{L}(\mathbb{R}^n \to \mathbb{R}^n)$;

(b) $(Df)_{\mathrm{id}}(H) = \operatorname{tr}(H)$ for all $H \in \mathcal{L}(\mathbb{R}^n \to \mathbb{R}^n)$;

(c) $(D\log|f|)_A(H) = \operatorname{tr}(A^{-1}H)$ for all $H \in \mathcal{L}(\mathbb{R}^n \to \mathbb{R}^n)$ and all invertible $A \in \mathcal{L}(\mathbb{R}^n \to \mathbb{R}^n)$.

Prove it.

Thus,

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

for small H.

1f28 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.

1f29 Exercise. Consider the vector space $V_{n+1} = \{f : f^{(n+1)}(\cdot) = 0\}$ and the mapping $\varphi : \mathbb{R}^n \to V_{n+1}$,

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

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] Shurman:Ex.4.4.9

FROM MEAN VALUE TO FINITE INCREMENT

Recall the 1-dimensional mean value theorem: if $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$ and differentiable on (a, b), then $f(b) - f(a) = f'(t)(b - a)$ for some $t \in (a, b)$.

Applying this to the function $t \mapsto f\big(a+t(b-a)\big)$ we get the n-dimensional mean value theorem: if $G \subset \mathbb{R}^n$ is open, $f : \overline{G} \to \mathbb{R}$ is continuous on $\overline{G}$ and differentiable on G, and $a, b \in \overline{G}$ are such that $a + t(b - a) \in G$ for all $t \in (0, 1)$, then

$$f(b) - f(a) = (Df)_{a+t(b-a)}(b - a) = \langle \nabla f\big(a + t(b - a)\big), b - a\rangle$$

for some $t \in (0, 1)$; and therefore

$$\text{(1f30)}\quad |f(b) - f(a)| \le |b - a| \sup_{t\in(0,1)} \|(Df)_{a+t(b-a)}\| = \\ = |b - a| \sup_{t\in(0,1)} \big|\nabla f\big(a + t(b - a)\big)\big| \,.$$

Given open $G \subset \mathbb{R}^n$; a, b as before; and $f : \overline{G} \to \mathbb{R}^m$ continuous on $\overline{G}$ and differentiable on G, $f(x) = \big(f_1(x), \dots, f_m(x)\big)$, we may apply (1f30) to f_1 and get

$$|f_1(b) - f_1(a)| \le |b - a| \underbrace{\sup_{t\in(0,1)} \|(Df)_{a+t(b-a)}\|}_{C}$$

since $\|(Df_1)_x\| \le \left\| \begin{pmatrix} (Df_1)_x \\ \dots \\ (Df_m)_x \end{pmatrix} \right\| = \|(Df)_x\|$. The same holds for $f_2, \dots, f_m$, which implies easily $|f(b) - f(a)| \le C\sqrt{n}(b - a)$; but we can get more,

$$\text{(1f31)}\qquad |f(b) - f(a)| \le C|b - a| \,, \qquad \text{finite increment theorem}^1$$

just by changing the basis in $\mathbb{R}^m$ such that $f(b) - f(a)$ is proportional to the first basis vector!

[1]Zorich vol. 2, Sect. 10.4.1, Th. 1.

Index