2

The Seven C's of Analysis

2.1 Introduction

The current chapter explains key concepts of mathematical analysis summarized by the six adjectives convergent, complete, closed, compact, continuous, and connected. Chapter 6 will add to these six c's the seventh c, convex. At first blush these concepts seem remote from practical problems of optimization. However, painful experience and exotic counterexamples have taught mathematicians to pay attention to details. Fortunately, we can benefit from the struggles of earlier generations and bypass many of the intellectual traps.

2.2 Vector and Matrix Norms

In multidimensional calculus, vector and matrix norms quantify notions of topology and convergence [48, 105, 117, 207]. Norms are also helpful in estimating rates of convergence of iterative methods for solving linear and nonlinear equations and optimizing functions. Functional analysis, which deals with infinite-dimensional vector spaces, uses norms on functions.

We have already met the Euclidean vector norm $\|\boldsymbol{x}\|$ on R^n. For most purposes, this norm suffices. It shares with other norms the four properties:

(a) $\|\boldsymbol{x}\| \geq 0$,

(b) $\|\boldsymbol{x}\| = 0$ if and only if $\boldsymbol{x} = \mathbf{0}$,

K. Lange, *Optimization*, Springer Texts in Statistics 95,
DOI 10.1007/978-1-4614-5838-8_2,

(c) $\|c\boldsymbol{x}\| = |c| \cdot \|\boldsymbol{x}\|$ for every real number c,

(d) $\|\boldsymbol{x}+\boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$.

Property (d) is known as the triangle inequality. To prove it for the Euclidean norm, we note that the Cauchy-Schwarz inequality implies

$$\begin{aligned} \|\boldsymbol{x}+\boldsymbol{y}\|^2 &= \|\boldsymbol{x}\|^2 + 2\boldsymbol{x}^*\boldsymbol{y} + \|\boldsymbol{y}\|^2 \\ &\leq \|\boldsymbol{x}\|^2 + 2\|\boldsymbol{x}\|\|\boldsymbol{y}\| + \|\boldsymbol{y}\|^2 \\ &= (\|\boldsymbol{x}\| + \|\boldsymbol{y}\|)^2 . \end{aligned}$$

One immediate consequence of the triangle inequality is the further inequality

$$\big|\, \|\boldsymbol{x}\| - \|\boldsymbol{y}\| \,\big| \quad \leq \quad \|\boldsymbol{x}-\boldsymbol{y}\|.$$

Two other simple but helpful norms are the ℓ_1 and ℓ_∞ norms

$$\begin{aligned} \|\boldsymbol{x}\|_1 &= \sum_{i=1}^n |x_i| \\ \|\boldsymbol{x}\|_\infty &= \max_{1\leq i\leq n} |x_i|. \end{aligned}$$

Some of the properties of these norms are explored in the problems. In the mathematical literature, the three norms are often referred to as the ℓ_2, ℓ_1, and ℓ_∞ norms.

An $m\times n$ matrix $\boldsymbol{A} = (a_{ij})$ can be viewed as a vector in R^{mn}. Accordingly, we define its Frobenius norm

$$\|\boldsymbol{A}\|_F = \left(\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2\right)^{1/2} = \sqrt{\operatorname{tr}(\boldsymbol{A}\boldsymbol{A}^*)} = \sqrt{\operatorname{tr}(\boldsymbol{A}^*\boldsymbol{A})},$$

where $\operatorname{tr}(\cdot)$ is the matrix trace function. Our reasons for writing $\|\boldsymbol{A}\|_F$ rather than $\|\boldsymbol{A}\|$ will soon be apparent. In the meanwhile, the Frobenius matrix norm satisfies the additional condition

(e) $\|\boldsymbol{A}\boldsymbol{B}\| \leq \|\boldsymbol{A}\| \cdot \|\boldsymbol{B}\|$

for any two compatible matrices $\boldsymbol{A} = (a_{ij})$ and $\boldsymbol{B} = (b_{ij})$. Property (e) is verified by invoking the Cauchy-Schwarz inequality in

$$\begin{aligned} \|\boldsymbol{A}\boldsymbol{B}\|_F^2 &= \sum_{i,j} \Big| \sum_k a_{ik} b_{kj} \Big|^2 \\ &\leq \sum_{i,j} \Big(\sum_k a_{ik}^2\Big)\Big(\sum_l b_{lj}^2\Big) \\ &= \Big(\sum_{i,k} a_{ik}^2\Big)\Big(\sum_{l,j} b_{lj}^2\Big) \qquad (2.1) \\ &= \|\boldsymbol{A}\|_F^2 \|\boldsymbol{B}\|_F^2 . \end{aligned}$$

The Frobenius norm does not satisfy the natural condition

(f) $\|\boldsymbol{I}\| = 1$

for an identity matrix $\boldsymbol{I}$. Indeed, an easy calculation shows that $\|\boldsymbol{I}\|_F = \sqrt{n}$ when $\boldsymbol{I}$ is $n \times n$.

To meet all of the conditions (a) through (f), we need to turn to induced matrix norms. Let $\|\cdot\|$ denote both the Euclidean norm on R^m and the Euclidean norm on R^n. The induced Euclidean norm on $m \times n$ matrices is defined by

$$\|\boldsymbol{A}\| \quad = \quad \sup_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\|\boldsymbol{A}\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \quad = \quad \sup_{\|\boldsymbol{x}\|=1} \|\boldsymbol{A}\boldsymbol{x}\|. \tag{2.2}$$

For reasons explained in Proposition 2.2.1, the induced norm (2.2) is called the spectral norm. The question of whether the indicated supremum exists definition (2.2) is settled by the inequalities

$$\|\boldsymbol{A}\boldsymbol{x}\| \quad \leq \quad \sum_{i=1}^{n} |x_i| \cdot \|\boldsymbol{A}\boldsymbol{e}_i\| \quad \leq \quad \left(\sum_{i=1}^{n} \|\boldsymbol{A}\boldsymbol{e}_i\| \right) \|\boldsymbol{x}\|,$$

where $\boldsymbol{x} = \sum_{i=1}^{n} x_i \boldsymbol{e}_i$ and $\boldsymbol{e}_i$ is the unit vector whose entries are all 0 except for $e_{ii} = 1$. More exotic induced matrix norms can be concocted by substituting non-Euclidean norms in the numerator and denominator of definition (2.2). For square matrices, the two norms ordinarily coincide. All of the defining properties of a matrix norm are trivial to check for an induced matrix norm. For instance, property (e) follows from

$$\begin{aligned} \|\boldsymbol{A}\boldsymbol{B}\| \quad &= \quad \sup_{\|\boldsymbol{x}\|=1} \|\boldsymbol{A}\boldsymbol{B}\boldsymbol{x}\| \\ &\leq \quad \|\boldsymbol{A}\| \sup_{\|\boldsymbol{x}\|=1} \|\boldsymbol{B}\boldsymbol{x}\| \\ &= \quad \|\boldsymbol{A}\| \cdot \|\boldsymbol{B}\|. \end{aligned}$$

Definition (2.2) also clearly entails the equality $\|\boldsymbol{I}\| = 1$ when $m = n$.

The next proposition determines the value of the Euclidean norm $\|\boldsymbol{A}\|$. In the proposition, $\rho(\boldsymbol{M})$ denotes the absolute value of the dominant eigenvalue of the square matrix $\boldsymbol{M}$. This quantity is called the spectral radius of $\boldsymbol{M}$.

Proposition 2.2.1 *If $\boldsymbol{A} = (a_{ij})$ is an $m \times n$ matrix, then*

$$\|\boldsymbol{A}\| \quad = \quad \sqrt{\rho(\boldsymbol{A}^*\boldsymbol{A})} \quad = \quad \sqrt{\rho(\boldsymbol{A}\boldsymbol{A}^*)} \quad = \quad \|\boldsymbol{A}^*\|.$$

When $\boldsymbol{A}$ is symmetric, $\|\boldsymbol{A}\|$ reduces to $\rho(\boldsymbol{A})$. The norms $\|\boldsymbol{A}\|$ and $\|\boldsymbol{A}\|_F$ satisfy

$$\|\boldsymbol{A}\| \quad \leq \quad \|\boldsymbol{A}\|_F \quad \leq \quad \sqrt{n}\|\boldsymbol{A}\|. \tag{2.3}$$

Finally, when $\boldsymbol{A}$ is a row or column vector, the Euclidean matrix and vector norms of $\boldsymbol{A}$ coincide.

