Contents

CHAPTER 1
Introduction

1.1 Overview

Geometry is the study of "shapes" or "spaces" (e.g., subsets of $\mathbb{R}^n$ or more general metric spaces), including quantities such as distances, angles, areas, and volumes. There is often a focus on *manifolds*: spaces that look locally like $\mathbb{R}^n$ for some n.

Example 1.1.1 The surface of the Earth is a two-dimensional manifold embedded in $\mathbb{R}^3$. Indeed, from our perspective, at any given point on the surface, it is difficult to tell whether we are on $\mathbb{R}^2$ or on a sphere.

Example 1.1.2 The shape of a Y in $\mathbb{R}^2$ is not a manifold because it does not look like any $\mathbb{R}^n$ at the intersection of the three lines, although it looks like $\mathbb{R}$ on each line individually.

Here are some of the questions that are tackled in this book:

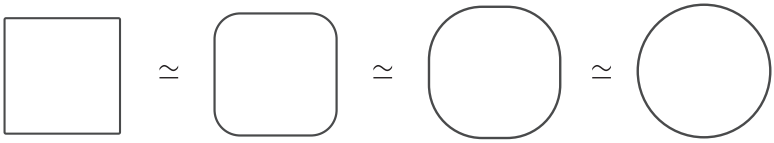
1. How can we tell if a given subset of $\mathbb{R}^n$ is a manifold?
2. For what values of n can we embed a given manifold in the Euclidean space $\mathbb{R}^n$?

With regard to question 1, we develop conditions for when the zero set or the image of a given smooth function $f : \mathbb{R}^m \to \mathbb{R}^n$ is a manifold. With regard to question 2, we will prove

Theorem 1.1.3 (Whitney's embedding theorem) *Every n-dimensional manifold can be embedded in* $\mathbb{R}^{2n+1}$.

Topology, or so-called rubber band geometry, is the study of topological properties of spaces. *Topological properties* are those that do not change under deformations like bending or stretching.

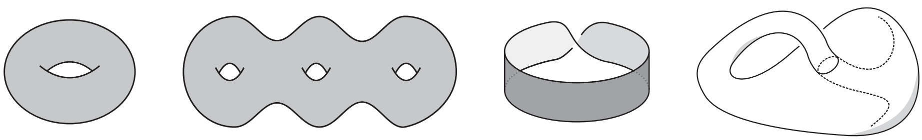
Example 1.1.4 The square and the circle have different geometry but the same topological properties, because one can be deformed into the other:



The triangle is also the same topologically as the square or the circle. Note that we are thinking of these objects (square, circle, and triangle) as the boundaries of the usual figures (not filled in). As such, they are one-dimensional manifolds.

Example 1.1.5 Two circles and one circle are not the same topologically because we cannot deform one circle into two without breaking it. This is intuitively obvious, and it can be made mathematically rigorous through the notions we develop later in the text.

Example 1.1.6 The surfaces of a cube and a sphere are again different in geometry but the same topologically. They are both two-dimensional manifolds. Here are a few other two-dimensional manifolds, all topologically different from the sphere and from each other:



The first is the surface of a donut, called a *torus*. The second is a three-holed torus. The third is the *Möbius band*, a band with a twist in it so that it ends up having a single face.

The fourth surface is the *Klein bottle*, which also has a single face, and it looks like it intersects itself, which makes it not quite look like a manifold. However, if we had one more dimension available (i.e., if we were in $\mathbb{R}^4$ rather than $\mathbb{R}^3$), we could lift the tube into the fourth dimension and make it not cut through the other part of the bottle. The result is a true two-dimensional manifold, the Klein bottle, embedded in $\mathbb{R}^4$. The Klein bottle cannot be embedded in $\mathbb{R}^3$, and it's hard to draw the fourth dimension, so the best we can do is to draw it as shown here.

Apart from the distinction between geometry and topology, we also distinguish between the study of *continuous* and *smooth* objects. Just as we have continuous and smooth (infinitely differentiable) functions, we have two notions of manifolds: *topological* (continuous) *manifolds* and *smooth manifolds*. For example, the square is a topological manifold (indeed, at any point we can move in two directions, just like on $\mathbb{R}$), but it has corners, so it does not look smooth.