Proof: Choose an orthonormal basis of eigenvectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ for the symmetric matrix $\boldsymbol{A}^*\boldsymbol{A}$ with corresponding eigenvalues arranged so that $0 \le \lambda_1 \le \cdots \le \lambda_n$. If $\boldsymbol{x} = \sum_{i=1}^n c_i \boldsymbol{u}_i$ is a unit vector, then $\sum_{i=1}^n c_i^2 = 1$, and

$$\begin{aligned} \|\boldsymbol{A}\|^2 &= \sup_{\|\boldsymbol{x}\|=1} \boldsymbol{x}^*\boldsymbol{A}^*\boldsymbol{A}\boldsymbol{x} \\ &= \sup_{\|\boldsymbol{x}\|=1} \sum_{i=1}^n \lambda_i c_i^2 \\ &\le \lambda_n. \end{aligned}$$

Equality is achieved when $c_n = \pm 1$ and all other $c_i = 0$. If $\boldsymbol{A}$ is symmetric with eigenvalues μ_i arranged so that $|\mu_1| \le \cdots \le |\mu_n|$, then the $\boldsymbol{u}_i$ can be chosen to be the corresponding eigenvectors. In this case, clearly $\lambda_i = \mu_i^2$.

To prove that $\rho(\boldsymbol{A}^*\boldsymbol{A}) = \rho(\boldsymbol{A}\boldsymbol{A}^*)$, choose an eigenvalue $\lambda \ne 0$ of $\boldsymbol{A}^*\boldsymbol{A}$ with corresponding eigenvector $\boldsymbol{v}$. Multiplying the equation $\boldsymbol{A}^*\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v}$ on the left by $\boldsymbol{A}$ produces $(\boldsymbol{A}\boldsymbol{A}^*)\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{A}\boldsymbol{v}$. Because $\boldsymbol{A}^*\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v}$, the vector $\boldsymbol{A}\boldsymbol{v} \ne \boldsymbol{0}$. Thus, λ is an eigenvalue of $\boldsymbol{A}\boldsymbol{A}^*$ with eigenvector $\boldsymbol{A}\boldsymbol{v}$. Likewise, any eigenvalue $\omega \ne 0$ of $\boldsymbol{A}\boldsymbol{A}^*$ is also an eigenvalue of $\boldsymbol{A}^*\boldsymbol{A}$.

To verify the left bound of the pair of bounds (2.3), apply inequality (2.1) with $\boldsymbol{B} = \boldsymbol{x}$ in the definition of $\|\boldsymbol{A}\|$. The right bound follows from

$$\textstyle\sum_{i=1}^m a_{ij}^2 \;=\; \|\boldsymbol{A}\boldsymbol{e}_j\|^2 \;\le\; \|\boldsymbol{A}\|^2$$

by summing on j. Finally, suppose that $\boldsymbol{A}$ is a column vector. The two bounds (2.3) with $n = 1$ show that $\|\boldsymbol{A}\| = \|\boldsymbol{A}\|_F$. If $\boldsymbol{A}$ is a row vector, the same reasoning applied to $\boldsymbol{A}^*$ gives $\|\boldsymbol{A}\| = \|\boldsymbol{A}^*\| = \|\boldsymbol{A}^*\|_F = \|\boldsymbol{A}\|_F$. ■

2.3 Convergence and Completeness

A sequence $\boldsymbol{x}_m \in \mathsf{R}^n$ converges to $\boldsymbol{x}$, written $\lim_{m\to\infty} \boldsymbol{x}_m = \boldsymbol{x}$, provided $\lim_{m\to\infty} \|\boldsymbol{x}_m - \boldsymbol{x}\| = 0$ in the standard Euclidean norm. For convergence of $\boldsymbol{x}_m$ to $\boldsymbol{x}$ to occur, it is necessary and sufficient that each component sequence x_{mi} converge to x_i. Convergence of a sequence of matrices is defined similarly using either the Frobenius norm $\|\boldsymbol{A}\|_F$ or the induced matrix norm $\|\boldsymbol{A}\|$. The pair of bounds (2.3) shows that the two norms are equivalent in testing convergence.

Convergent sequences of vectors or matrices enjoy many useful properties. Some of these are mentioned in the next proposition.

Proposition 2.3.1 *In the following list, once a limit is assumed to exist for an item, it is assumed to exist for all subsequent items. With this proviso, we have:*

(a) *If* $\lim_{m\to\infty} \boldsymbol{x}_m = \boldsymbol{x}$*, then* $\lim_{m\to\infty} \|\boldsymbol{x}_m\| = \|\boldsymbol{x}\|$.

(b) *If* $\lim_{m\to\infty} \boldsymbol{y}_m = \boldsymbol{y}$, *then*

$$\lim_{m\to\infty} \boldsymbol{x}_m^* \boldsymbol{y}_m = \boldsymbol{x}^* \boldsymbol{y}.$$

(c) *If* a *and* b *are real scalars, then*

$$\lim_{m\to\infty} [a\boldsymbol{x}_m + b\boldsymbol{y}_m] = a\boldsymbol{x} + b\boldsymbol{y}.$$

(d) *If* $\lim_{m\to\infty} \boldsymbol{M}_m = \boldsymbol{M}$ *for a sequence of matrices compatible with* $\boldsymbol{x}$, *then*

$$\lim_{m\to\infty} \boldsymbol{M}_m \boldsymbol{x}_m = \boldsymbol{M}\boldsymbol{x}.$$

(e) *If* $\boldsymbol{M}$ *is square and invertible, then* $\boldsymbol{M}_m^{-1}$ *exists for large* m *and*

$$\lim_{m\to\infty} \boldsymbol{M}_m^{-1} = \boldsymbol{M}^{-1}.$$

(f) *Finally, if* $\lim_{m\to\infty} \boldsymbol{N}_m = \boldsymbol{N}$ *for a sequence of matrices compatible with* $\boldsymbol{M}$, *then*

$$\lim_{m\to\infty} \boldsymbol{M}_m \boldsymbol{N}_m = \boldsymbol{M}\boldsymbol{N}.$$

Proof: As a sample proof, part (d) follows from the inequalities

$$\begin{aligned} \|\boldsymbol{M}_m \boldsymbol{x}_m - \boldsymbol{M}\boldsymbol{x}\| &\leq \|\boldsymbol{M}_m \boldsymbol{x}_m - \boldsymbol{M}_m \boldsymbol{x}\| + \|\boldsymbol{M}_m \boldsymbol{x} - \boldsymbol{M}\boldsymbol{x}\| \\ &\leq \|\boldsymbol{M}_m\| \cdot \|\boldsymbol{x}_m - \boldsymbol{x}\| + \|\boldsymbol{M}_m - \boldsymbol{M}\| \cdot \|\boldsymbol{x}\| \\ \|\boldsymbol{M}_m\| &\leq \|\boldsymbol{M}_m - \boldsymbol{M}\| + \|\boldsymbol{M}\|. \end{aligned}$$

Part (e) will be proved after Example 2.3.3. ■

In some situations, we know that the members of a sequence become progressively closer together. A Cauchy sequence $\boldsymbol{x}_m$ exhibits a strong form of this phenomenon; namely, for every $\epsilon > 0$, there is an m such that $\|\boldsymbol{x}_p - \boldsymbol{x}_q\| \leq \epsilon$ for all $p, q \geq m$. The real line R is complete in the sense that every Cauchy sequence possesses a limit. The rational numbers are incomplete by contrast because a sequence of rationals can converge to an irrational. The completeness of R carries over to R^n. Indeed, if $\boldsymbol{x}_m$ is a Cauchy sequence, then under the Euclidean norm we have

$$|x_{pi} - x_{qi}| \leq \|\boldsymbol{x}_p - \boldsymbol{x}_q\|.$$

This shows that each component sequence is Cauchy and consequently possesses a limit x_i. The vector $\boldsymbol{x}$ with components x_i then furnishes a limit for the vector sequence $\boldsymbol{x}_m$.