Altogether, this justifies splitting the book into four parts:

- Chapter 2, "Metric Geometry," discusses metric spaces, which are generalizations of subsets of $\mathbb{R}^n$. The notion of distance plays an essential role. On the other hand, we do not involve any differentiability.
- Chapter 3, "Point-Set Topology," discusses topological spaces, which are more general than metric spaces and may not have a good notion of distance. We discuss their topological properties and introduce topological manifolds.
- Chapter 4, "Differential Topology," is focused on smooth manifolds, and smooth maps between them, without discussing distances.
- Chapter 5, "Differential Geometry," is about a smooth notion of distance (Riemannian metric), which applies to smooth manifolds.

There are many applications of these concepts. To list a few:

- **Physics**: *Cosmology* studies the shape of the universe, which is a three-dimensional manifold (four-dimensional if we include time). The more speculative *string theory* posits that the universe is ten-dimensional, with the extra six dimensions (beyond the usual space-time) curling up in a shape called a Calabi-Yau manifold.
- **Computer science**: In *data analysis*, data often clusters in the shape of a manifold, and we are interested in its metric and topological properties.
- **Biology and chemistry**: Molecules such as DNA, RNA, and various polymers can be knotted up in three-dimensional space. They are studied using mathematical knot theory, which deals with one-dimensional manifolds (knots and links) embedded in $\mathbb{R}^3$.
- **Economics**: Models for the economy take the form of manifolds, and a general equilibrium is the fixed point of a continuous transformation.
- **Math**: Geometry and topology are at the core of modern mathematics, being used in such disparate fields as number theory (e.g., elliptic curves), partial differential equations, combinatorics, and many others.

1.2 Prerequisites

The book assumes knowledge of multivariable calculus, real analysis, and linear algebra. The reader should be familiar with rigorous mathematical reasoning (proofs).

We list in the following a few concepts and facts that are used in later chapters and that are not always covered in standard courses.

1.2.1 Set Theory

A set X is called *countably infinite* if there exists a bijection $f: \mathbb{N} \to X$. A set X is called *countable* if it is either finite or countably infinite.

Proposition 1.2.1 *A subset of a countable set is countable.*

Proposition 1.2.2 *A countable union of countable sets is countable.*

Example 1.2.3 The set $\mathbb{Q}$ of rational numbers is countable because

$$\mathbb{Q} = \bigcup_{n \in \mathbb{N}} \{\tfrac{a}{n} \mid a \in \mathbb{Z}\}$$

is a countable union of countable sets.

Theorem 1.2.4 (Cantor, 1874) *The set $\mathbb{R}$ of real numbers is not countable.*

Proof. This is known as Cantor's diagonal argument. Assume $\mathbb{R}$ is countable. Then any $X \subseteq \mathbb{R}$ is also countable. For simplicity, we want to avoid numbers with two different decimal expansions, such as $.341\bar{9} = .342$, so let

$$X = \{x \in (0, 1) \mid x \text{ has no 9 in its decimal expansion}\}.$$

Given the bijection $f: \mathbb{N} \to X$, consider the decimal expansions of all $f(i) = x_i$:

$$\begin{aligned} x_1 &= .a_{11}a_{12}a_{13}\cdots \\ x_2 &= .a_{21}a_{22}a_{23}\cdots \\ x_3 &= .a_{31}a_{32}a_{33}\cdots \\ &\cdots \end{aligned}$$

For each i, pick some digit $y_i \notin \{a_{ii}, 9\}$. Let $y = .y_1y_2y_3\cdots$. Because $y \in X$, we must have $y = x_i$ for some i. But $y_i \neq a_{ii}$ gives a contradiction. $\square$

Let us also review the concept of equivalence relations.

Definition 1.2.5 A **binary relation** R on a set X is a subset $R \subseteq X \times X$. It is customary to write xRy instead of $(x, y) \in R$.

Definition 1.2.6 An **equivalence relation** on a set X is a binary relation $\sim$ satisfying the following properties:

1. Symmetry: For all $x \in X$, we have $x \sim x$.
2. Reflexivity: If $x \sim y$, then $y \sim x$.
3. Transitivity: If $x, y, z \in X$ are such that $x \sim y$ and $y \sim z$, then $x \sim z$.

Definition 1.2.7 Given a binary relation R on X, the equivalence relation generated by R is the smallest equivalence relation that contains R.

An equivalence relation determines a splitting of X as a disjoint union of equivalence classes, each of the form

$$[x_0] = \{x \in X \mid x \sim x_0\}.$$

One way to produce an equivalence relation on a set X is from a surjective function $p : X \to A$: We set $x \sim x'$ if $p(x) = p(x')$. Conversely, given an equivalence relation, we can associate to it the surjective function

$$p : X \to X^*, \quad p(x) = [x],$$

where X^* is the set of equivalence classes.

1.2.2 Linear Algebra

Let V and W be vector spaces with $\dim V = n$, $\dim W = m$. After choosing bases for V and W, a linear transformation $A : V \to W$ is represented by an $m \times n$ matrix,

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

If $m = n$, two important quantities associated to A are the *trace*,

$$\mathrm{tr}(A) = a_{11} + a_{22} + \cdots + a_{nn},$$

and the *determinant*,

$$\det(A) = \sum_{\sigma \in S_n} \varepsilon(\sigma) \cdot a_{1\sigma(1)} a_{2\sigma(2)} \dots a_{n\sigma(n)},$$

where

$$S_n = \{\sigma : \{1, \dots, n\} \to \{1, \dots, n\} \mid \sigma \text{ bijective}\}$$

is the set of permutations of n elements and

$$\varepsilon(\sigma) = (-1)^{\#\{(i,j) \mid i<j, \sigma(i)>\sigma(j)\}}$$

is the sign of the permutation σ.

Let us now go back to the general case of an $m \times n$ matrix representing $A : V \to W$.

Definition 1.2.8 The kernel of A is $\ker(A) = \{v \in V \mid A(v) = 0\}$. The dimension of $\ker(A)$ is called the **nullity** of A, denoted $\mathrm{null}(A)$.

Definition 1.2.9 The dimension of the image of A is called the **rank** of A, denoted $\mathrm{rk}(A)$.

Proposition 1.2.10 *The rank of A is the maximum number of linearly independent columns in A and the maximum number of linearly independent rows in A.*

Theorem 1.2.11 (rank-nullity theorem) *For any linear transformation $A : V \to W$, we have*

$$\mathrm{rk}(A) + \mathrm{null}(A) = \dim(V).$$

From Proposition 1.2.10, it follows that $\mathrm{rk}(A) \leq \min(m, n)$. A matrix A with $\mathrm{rk}(A) = \min(m, n)$ is called *of maximal rank*.

The following is an immediate consequence of the definition of rank:

Proposition 1.2.12 *When $m \leq n$, a matrix A is of maximal rank if and only if $A : V \to W$ is surjective.*

The following is a consequence of Theorem 1.2.11:

Proposition 1.2.13 *When $m \geq n$, a matrix A is of maximal rank if and only if $A : V \to W$ is injective.*

When $m = n$, the following conditions are equivalent:

- The matrix A is *invertible*, that is, there exists a matrix A^{-1} such that $AA^{-1} = A^{-1}A = I_n$.
- A is nonsingular, that is, the determinant is nonzero: $\det(A) \neq 0$.
- A is of maximal rank.
- A has zero kernel.
- The columns of A are linearly independent.
- The rows of A are linearly independent.

When $m \neq n$, the matrix A does not have a determinant, but we can use its minors to check its rank.

Definition 1.2.14 A $k \times k$ **minor** of an $m \times n$ matrix A is the determinant of a $k \times k$ matrix obtained from A by deleting $m - k$ rows and $n - k$ columns.

Theorem 1.2.15 *The rank of A is the maximum k such that A admits a nonzero $k \times k$ minor.*

Proof. This is a consequence of Proposition 1.2.10. If A has rank r, then there are r linearly independent rows and r linearly independent columns. The $r \times r$ matrix formed of those rows and columns must also have rank r, and therefore the respective minor is nonzero. Thus, $r \leq k$. Conversely, if a $k \times k$ minor is nonzero, then those k rows must be linearly independent, so $k \leq r$. □

Theorem 1.2.16 *Let $A : V \to W$ have rank k. Then there exists an invertible $n \times n$ matrix S and an invertible $m \times m$ matrix T such that, in block form,*

$$T^{-1}AS = \left(\begin{array}{c|c} I_k & O_{k,n-k} \\ \hline O_{m-k,k} & O_{m-k,n-k} \end{array}\right),$$

where $O_{i,j}$ denotes the zero $i \times j$ matrix.