Example 2.3.1 *Existence of Suprema and Infima*

The completeness of the real line is equivalent to the existence of least upper bounds or suprema. Consider a nonempty set $S \subset \mathsf{R}$ that is bounded above. If the set is finite, then its least upper bound is just its largest element. If the set is infinite, we choose a and b such that the interval $[a, b]$ contains an element of S and b is an upper bound of S. We can generate $\sup S$ by a bisection strategy. Bisect $[a, b]$ into the two subintervals $[a, (a + b)/2]$ and $[(a + b)/2, b]$. Let $[a_1, b_1]$ denote the left subinterval if $(a + b)/2$ provides an upper bound. Otherwise, let $[a_1, b_1]$ denote the right subinterval. In either case, $[a_1, b_1]$ contains an element of S. Now bisect $[a_1, b_1]$ and generate a subinterval $[a_2, b_2]$ by the same criterion. If we continue bisecting and choosing a left or right subinterval ad infinitum, then we generate two Cauchy sequences a_i and b_i with common limit c. By the definition of the sequence b_i, c furnishes an upper bound of S. By the definition of the sequence a_i, no bound of S is smaller than c. Establishing the existence of the greatest lower bound $\inf S$ for S bounded below proceeds similarly. If S is unbounded above, then $\sup S = \infty$, and if it is unbounded below, then $\inf S = -\infty$. ■

Example 2.3.2 *Limit Superior and Limit Inferior*

For a real sequence x_n. we define the limit superior and limit inferior by

$$\begin{aligned} \limsup_{n\to\infty} x_n &= \inf_m \sup_{n\geq m} x_n = \lim_{m\to\infty} \sup_{n\geq m} x_n \\ \liminf_{n\to\infty} x_n &= \sup_m \inf_{n\geq m} x_n = \lim_{m\to\infty} \inf_{n\geq m} x_n. \end{aligned}$$

If $\sup_n x_n = \infty$, then $\limsup_{n\to\infty} x_n = \infty$, and if $\lim_{n\to\infty} x_n = -\infty$, then $\limsup_{n\to\infty} x_n = -\infty$. From these definitions, one can also deduce that

$$\limsup_{n\to\infty} -x_n = -\liminf_{n\to\infty} x_n \tag{2.4}$$

and that

$$\liminf_{n\to\infty} x_n \leq \limsup_{n\to\infty} x_n. \tag{2.5}$$

The sequence x_n has a limit if and only if equality prevails in inequality (2.5). In this situation, the common value of the limit superior and inferior furnishes the limit of x_n. ■

Example 2.3.3 *Series Expansion for a Matrix Inverse*

If a square matrix $\boldsymbol{M}$ has norm $\|\boldsymbol{M}\| < 1$, then we can write

$$(\boldsymbol{I} - \boldsymbol{M})^{-1} = \sum_{i=0}^{\infty} \boldsymbol{M}^i.$$

To verify this claim, we first prove that the partial sums $\boldsymbol{S}_j = \sum_{i=0}^{j} \boldsymbol{M}^i$ form a Cauchy sequence. This fact is a consequence of the inequalities

$$\begin{aligned} \|\boldsymbol{S}_k - \boldsymbol{S}_j\| &= \Big\| \sum_{i=j+1}^{k} \boldsymbol{M}^i \Big\| \\ &\leq \sum_{i=j+1}^{k} \|\boldsymbol{M}^i\| \\ &\leq \sum_{i=j+1}^{k} \|\boldsymbol{M}\|^i \end{aligned}$$

for $k \geq j$ and the assumption $\|\boldsymbol{M}\| < 1$. If we let $\boldsymbol{S}$ represent the limit of the $\boldsymbol{S}_j$, then part (f) of Proposition 2.3.1 implies that $(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{S}_j$ converges to $(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{S}$. But $(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{S}_j = \boldsymbol{I} - \boldsymbol{M}^{j+1}$ also converges to $\boldsymbol{I}$. Hence, $(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{S} = \boldsymbol{I}$, and this verifies the claim $\boldsymbol{S} = (\boldsymbol{I}-\boldsymbol{M})^{-1}$. ■

With this result under our belts, we now demonstrate part (e) of Proposition 2.3.1. Because $\|\boldsymbol{M}^{-1}(\boldsymbol{M}-\boldsymbol{M}_m)\| \leq \|\boldsymbol{M}^{-1}\| \cdot \|\boldsymbol{M}-\boldsymbol{M}_m\|$, the matrix inverse $[\boldsymbol{I}-\boldsymbol{M}^{-1}(\boldsymbol{M}-\boldsymbol{M}_m)]^{-1}$ exists for large m. Therefore, we can write the inverse of

$$\begin{aligned} \boldsymbol{M}_m &= \boldsymbol{M} - (\boldsymbol{M} - \boldsymbol{M}_m) \\ &= \boldsymbol{M}[\boldsymbol{I} - \boldsymbol{M}^{-1}(\boldsymbol{M} - \boldsymbol{M}_m)] \end{aligned}$$

as

$$\boldsymbol{M}_m^{-1} = [\boldsymbol{I} - \boldsymbol{M}^{-1}(\boldsymbol{M} - \boldsymbol{M}_m)]^{-1}\boldsymbol{M}^{-1}.$$

The proof of convergence is completed by noting the bound

$$\begin{aligned} \|\boldsymbol{M}_m^{-1} - \boldsymbol{M}^{-1}\| &= \Big\| \sum_{i=1}^{\infty} [\boldsymbol{M}^{-1}(\boldsymbol{M} - \boldsymbol{M}_m)]^i \boldsymbol{M}^{-1} \Big\| \\ &\leq \sum_{i=1}^{\infty} \|\boldsymbol{M}^{-1}\|^i \|\boldsymbol{M} - \boldsymbol{M}_m\|^i \|\boldsymbol{M}^{-1}\| \\ &= \frac{\|\boldsymbol{M}^{-1}\|^2 \|\boldsymbol{M} - \boldsymbol{M}_m\|}{1 - \|\boldsymbol{M}^{-1}\| \cdot \|\boldsymbol{M} - \boldsymbol{M}_m\|}, \end{aligned}$$

applying in the process the matrix analog of part (a) of Proposition 2.3.1.

Example 2.3.4 *Matrix Exponential Function*

The exponential of a square matrix $\boldsymbol{M}$ is given by the series expansion

$$e^{\boldsymbol{M}} = \sum_{i=0}^{\infty} \frac{1}{i!} \boldsymbol{M}^i.$$

To prove the convergence of the series, it again suffices to show that the partial sums $\boldsymbol{S}_j = \sum_{i=0}^{j} \frac{1}{i!}\boldsymbol{M}^i$ form a Cauchy sequence. The bound

$$\|\boldsymbol{S}_k - \boldsymbol{S}_j\| \quad = \quad \Big\| \sum_{i=j+1}^{k} \frac{1}{i!}\boldsymbol{M}^i \Big\| \quad \leq \quad \sum_{i=j+1}^{k} \frac{1}{i!}\|\boldsymbol{M}\|^i$$

for $k \geq j$ is just what we need.

The matrix exponential function has many interesting properties. For example, the function $N(t) = e^{t\boldsymbol{M}}$ solves the differential equation

$$N'(t) \quad = \quad \boldsymbol{M}N(t)$$

subject to the initial condition $N(0) = \boldsymbol{I}$. Here t is a real parameter, and we differentiate the matrix $N(t)$ entry by entry. In Example 4.2.2 of Chap. 4, we will prove that $N(t) = e^{t\boldsymbol{M}}$ is the one and only solution. The law of exponents $e^{\boldsymbol{A}+\boldsymbol{B}} = e^{\boldsymbol{A}}e^{\boldsymbol{B}}$ for commuting matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ is another interesting property of the matrix exponential function. One way of proving the law of exponents is to observe that $e^{t(\boldsymbol{A}+\boldsymbol{B})}$ and $e^{t\boldsymbol{A}}e^{t\boldsymbol{B}}$ both solve the differential equation

$$N'(t) \quad = \quad (\boldsymbol{A}+\boldsymbol{B})N(t)$$

subject to the initial condition $N(0) = \boldsymbol{I}$. Since the solution to such an initial value problem is unique, the two solutions must coincide at $t = 1$. ■

2.4 The Topology of R^n

Mathematics involves a constant interplay between the abstract and the concrete. We now consider some qualitative features of sets in R^n that generalize to more abstract spaces. For instance, there is the matter of boundedness. A set $S \subset \mathsf{R}^n$ is said to be bounded if it is contained in some ball $B(\mathbf{0}, r) = \{\boldsymbol{x} \in \mathsf{R}^n : \|\boldsymbol{x}\| < r\}$ of radius r centered at the origin $\mathbf{0}$. As we shall see in our discussion of compactness, boundedness takes on added importance when it is combined with the notion of closedness. A closed set is closed under the formation of limits. Thus, $S \subset \mathsf{R}^n$ is closed if for every convergent sequence $\boldsymbol{x}_m$ taken from S, we have $\lim_{m\to\infty} \boldsymbol{x}_m \in S$ as well.

It takes time and effort to appreciate the ramifications of these ideas. A few of the most pertinent ones for closedness are noted in the next proposition.

Proposition 2.4.1 *The collection of closed sets satisfy the following:*

(a) *The whole space* R^n *is closed.*

(b) *The empty set* $\emptyset$ *is closed.*

(c) *The intersection $S = \cap_\alpha S_\alpha$ of an arbitrary number of closed sets S_α is closed.*

(d) *The union $S = \cup_\alpha S_\alpha$ of a finite number of closed sets S_α is closed.*

Proof: All of these are easy. For part (d), observe that for any convergent sequence $\boldsymbol{x}_m$ taken from S, one of the sets S_α must contain an infinite subsequence $\boldsymbol{x}_{m_k}$. The limit of this subsequence exists and falls in S_α. ■

Some examples of closed sets are closed intervals $(-\infty, a]$, $[a, b]$, and $[b, \infty)$; closed balls $\{\boldsymbol{y} \in \mathsf{R}^n : \|\boldsymbol{y} - \boldsymbol{x}\| \leq r\}$; spheres $\{\boldsymbol{y} \in \mathsf{R}^n : \|\boldsymbol{y} - \boldsymbol{x}\| = r\}$; hyperplanes $\{\boldsymbol{x} \in \mathsf{R}^n : \boldsymbol{z}^*\boldsymbol{x} = c\}$; and closed halfspaces $\{\boldsymbol{x} \in \mathsf{R}^n : \boldsymbol{z}^*\boldsymbol{x} \leq c\}$. A closed set S of R^n is complete in the sense that all Cauchy sequences from S possess limits in S.

Example 2.4.1 *Finitely Generated Convex Cones*

A set C is a convex cone provided $\alpha\boldsymbol{u} + \beta\boldsymbol{v}$ is in C whenever the vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are in C and the scalars α and β are nonnegative. A finitely generated convex cone can be written as

$$C = \Big\{\sum_{i=1}^m \alpha_i \boldsymbol{v}_i : \alpha_i \geq 0, \; i = 1, \ldots, m\Big\}.$$

Demonstrating that C is a closed set is rather subtle. Consider a sequence $\boldsymbol{u}_j = \sum_{i=1}^m \alpha_{ji}\boldsymbol{v}_i$ in C converging to a point $\boldsymbol{u}$. If the vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$ are linearly independent, then the coefficients α_{ji} are the unique coordinates of $\boldsymbol{u}_j$ in the finite-dimensional subspace spanned by the $\boldsymbol{v}_i$. To recover the α_{ji}, we introduce the matrix $\boldsymbol{V}$ with columns $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$ and rewrite the original equation as $\boldsymbol{u}_j = \boldsymbol{V}\boldsymbol{\alpha}_j$. Multiplying this equation by first $\boldsymbol{V}^*$ and then by $(\boldsymbol{V}^*\boldsymbol{V})^{-1}$ on the left gives $\boldsymbol{\alpha}_j = (\boldsymbol{V}^*\boldsymbol{V})^{-1}\boldsymbol{V}^*\boldsymbol{u}_j$. This representations allows us to conclude that $\boldsymbol{\alpha}_j$ possesses a limit $\boldsymbol{\alpha}$ with nonnegative entries. Therefore, the limit $\boldsymbol{u} = \boldsymbol{V}\boldsymbol{\alpha}$ lies in C.

If we relax the assumption that the vectors are linearly independent, we must resort to an inductive argument to prove that C is closed. The case $m = 1$ is true because a single vector $\boldsymbol{v}_1$ is linearly independent. Assume that the claim holds for $m - 1$ vectors. If the vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$ are linearly independent, then we are done. If the vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$ are linearly dependent, then there exist scalars $\beta_1, \ldots, \beta_m$, not all 0, such that $\sum_{i=1}^m \beta_i\boldsymbol{v}_i = \boldsymbol{0}$. Without loss of generality, we can assume that $\beta_i < 0$ for at least one index i. We can express any point $\boldsymbol{u} \in C$ as

$$\boldsymbol{u} = \sum_{i=1}^m \alpha_i\boldsymbol{v}_i = \sum_{i=1}^m (\alpha_i + t\beta_i)\boldsymbol{v}_i$$

for an arbitrary scalar t. If we increase t gradually from 0, then there is a first value at which $\alpha_j + t\beta_j = 0$ for some index j. This shows that C can be decomposed as the union

$$C = \bigcup_{j=1}^{m} \Big\{ \sum_{i \neq j} \gamma_i \boldsymbol{v}_i : \gamma_i \geq 0,\ i \neq j \Big\}.$$

Each of the convex cones $\{\sum_{i \neq j} \gamma_i \boldsymbol{v}_i : \gamma_i \geq 0,\ i \neq j\}$ is closed by the induction hypothesis. Since a finite union of closed sets is closed, C itself is closed. A straightforward extension of this argument establishes the stronger claim that every point in the cone can be represented as a positive combination of a linearly independent subset of $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$. ∎

The complement $S^c = \mathsf{R}^n \setminus S$ of a closed set S is called an open set. Every $\boldsymbol{x} \in S^c$ is surrounded by a ball $B(\boldsymbol{x}, r)$ completely contained in S^c. If this were not the case, then we could construct a sequence of points $\boldsymbol{x}_m$ from S converging to $\boldsymbol{x}$, contradicting the closedness of S. This fact is the first of several mentioned in the next proposition.

Proposition 2.4.2 *The collection of open sets satisfy the following:*

(a) *Every open set is a union of balls, and every union of balls is an open set.*

(b) *The whole space* R^n *is open.*

(c) *The empty set* $\emptyset$ *is open.*

(d) *The union* $S = \cup_\alpha S_\alpha$ *of an arbitrary number of open sets* S_α *is open.*

(e) *The intersection* $S = \cap_\alpha S_\alpha$ *of a finite number of open sets* S_α *is open.*

Proof: Again these are easy. Parts (d) and (e) are consequences of the set identities

$$\begin{aligned} (\cap_\alpha S_\alpha)^c &= \cup_\alpha S_\alpha^c \\ (\cup_\alpha S_\alpha)^c &= \cap_\alpha S_\alpha^c \end{aligned}$$

and parts (c) and (d) of Proposition 2.4.1. ∎

Some examples of open sets are open intervals $(-\infty, a)$, (a, b), and (b, ∞); balls $\{\boldsymbol{y} \in \mathsf{R}^n : \|\boldsymbol{y} - \boldsymbol{x}\| < r\}$, and open halfspaces $\{\boldsymbol{x} \in \mathsf{R}^n : \boldsymbol{z}^* \boldsymbol{x} < c\}$. Any open set surrounding a point is called a neighborhood of the point. Some examples of sets that are neither closed nor open are the unbalanced intervals $(a, b]$ and $[a, b)$, the discrete set $V = \{n^{-1} : n = 1, 2, \ldots\}$, and the rational numbers. If we append the limit 0 to the set V, then it becomes closed.

A boundary point $\boldsymbol{x}$ of a set S is the limit of a sequence of points from S and also the limit of a different sequence of points from S^c. Closed sets contain all of their boundary points, and open sets contain none of their boundary points. The interior of S is the largest open set contained within S. The closure of S is the smallest closed set containing S. For instance, the boundary of the ball $B(\boldsymbol{x}, r) = \{\boldsymbol{y} \in \mathsf{R}^n : \|\boldsymbol{y} - \boldsymbol{x}\| < r\}$ is the sphere $S(\boldsymbol{x}, r) = \{\boldsymbol{y} \in \mathsf{R}^n : \|\boldsymbol{y} - \boldsymbol{x}\| = r\}$. The closure of $B(\boldsymbol{x}, r)$ is the closed ball $C(\boldsymbol{x}, r) = \{\boldsymbol{y} \in \mathsf{R}^n : \|\boldsymbol{y} - \boldsymbol{x}\| \leq r\}$, and the interior of $C(\boldsymbol{x}, r)$ is $B(\boldsymbol{x}, r)$.