Proof. The matrices S and T represent changes of basis on V and W. We choose S such that the last $n - k$ new basis vectors for V are a basis for $\ker(A)$ and the first k is a basis for a complement of $\ker(A)$. We then choose T such that the new basis for W starts with the images under A of the first k basis vectors for V. □

Let us also recall the explicit formula for the inverse.

Theorem 1.2.17 *Let $A = (a_{ij})$ be an $n \times n$ invertible matrix. Then the coefficient of A^{-1} in row i and column j is given by*

$$(A^{-1})_{ij} = (-1)^{i+j} \frac{\det(M_{ji})}{\det(A)},$$

where M_{ji} is the matrix obtained from A by deleting row j and column i.

In addition to representing linear maps, matrices can also be used to describe bilinear forms.

Definition 1.2.18 Let V be a real vector space. A **bilinear form** on V is a map $B: V \times V \to \mathbb{R}$ such that, for all $v \in V$, the maps

$$B(v, -) : V \to \mathbb{R}, \quad w \mapsto B(v, w)$$

and

$$B(-, v) : V \to \mathbb{R}, \quad w \mapsto B(w, v)$$

are linear. A bilinear form is called *symmetric* if

$$B(v, w) = (w, v) \text{ for all } v, w \in V.$$

Definition 1.2.19 An **inner product** on V is a symmetric bilinear form B: $V \times V \to \mathbb{R}$ that is positive definite; that is,

$$B(v, v) > 0$$

for all $v \in V$ with $v \neq 0$.

Given a basis $\{e_1, \ldots, e_n\}$ for V, the $n \times n$ matrix associated to a bilinear form B is

$$A = (a_{ij}), \quad a_{ij} = B(e_i, e_j).$$

The matrix A determines the bilinear form, by the formula

$$B(v, w) = v^T Aw.$$

If we change the basis to

$$f_1 = S(e_1), \ldots, f_n = S(e_n)$$

using a change-of-basis matrix S, the bilinear form B is represented in terms of the new basis by the matrix $S^T AS$.

Proposition 1.2.20 *Let A be the matrix representing a bilinear form B in terms of a basis $\{e_1, \ldots, e_n\}$. Then*

(a) *B is symmetric if and only if A is symmetric, that is, $A^T = A$.*

(b) *B is an inner product if and only if $A^T = A$ and the n diagonal minors,*

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1j} \\ a_{21} & a_{22} & \cdots & a_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ a_{j1} & a_{j2} & \cdots & a_{jj} \end{vmatrix},$$

are all positive, for $j=1,\dots,n$. (We then say that A is a symmetric positive definite matrix.)

Theorem 1.2.21 (Sylvester's law of inertia) *Let V be an n-dimensional vector space. To every symmetric bilinear form B on V, we can assign a unique triple of numbers (n_0, n_+, n_-), called the signature of B, such that there exists a basis of V in which B is represented by a diagonal matrix D whose diagonal entries are all 0 (appearing n_0 times), 1 (appearing n_+ times), and -1 (appearing n_- times).*

Note that there may be different bases in which B is represented by a diagonal matrix D as above. The uniqueness part of the theorem says that the numbers n_0, n_+, and n_- do not depend on the choice of such a basis.

Observe also that

$$n = n_0 + n_+ + n_-.$$

The symmetric bilinear form B is called *nondegenerate* if $n_0 = 0$; this is equivalent to the condition

$$(B(v,w)=0,\ \forall w \in V) \Rightarrow v = 0.$$

Furthermore, B is an inner product if and only if $n_0 = n_- = 0$ and $n_+ = n$.

Let us now suppose that V is equipped with an inner product $\langle\, ,\, \rangle$. A basis $\{e_1,\dots,e_n\}$ is called *orthogonal* if $\langle e_i, e_j\rangle = 0$ for all $i \neq j$. It is called *orthonormal* if it is orthogonal and $\langle e_i, e_i\rangle = 1$ for all i. A linear transformation $S: V \to V$ is orthogonal if it preserves the inner product, that is,

$$\langle Sv, Sw\rangle = \langle v, w\rangle \ \text{ for all } w, w \in V.$$

Equivalently, the corresponding matrix S (in terms of an orthogonal basis) should satisfy $S^T = S^{-1}$. Note that S is orthogonal if and only if it takes orthonormal bases to orthonormal bases.