A closed bounded set is said to be compact. Finite intervals $[a, b]$ are typical compact sets. Compact sets can be defined in several equivalent ways. The most important of these is the Bolzano-Weierstrass characterization. In preparation for this result, let us define a multidimensional interval $[\boldsymbol{a}, \boldsymbol{b}]$ in R^n to be the Cartesian product

$$[\boldsymbol{a}, \boldsymbol{b}] \quad = \quad [a_1, b_1] \times \cdots \times [a_n, b_n]$$

of n one-dimensional intervals. We will only consider closed intervals. The diameter of $[\boldsymbol{a}, \boldsymbol{b}]$ is the greatest separation between any two of its points; this clearly reduces to the distance $\|\boldsymbol{a} - \boldsymbol{b}\|$ between its extreme corners.

Proposition 2.4.3 (Bolzano-Weierstrass) *A set $S \subset \mathsf{R}^n$ is compact if and only if every sequence $\boldsymbol{x}_m$ in S has a convergent subsequence $\boldsymbol{x}_{m_i}$ with limit in S.*

Proof: Suppose every sequence $\boldsymbol{x}_m$ in S has a convergent subsequence $\boldsymbol{x}_{m_i}$ with limit in S. If S is unbounded, then we can define a sequence $\boldsymbol{x}_m$ with $\|\boldsymbol{x}_m\| \geq m$. Clearly, this sequence has no convergent subsequence. If S is not closed, then there is a convergent sequence $\boldsymbol{x}_m$ with limit $\boldsymbol{x}$ outside S. Clearly, no subsequence of $\boldsymbol{x}_m$ can converge to a limit in S. Thus, the subsequence property implies compactness.

For the converse, let $\boldsymbol{x}_m$ be a sequence in the compact set S. Because S is bounded, it is contained in a multidimensional interval $[\boldsymbol{a}, \boldsymbol{b}]$. If infinitely many of the $\boldsymbol{x}_m$ coincide, then these can be used to construct a constant subsequence that trivially converges to a point of S. Otherwise, let T_0 denote the infinite set $\cup_{m=1}^{\infty}\{\boldsymbol{x}_m\}$.

The rest of the proof adapts the bisection strategy of Example 2.3.1. The first stage of the bisection divides $[\boldsymbol{a}, \boldsymbol{b}]$ into 2^n subintervals of equal volume. Each of these subintervals can be written as $[\boldsymbol{a}_1, \boldsymbol{b}_1]$, where $a_{1j} = a_j$ and $b_{1j} = (a_j + b_j)/2$ or $a_{1j} = (a_j + b_j)/2$ and $b_{1j} = b_j$. There is no harm in the fact that these subintervals overlap along their boundaries. It is only vital to observe that one of the subintervals contains an infinite subset $T_1 \subset T_0$. Let us choose such a subinterval and label it using the generic notation $[\boldsymbol{a}_1, \boldsymbol{b}_1]$. We now inductively repeat the process. At stage $i + 1$ we divide the previously chosen subinterval $[\boldsymbol{a}_i, \boldsymbol{b}_i]$ into 2^n subintervals of equal volume. Each of these subintervals can be written as $[\boldsymbol{a}_{i+1}, \boldsymbol{b}_{i+1}]$, where either $a_{i+1,j} = a_{ij}$ and $b_{i+1,j} = (a_{ij} + b_{ij})/2$ or $a_{i+1,j} = (a_{ij} + b_{ij})/2$ and $b_{i+1,j} = b_{ij}$. One of these subintervals, which we label $[\boldsymbol{a}_{i+1}, \boldsymbol{b}_{i+1}]$ for convenience, contains an infinite subset $T_{i+1} \subset T_i$.

We continue this process ad infinitum. in the process choosing $\boldsymbol{x}_{m_i}$ from T_i and $m_i > m_{i-1}$. Because $T_i \subset [\boldsymbol{a}_i, \boldsymbol{b}_i]$ and the diameter of $[\boldsymbol{a}_i, \boldsymbol{b}_i]$ tends to 0, the subsequence $\boldsymbol{x}_{m_i}$ is Cauchy. By virtue of the completeness of R^n, this subsequence converges to some point $\boldsymbol{x}$, which necessarily belongs to the closed set S. ■

In many instances it is natural to consider a subset S of R^n as a topological space in its own right. Notions of distance and convergence carry over immediately, but we must exercise some care in defining closed and open sets. In the relative topology, a subset $T \subset S$ is closed if and only if it can be represented as the intersection $T = S \cap C$ of S with a closed set C of R^n. If T is closed in S, then the obvious choice of C is the closure of T in R^n. Likewise, $T \subset S$ is open in the relative topology if and only if it can be represented as the intersection $T = S \cap O$ of S with an open set O of R^n. These two definitions are consistent with an open set being the relative complement of a closed set and vice versa. They are also consistent with the development of continuous functions sketched in the next section.

2.5 Continuous Functions

Continuous functions are the building blocks of mathematical analysis. Continuity is such an intuitive notion that ancient mathematicians did not even bother to define it. Proper recognition of continuity had to wait until differentiability was thoroughly explored. Our approach to continuity emphasizes convergent sequences. A function $f(\boldsymbol{x})$ from R^m to R^n is said to be continuous at $\boldsymbol{y}$ if $f(\boldsymbol{x}_i)$ converges to $f(\boldsymbol{y})$ for every sequence $\boldsymbol{x}_i$ that converges to $\boldsymbol{y}$. If the domain of $f(\boldsymbol{x})$ is a subset S of R^m, then the sequences $\boldsymbol{x}_i$ and the point $\boldsymbol{y}$ are confined to S. Finally, $f(\boldsymbol{x})$ is said to be continuous if it is continuous at every point $\boldsymbol{y}$ of its domain.

The definition of continuity through convergent sequences tends to be simpler to apply than the competing ϵ and δ approach of calculus. We leave it to the reader to show that the two definitions are fully equivalent. Either definition has powerful consequences. For instance, it is clear that a vector-valued function is continuous if and only if each of its component functions is continuous. Before enumerating other less obvious consequences, it is helpful to forge a few tools for recognizing and constructing continuous functions. Fortunately, the collection of continuous functions is closed under many standard algebraic operations. Here are a few examples.

Proposition 2.5.1 *Given that the vector-valued functions $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$ and matrix-valued function $M(\boldsymbol{x})$ and $N(\boldsymbol{x})$ are continuous and compatible whenever necessary, the following algebraic combinations are continuous:*

(a) *The norm $\|f(\boldsymbol{x})\|$.*

(b) *The inner product $f(\boldsymbol{x})^* g(\boldsymbol{x})$.*

(c) *The linear combination $\alpha f(\boldsymbol{x}) + \beta g(\boldsymbol{x})$ for real scalars α and β.*

(d) *The matrix-vector product* $M(\boldsymbol{x})f(\boldsymbol{x})$.

(e) *The matrix inverse* $M^{-1}(\boldsymbol{x})$ *when* $M(\boldsymbol{x})$ *is square and invertible.*

(f) *The matrix product* $M(\boldsymbol{x})N(\boldsymbol{x})$.

(g) *The functional composition* $f \circ g(\boldsymbol{x}) = f[g(\boldsymbol{x})]$.

Proof: Parts (a) through (f) are all immediate by-products of Proposition 2.3.1 and the definition of continuity. For part (g), suppose $\boldsymbol{x}_i$ tends to $\boldsymbol{x}$. Then $f(\boldsymbol{x}_i)$ tends to $f(\boldsymbol{x})$, and so $f \circ g(\boldsymbol{x}_i)$ tends to $f \circ g(\boldsymbol{x})$. ■

Example 2.5.1 *Rational Functions*

Because the coordinate variables x_i of $\boldsymbol{x} \in \mathsf{R}^n$ are continuous, all polynomials in these variables are continuous as well. For example, the determinant of a square matrix is a continuous function of the entries of the matrix. A quotient of two polynomials (rational function) in the coordinate variables x_i of $\boldsymbol{x} \in \mathsf{R}^n$ is continuous where its denominator does not vanish. Finally, any linear transformation of one vector space into another is continuous. ■

Example 2.5.2 *Distance to a Set*

The distance $\text{dist}(\boldsymbol{x}, S)$ from a point $\boldsymbol{x} \in \mathsf{R}^n$ to a set S is defined by

$$\text{dist}(\boldsymbol{x}, S) \quad = \quad \inf_{\boldsymbol{z} \in S} \|\boldsymbol{z} - \boldsymbol{x}\|.$$

To prove that the function $\text{dist}(\boldsymbol{x}, S)$ is continuous in $\boldsymbol{x}$, take the infimum over $\boldsymbol{z} \in S$ of both sides of the triangle inequality

$$\|\boldsymbol{z} - \boldsymbol{x}\| \quad \leq \|\boldsymbol{z} - \boldsymbol{y}\| + \|\boldsymbol{y} - \boldsymbol{x}\|.$$

This demonstrates that $\text{dist}(\boldsymbol{x}, S) \leq \text{dist}(\boldsymbol{y}, S) + \|\boldsymbol{y} - \boldsymbol{x}\|$. Reversing the roles of $\boldsymbol{x}$ and $\boldsymbol{y}$ then leads to the inequality

$$|\,\text{dist}(\boldsymbol{x}, S) - \text{dist}(\boldsymbol{y}, S)| \quad \leq \quad \|\boldsymbol{y} - \boldsymbol{x}\|,$$

establishing continuity. ■

In generalizing continuity to more abstract topological spaces, the characterizations in the next proposition are crucial.

Proposition 2.5.2 *The following conditions are equivalent for a function* $f(\boldsymbol{x})$ *from* $T \subset \mathsf{R}^m$ *to* R^n*:*

(a) $f(\boldsymbol{x})$ *is continuous.*

(b) *The inverse image* $f^{-1}(S)$ *of every closed set* S *is closed.*

(c) *The inverse image* $f^{-1}(S)$ *of every open set* S *is open.*

Proof: To prove that (a) implies (b), suppose $\boldsymbol{x}_i$ is a sequence in $f^{-1}(S)$ tending to $\boldsymbol{x} \in T$. Then the conclusion $\lim_{i\to\infty} f(\boldsymbol{x}_i) = f(\boldsymbol{x})$ identifies $f(\boldsymbol{x})$ as an element of the closed set S and therefore $\boldsymbol{x}$ as belonging to $f^{-1}(S)$. Conditions (b) and (c) are equivalent because of the relation

$$f^{-1}(S)^c \quad = \quad f^{-1}(S^c)$$

between inverse images and set complements. Finally, to prove that (c) entails (a), suppose that $\lim_{i\to\infty} \boldsymbol{x}_i = \boldsymbol{x}$. For any $\epsilon > 0$, the inverse image of the ball $B[f(\boldsymbol{x}), \epsilon]$ is open by assumption. Consequently, there exists a neighborhood $T \cap B(\boldsymbol{x}, \delta)$ mapped into $B[f(\boldsymbol{x}), \epsilon]$. In other words,

$$\|f(\boldsymbol{x}_i) - f(\boldsymbol{x})\| \quad < \quad \epsilon$$

whenever $\|\boldsymbol{x}_i - \boldsymbol{x}\| < \delta$, which is sufficient to validate continuity. ∎

Example 2.5.3 *Continuity of* $\sqrt[m]{x}$

The root function $f(x) = \sqrt[m]{x}$ is the functional inverse of the power function $g(x) = x^m$. We have already noted that $g(x)$ is continuous. On the interval $(0, \infty)$, it is also strictly increasing and maps the open interval (a, b) onto the open interval (a^m, b^m). Put another way, $f^{-1}[(a, b)] = (a^m, b^m)$. (Here we implicitly invoke the intermediate value property proved in Proposition 2.7.1.) Because the inverse image of a union of open intervals is a union of open intervals, application of part (c) of Proposition 2.5.2 establishes the continuity of $f(x)$. ∎

Example 2.5.4 *The Set of Positive Definite Matrices*

A symmetric $n \times n$ matrix $\boldsymbol{M} = (m_{ij})$ can be viewed as a point in R^m for $m = \binom{n}{2} + n$. To demonstrate that the subset S of positive definite matrices is open in R^m, we invoke the classical criterion of Sylvester. (See Problem 29 of Chap. 5 or [136].) This test for positive definiteness uses the determinants of the principal submatrices of $\boldsymbol{M}$. The kth of these submatrices $\boldsymbol{M}_k$ is the $k \times k$ upper left block of $\boldsymbol{M}$. If $\boldsymbol{M}$ is positive definite, then one can show that $\boldsymbol{M}_k$ is positive definite by taking a nontrivial $k \times 1$ vector $\boldsymbol{x}_k$ and padding it with zeros to construct a nontrivial $n \times 1$ vector $\boldsymbol{x}$. It is then clear that $\boldsymbol{x}_k^* \boldsymbol{M}_k \boldsymbol{x}_k = \boldsymbol{x}^* \boldsymbol{M} \boldsymbol{x} > 0$. Because $\boldsymbol{M}_k$ is positive definite, its determinant $\det \boldsymbol{M}_k > 0$. Conversely, if all of the $\det \boldsymbol{M}_k > 0$, then $\boldsymbol{M}$ itself is positive definite.

Given this background, we write

$$S \quad = \quad \bigcap_{k=1}^{n} \{\boldsymbol{M} : \det \boldsymbol{M}_k > 0\}.$$

Because the functions $\det \boldsymbol{M}_k$ are continuous in the entries of $\boldsymbol{M}$, the inverse images $\{\boldsymbol{M} : \det \boldsymbol{M}_k > 0\}$ of the open set $(0, \infty)$ are open. Since a finite intersection of open sets is open, S itself is an open set. ∎

As opposed to inverse images, the image of a closed (open) set under a continuous function need not be closed (open). However, continuous functions do preserve compactness.

Proposition 2.5.3 *Suppose the continuous function $f(\boldsymbol{x})$ maps the compact set $S \subset \mathsf{R}^m$ into R^n. Then the image $f(S)$ is compact.*

Proof: The key is to apply Proposition 2.4.3. Let $f(\boldsymbol{x}_i)$ be a sequence in $f(S)$. Extract a convergent subsequence $\boldsymbol{x}_{i_j}$ of $\boldsymbol{x}_i$ with limit $\boldsymbol{y} \in S$. Then the continuity of $f(\boldsymbol{x})$ compels $f(\boldsymbol{x}_{i_j})$ to converge to $f(\boldsymbol{y})$. ■

We now come to one of the most important results in optimization theory.

Proposition 2.5.4 (Weierstrass) *Let $f(\boldsymbol{x})$ be a continuous real-valued function defined on a set S of R^n. If the set $T = \{\boldsymbol{x} \in S : f(\boldsymbol{x}) \geq f(\boldsymbol{y})\}$ is compact for some $\boldsymbol{y} \in S$, then $f(\boldsymbol{x})$ attains its supremum on S. Similarly, if $T = \{\boldsymbol{x} \in S : f(\boldsymbol{x}) \leq f(\boldsymbol{y})\}$ is compact for some $\boldsymbol{y} \in S$, then $f(\boldsymbol{x})$ attains its infimum on S. Both conclusions apply when S itself is compact.*

Proof: Consider the question of whether the function $f(\boldsymbol{x})$ attains its supremum $u = \sup_{\boldsymbol{x} \in S} f(\boldsymbol{x})$. The set $f(T)$ is bounded by virtue of Proposition 2.5.3, and the supremum of $f(\boldsymbol{x})$ on T coincides with u. For every positive integer i choose a point $\boldsymbol{x}_i \in T$ such that $f(\boldsymbol{x}_i) \geq u - 1/i$. In view of the compactness of T, we can extract a convergent subsequence of $\boldsymbol{x}_i$ with limit $\boldsymbol{z} \in T$. The continuity of $f(\boldsymbol{x})$ along this subsequence then implies that $f(\boldsymbol{z}) = u$. ■

Example 2.5.5 *Closest Point in a Set*

To prove that the distance $\operatorname{dist}(\boldsymbol{x}, S)$ is achieved for some $\boldsymbol{z} \in S$, we must assume that S is closed. In finding the closest point to $\boldsymbol{x}$ in S, choose any point $\boldsymbol{y} \in S$. The set $T = S \cap \{\boldsymbol{z} : \|\boldsymbol{z} - \boldsymbol{x}\| \leq \|\boldsymbol{y} - \boldsymbol{x}\|\}$ is both closed and bounded and therefore compact. Proposition 2.5.4 now informs us that the continuous function $\boldsymbol{z} \mapsto \|\boldsymbol{z} - \boldsymbol{x}\|$ attains its infimum on S. ■

Example 2.5.6 *Equivalence of Norms*

Every norm $\|\boldsymbol{x}\|_\dagger$ on R^n is equivalent to the Euclidean norm $\|\boldsymbol{x}\|$ in the sense that there exist positive constants a and b such that the inequalities

$$a\|\boldsymbol{x}\| \quad \leq \quad \|\boldsymbol{x}\|_\dagger \quad \leq \quad b\|\boldsymbol{x}\| \tag{2.6}$$

hold for all $\boldsymbol{x}$. To prove the right inequality in (2.6), let $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ denote the standard basis. Then conditions (c) and (d) defining a norm indicate that $\boldsymbol{x} = \sum_{i=1}^n x_i \boldsymbol{e}_i$ satisfies

$$\|\boldsymbol{x}\|_\dagger \quad \leq \quad \sum_{i=1}^n |x_i| \cdot \|\boldsymbol{e}_i\|_\dagger \quad = \quad \|\boldsymbol{x}\| \sum_{i=1}^n \|\boldsymbol{e}_i\|_\dagger.$$

This proves the upper bound with $b = \sum_{i=1}^n \|\boldsymbol{e}_i\|_\dagger$.