Further, a linear transformation $A: V \to V$ is called *self-adjoint* if

$$\langle Av, w\rangle = \langle v, Aw\rangle \ \text{ for all } w, w \in V.$$

Equivalently, the corresponding matrix A (in terms of an orthogonal basis) should be symmetric; that is, it should satisfy $A^T = A$.

Theorem 1.2.22 (spectral theorem) *Every self-adjoint linear transformation A admits an orthonormal basis of eigenvectors; that is, in terms of matrices, if $A^T = A$, then there exists a change-of-basis matrix S such that $S^T = S^{-1}$ and $S^{-1}AS = D$ is diagonal.*

A symmetric matrix A can also be used to represent a symmetric bilinear form. Also, orthogonal matrices (with $S^T = S^{-1}$) change the corresponding linear transformations and bilinear forms in the same way: $A \mapsto S^{-1}AS$ and $A \mapsto S^T AS$ are the same. Thus, we can rephrase the spectral theorem as follows:

Proposition 1.2.23 *A symmetric bilinear form B admits an orthonormal basis in which B is represented by a diagonal matrix.*

1.2.3 Multivariate Analysis

We denote the Euclidean norm of vectors, $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, by

$$\|x\| = (x_1^2 + x_2^2 + \cdots + x_n^2)^{1/2}.$$

Let $\mathrm{Hom}(\mathbb{R}^m, \mathbb{R}^n)$ denote the space of linear maps from $\mathbb{R}^m$ to $\mathbb{R}^n$. This can be identified with the space $\mathbb{R}^{mn}$ of $m \times n$ matrices.

Definition 1.2.24 Let $U \subseteq \mathbb{R}^n$ be an open set and $x \in U$. A function $f\colon U \to \mathbb{R}^m$ is called **differentiable** at x if there exists a linear map $A \in \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m)$ such that

$$\lim_{h \to 0} \frac{\|f(x+h) - f(x) - A(h)\|}{\|h\|} = 0.$$

Then A is called the derivative of f at x, denoted $Df(x)$.

Example 1.2.25 If f is linear, then its derivative $Df(x)$ at any point is the function f itself. Indeed, by linearity, we have $f(x+h) - f(x) - f(h) = 0$, so the condition in the definition holds. This is somewhat confusing because linear maps are represented by matrices, which look constant, and one might expect their derivative to be 0. However, for example, the 1×1 matrix (6) is really representing the linear function $x \mapsto 6x$, whose derivative is 6.

A necessary (but not sufficient) condition for the existence of the derivative is that f is continuous at x.

If f is differentiable at all points $x \in U$, we say that f is differentiable on U. Its derivative takes the form of a map:

$$Df\colon U \to \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m) \cong \mathbb{R}^{mn}.$$

If Df is continuous, then f is called *continuously differentiable*, or of class C^1.

We can also ask about the existence of a second derivative, which would be a map

$$D^2f\colon U \to \mathrm{Hom}(\mathbb{R}^n, \mathrm{Hom}(\mathbb{R}^m, \mathbb{R}^n)) \cong \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^{mn}) \cong \mathbb{R}^{mn^2}.$$

Definition 1.2.26 A function $f\colon U \to \mathbb{R}^m$ is called **of class** C^k (or a C^k **function**) if it has k continuous derivatives. (When $k = 0$, this just means that f is continuous.)

The function f is called **smooth** (or **infinitely differentiable**, or of class C^∞) if it is in C^k for all $k \geq 0$, that is, if it has infinitely many derivatives.

Pick a nonzero vector $u = (u_1, \ldots, u_m)$. The *directional derivative* of f in the direction of u at x is defined to be

$$D_u f(x) = \lim_{t \to 0} \frac{f(x + tu) - f(x)}{t}.$$

When f is differentiable at x, this limit (the directional derivative) exists and is simply given by the formula

$$D_u f(x) = Df(x)(u).$$

In particular, when u is one of the basis vectors, $e_1, \ldots, e_n$, if we write f in terms of its components as $f = (f_1, \ldots, f_m)$, then

$$D_{e_j} f = (D_{e_j} f_1, \ldots, D_{e_j} f_m)$$

is the vector consisting of the *partial derivatives*

$$\partial_j f_i := D_{e_j} f_i = \frac{\partial f_i}{\partial x_j}.$$

Then, as a linear map, the derivative $Df(x)$ is represented by the matrix

$$Df(x) = \begin{pmatrix} (\partial_1 f_1)(x) & (\partial_2 f_1)(x) & \ldots & (\partial_n f_1)(x) \\ (\partial_1 f_2)(x) & (\partial_2 f_2)(x) & \ldots & (\partial_n f_2)(x) \\ \vdots & \vdots & \ddots & \vdots \\ (\partial_1 f_m)(x) & (\partial_2 f_m)(x) & \ldots & (\partial_n f_m)(x) \end{pmatrix}.$$