To establish the lower bound, we note that property (c) of a norm allows us to restrict attention to the sphere $S = \{\boldsymbol{x} : \|\boldsymbol{x}\| = 1\}$. Now the function $\boldsymbol{x} \mapsto \|\boldsymbol{x}\|_\dagger$ is uniformly continuous on R^n because

$$\big| \|\boldsymbol{x}\|_\dagger - \|\boldsymbol{y}\|_\dagger \big| \quad \leq \quad \|\boldsymbol{x} - \boldsymbol{y}\|_\dagger \quad \leq \quad b\|\boldsymbol{x} - \boldsymbol{y}\|$$

follows from the upper bound just demonstrated. Since the sphere S is compact, the continuous function $\boldsymbol{x} \to \|\boldsymbol{x}\|_\dagger$ attains its lower bound a on S. In view of property (b) defining a norm, $a > 0$. ■

Example 2.5.7 *The Fundamental Theorem of Algebra*

Consider a polynomial $p(z) = c_n z^n + c_{n-1} z^{n-1} + \cdots + c_0$ in the complex variable z with $c_n \neq 0$. The fundamental theorem of algebra says that $p(z)$ has a root. d'Alembert suggested an interesting optimization proof of this fact [30]. We begin by observing that if we identify a complex number with an ordered pair of real numbers, then the domain of the real-valued function $|p(z)|$ is R^2. The identity

$$|p(z)| \quad = \quad |z|^n \left| c_n + \frac{c_{n-1}}{z} + \cdots + \frac{c_0}{z^n} \right|$$

shows that $|p(z)|$ tends to ∞ whenever $|z|$ tends to ∞. Therefore, the set $T = \{z : |p(z| \leq d\}$ is compact for any d, and Proposition 2.5.4 implies that $|p(z)|$ attains its minimum at some point y. Expanding $p(z)$ around y gives a polynomial

$$q(z) \quad = \quad p(z + y) \quad = \quad b_n z^n + b_{n-1} z^{n-1} + \cdots + b_0$$

with the same degree as $p(z)$. Furthermore, the minimum of $|q(z)|$ occurs at $z = 0$. Suppose $b_1 = \cdots = b_{k-1} = 0$ and $b_k \neq 0$. For some angle $\theta \in [0, 2\pi)$, the scaled complex exponential

$$u \quad = \quad \left| \frac{b_0}{b_k} \right|^{1/k} e^{i\theta/k}$$

is a root of the equation $b_k u^k + b_0 = 0$. The function $f(t) = |q(tu)|$ clearly satisfies $f(t) \geq |b_0|$ and

$$f(t) \quad = \quad |b_k t^k u^k + b_0| + o(t^k) \quad = \quad |b_0(1 - t^k)| + o(t^k)$$

for t small and positive. These two conditions are compatible only if $b_0 = 0$. Hence, the minimum of $|q(z)| = |p(z + y)|$ is 0. ■

Example 2.5.8 *Continuity of the Roots of a Polynomial*

As a followup to the previous example, let us prove that the roots of a polynomial depend continuously on its coefficients [261]. One has to exercise caution in stating this result. First, we limit ourselves to monic polynomials $p(z) = z^n + c_{n-1}z^{n-1} + \cdots + c_0$. Second, we rely on the fact that a monic polynomial can be written in factored form as

$$p(z) = (z - r_1) \cdots (z - r_n) \tag{2.7}$$

based on the roots guaranteed by the fundamental theorem of algebra. Let r be a root of $p(z)$ of multiplicity m and $q(z) = z^n + d_{n-1}z^{n-1} + \cdots + d_0$ be a second monic polynomial of the same degree n as $p(z)$. We now interpret continuity to mean that for every $\epsilon > 0$, there exists a $\delta > 0$ such that $q(z)$ has at least m roots within ϵ of r whenever the coefficient vector $\boldsymbol{d}$ of $q(z)$ satisfies $\|\boldsymbol{d} - \boldsymbol{c}\| < \delta$. Here we use the Euclidean norm on R^{2n}. In proving this result, we need the simple bound

$$|r_j| \leq \max\Big\{1, \sum_{i=0}^{n-1} |c_i|\Big\}.$$

on the roots of a monic polynomial $p(z)$ in terms of its coefficients. The proof of the bound is an immediate consequence of the identity

$$r_j = -\sum_{i=0}^{n-1} c_i r_j^{i-n+1}.$$

We are now in a position to verify the asserted continuous dependence. Suppose it fails for the polynomial $p(z)$ and the specified root r. Then for some $\epsilon > 0$ there exists a sequence $q_k(z)$ of monic polynomials of degree n with fewer than m roots within ϵ of r but whose coefficients d_{ki} converge to the coefficients c_i. Since the coefficients of the $q_k(z)$ converge, by the above inequality, the roots s_{ki} of the $q_k(z)$ are bounded. We can therefore extract a subsequence $q_{k_l}(z)$ whose roots converge to the complex numbers t_i. At most $m - 1$ of the t_i equal r. The representation

$$p(z) = \lim_{l\to\infty} q_{k_l}(z) = (z - t_1) \cdots (z - t_n)$$

is at odds with the representation (2.7) of $p(z)$. Indeed, one has m roots equal to r, and the other has at most $m - 1$ roots equal to r. This contradiction proves the claimed continuity of the roots.

As an illustration consider the quadratic $p(z) = z^2 - 2z + 1 = (z - 1)^2$ with the root 1 of multiplicity 2. For $\delta > 0$ small the related polynomial $z^2 - 2z + 1 - \delta$ has the real roots $1 \pm \sqrt{\delta}$ while the polynomial $z^2 - 2z + 1 + \delta$ has the complex roots $1 \pm \sqrt{-\delta}$. A more important application concerns the continuity of the eigenvalues of a matrix. Suppose the sequence $\boldsymbol{M}_k$ of square matrices converges to the square matrix $\boldsymbol{M}$. Then the sequence of characteristic polynomials $\det(z\boldsymbol{I} - \boldsymbol{M}_k)$ converges to the characteristic polynomial $\det(z\boldsymbol{I} - \boldsymbol{M})$. It follows that the eigenvalues of $\boldsymbol{M}_k$ converge to the eigenvalues of $\boldsymbol{M}$ in the sense just explained. ■

A function $f(\boldsymbol{x})$ is said to be uniformly continuous on its domain S if for every $\epsilon > 0$ there exists a $\delta > 0$ such that $\|f(\boldsymbol{y}) - f(\boldsymbol{x})\| < \epsilon$ whenever $\|\boldsymbol{y} - \boldsymbol{x}\| < \delta$. This sounds like ordinary continuity, but the chosen δ does not depend on the pivotal point $\boldsymbol{x} \in S$. One of the virtues of a compact domain is that it forces uniform continuity.