Remark 1.2.27 The partial derivatives (or, more generally, all directional derivatives) of a function at a point may exist without the function being differentiable at that point. For example, this happens for the function $f: \mathbb{R}^2 \to \mathbb{R}$, given by $f(0,0)=0$ and

$$f(x,y) = \frac{x^3}{x^2+y^2} \quad \text{if } (x,y) \neq (0,0).$$

Theorem 1.2.28 (the chain rule) *Let $U \subseteq \mathbb{R}^n$, $V \subseteq \mathbb{R}^m$, and $W \subseteq \mathbb{R}^p$ be open. If $f: U \to V$ is differentiable at x and $g: V \to W$ is differentiable at $f(x)$, then the composition $g \circ f: U \to W$ is differentiable at x, and we have*

$$D(g \circ f)(x) = (Dg)(f(x)) \cdot Df(x). \tag{1.1}$$

Definition 1.2.29 The **norm** of a linear map, $A \in \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m)$, is

$$\|A\| = \max_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|Ax\|}{\|x\|}. \tag{1.2}$$

One can show that the norm is indeed well defined; that is, the maximum in Equation (1.2) exists (as a finite real number): see problem 6 at the end of Section 2.4.

If $A = (a_{ij})$, note that we can get a rough bound on the norm in terms of the coefficients:

$$\|A\| \leq \sum_{i,j} |a_{ij}|.$$

Recall that for functions of a single variable we have the mean value theorem: If $f: [a,b] \to \mathbb{R}$ is continuous on $[a,b]$ and differentiable on (a,b), then

$$f(b) - f(a) = f'(c)(b-a)$$

for some $c \in (a,b)$. Therefore, if we have a bound $|f'| \leq M$ on the derivative, we get the inequality

$$|f(b) - f(a)| \leq M|b-a|. \tag{1.3}$$

Here is the generalization of this result to functions of several variables.

Theorem 1.2.30 *Let $U \subseteq \mathbb{R}^n$ be a convex open set and $f: U \to \mathbb{R}^m$ a differentiable function. Suppose there is a real number M such that*

$$\|Df(x)\| \leq M$$

for every $x \in U$. Then

$$\|f(b) - f(a)\| \leq M\|b - a\|$$

for all $a, b \in U$.

Proof. Consider the linear path between a and b:

$$\gamma : [0, 1] \to U, \quad \gamma(t) = (1 - t)a + tb,$$

which is contained in U because U is convex. Let $g(t) = f(\gamma(t))$. Using the chain rule, we have

$$\|g'(t)\| = \|Df(\gamma(t))(b - a)\| \leq \|Df(\gamma(t))\| \cdot \|b - a\| \leq M\|b - a\|.$$

From here we get

$$\|f(b) - f(a)\| = \|g(1) - g(0)\| = \left\| \int_0^1 g'(t)dt \right\| \leq \int_0^1 \|g'(t)\|dt \leq M\|b - a\|,$$

as desired. □

Remark 1.2.31 In this book, we often emphasize the derivative as a function of the direction u instead of the position x. When we do this (from Section 4.5 onward), we write $D_x f(u)$ in place of $D_u f(x)$. Then $D_x f$ is a linear function from $\mathbb{R}^n$ to $\mathbb{R}^m$, and the chain rule (1.1) becomes

$$D_x(g \circ f) = D_{f(x)} g \circ D_x f. \tag{1.4}$$

On the right-hand side, we use the symbol $\circ$ for composing linear functions, which corresponds to the multiplication symbol $\cdot$ for matrices in (1.1). Written like this, the chain rule looks quite different from what is learned in a typical calculus class, but it should be easier to remember: It just says that the linearization (i.e., the derivative) of a composition of functions is equal to the composition of their linearizations.

(continued...)

Index