Proposition 2.5.5 (Heine) *Every continuous function $f(\boldsymbol{x})$ from a compact set S of R^m into R^n is uniformly continuous.*

Proof: Suppose $f(\boldsymbol{x})$ fails to be uniformly continuous. Then for some $\epsilon > 0$, there exist sequences $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ from S such that $\lim_{i\to\infty} \|\boldsymbol{x}_i - \boldsymbol{y}_i\| = 0$ and $\|f(\boldsymbol{x}_i) - f(\boldsymbol{y}_i)\| \geq \epsilon$. Since S is compact, we can extract a subsequence of $\boldsymbol{x}_i$ that converges to a point $\boldsymbol{u} \in S$. Along the corresponding subsequence of $\boldsymbol{y}_i$ we can extract a subsubsequence that converges to a point $\boldsymbol{v} \in S$. Substituting the constructed subsubsequences for $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ if necessary, we may assume that $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ both converge to the same limit $\boldsymbol{u} = \boldsymbol{v}$. The condition $\|f(\boldsymbol{x}_i) - f(\boldsymbol{y}_i)\| \geq \epsilon$ now contradicts the continuity of $f(\boldsymbol{x})$ at $\boldsymbol{u}$. ■

Example 2.5.9 *Rigid Motions*

Uniform continuity certainly appears in the absence of compactness. One spectacular example is a rigid motion. By this we mean a function $f(\boldsymbol{x})$ of R^n into itself with the property $\|f(\boldsymbol{y}) - f(\boldsymbol{x})\| = \|\boldsymbol{y} - \boldsymbol{x}\|$ for every choice of $\boldsymbol{x}$ and $\boldsymbol{y}$. We can better understand the rigid motion $f(\boldsymbol{x})$ by investigating the translated rigid motion $g(\boldsymbol{x}) = f(\boldsymbol{x}) - f(\mathbf{0})$ that maps the origin $\mathbf{0}$ into itself. Because $g(\boldsymbol{x})$ preserves distances, it also preserves inner products. This fact is evident from the equalities

$$\begin{aligned}
\|\boldsymbol{y} - \boldsymbol{x}\|^2 &= \|\boldsymbol{y}\|^2 - 2\boldsymbol{y}^*\boldsymbol{x} + \|\boldsymbol{x}\|^2 \\
\|g(\boldsymbol{y}) - g(\boldsymbol{x})\|^2 &= \|g(\boldsymbol{y})\|^2 - 2g(\boldsymbol{y})^*g(\boldsymbol{x}) + \|g(\boldsymbol{x})\|^2 \\
\|g(\boldsymbol{y})\|^2 &= \|\boldsymbol{y}\|^2 \\
\|g(\boldsymbol{x})\|^2 &= \|\boldsymbol{x}\|^2.
\end{aligned}$$

The inner product identity

$$g(\boldsymbol{y})^*g(\boldsymbol{x}) = \boldsymbol{y}^*\boldsymbol{x}$$

is only possible if $g(\boldsymbol{y})$ is linear. To demonstrate this assertion, note that $g(\boldsymbol{x})$ maps the standard orthonormal basis $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ onto the orthonormal basis $g(\boldsymbol{e}_1), \ldots, g(\boldsymbol{e}_n)$. Because

$$\begin{aligned}
g(\alpha\boldsymbol{x} + \beta\boldsymbol{y})^*g(\boldsymbol{e}_i) &= (\alpha\boldsymbol{x} + \beta\boldsymbol{y})^*\boldsymbol{e}_i \\
&= \alpha\boldsymbol{x}^*\boldsymbol{e}_i + \beta\boldsymbol{y}^*\boldsymbol{e}_i \\
&= [\alpha g(\boldsymbol{x}) + \beta g(\boldsymbol{y})]^*g(\boldsymbol{e}_i)
\end{aligned}$$

holds for all i, it follows that $g(\alpha \boldsymbol{x} + \beta \boldsymbol{y}) = \alpha g(\boldsymbol{x}) + \beta g(\boldsymbol{y})$. In other words, $g(\boldsymbol{x})$ is linear. The linear transformations that preserve angles and distances are precisely the orthogonal transformations. Thus, the rigid motion $f(\boldsymbol{x})$ reduces to an orthogonal transformation $\boldsymbol{U}\boldsymbol{x}$ followed by the translation $f(\boldsymbol{0})$. Conversely, it is trivial to prove that every such transformation

$$f(\boldsymbol{x}) \quad = \quad \boldsymbol{U}\boldsymbol{x} + f(\boldsymbol{0})$$

is a rigid motion. ■

Example 2.5.10 *Multilinear Maps*

A k-linear map $\boldsymbol{M}[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k]$ transforms points from the k-fold Cartesian product $\mathsf{R}^m \times \mathsf{R}^m \times \cdots \times \mathsf{R}^m$ into points in R^n and satisfies the rules

$$\begin{aligned} \boldsymbol{M}[\boldsymbol{u}_1, \ldots, c\boldsymbol{u}_j, \ldots, \boldsymbol{u}_k] &= c\boldsymbol{M}[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_j, \ldots, \boldsymbol{u}_k] \\ \boldsymbol{M}[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_j + \boldsymbol{v}_j, \ldots, \boldsymbol{u}_k] &= \boldsymbol{M}[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_j, \ldots, \boldsymbol{u}_k] \\ &\quad + \boldsymbol{M}[\boldsymbol{u}_1, \ldots, \boldsymbol{v}_j, \ldots, \boldsymbol{u}_k] \end{aligned}$$

for every scalar c, index j, vector $\boldsymbol{v}_j$, and combination of vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$. For example, matrix multiplication $\boldsymbol{u} \mapsto \boldsymbol{A}\boldsymbol{u}$ is 1-linear and the determinant map $\boldsymbol{U} \mapsto \det \boldsymbol{U}$ is k-linear on the columns $\boldsymbol{u}_j$ of a $k \times k$ matrix. A k-linear map into the real line ($n = 1$) is called a k-linear form. The k-linear form

$$[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k] \quad \mapsto \quad \prod_{j=1}^{k} \boldsymbol{v}_j^* \boldsymbol{u}_j \tag{2.8}$$

for any fixed combination $[\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k]$ of vectors is often useful in applications. A k-linear map $\boldsymbol{M}[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_j, \ldots, \boldsymbol{u}_k]$ is said to be symmetric if

$$\boldsymbol{M}[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_i, \ldots, \boldsymbol{u}_j, \ldots, \boldsymbol{u}_k] \quad = \quad \boldsymbol{M}[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_j, \ldots, \boldsymbol{u}_i, \ldots, \boldsymbol{u}_k]$$

for all pairs of indices i and j and antisymmetric if

$$\boldsymbol{M}[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_i, \ldots, \boldsymbol{u}_j, \ldots, \boldsymbol{u}_k] \quad = \quad -\boldsymbol{M}[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_j, \ldots, \boldsymbol{u}_i, \ldots, \boldsymbol{u}_k].$$

The determinant function is antisymmetric.

One can easily check that the collection $\mathsf{L}^k(\mathsf{R}^m, \mathsf{R}^n)$ of k-linear maps from $\mathsf{R}^m \times \mathsf{R}^m \times \cdots \times \mathsf{R}^m$ to R^n forms a vector space under pointwise addition and scalar multiplication. Its dimension is $m^k n$. Indeed, let $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m$ be a basis for R^m and $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n$ be a basis for R^n. If $\boldsymbol{u}_i = \sum_{j=1}^m c_{ij} \boldsymbol{e}_j$, then the expansion

$$\boldsymbol{M}[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k] \quad = \quad \sum_{j_1=1}^{m} \cdots \sum_{j_k=1}^{m} \Big(\prod_{i=1}^{k} c_{i,j_i} \Big) \boldsymbol{M}[\boldsymbol{e}_{j_1}, \ldots, \boldsymbol{e}_{j_k}] \tag{2.9}$$

correctly suggests that the k-linear maps with

$$\boldsymbol{M}_{j_1,\ldots,j_k,l}[\boldsymbol{e}_{i_1},\ldots,\boldsymbol{e}_{i_k}] = \begin{cases} \boldsymbol{f}_l & i_1 = j_1,\ldots,i_k = j_k \\ \boldsymbol{0} & \text{otherwise} \end{cases}$$

constitute a basis of $\mathsf{L}^k(\mathsf{R}^m,\mathsf{R}^n)$. For a linear form $\boldsymbol{M}[\boldsymbol{u}_1,\ldots,\boldsymbol{u}_k]$, it is helpful to think of the numbers $\boldsymbol{M}[\boldsymbol{e}_{i_1},\ldots,\boldsymbol{e}_{i_k}]$ as coefficients that define the linear form just as the coefficients of a matrix define the corresponding linear transformation.

Equation (2.9) also implies that $\boldsymbol{M}[\boldsymbol{u}_1,\ldots,\boldsymbol{u}_k]$ is a continuous function of its arguments. Therefore, the norm

$$\|\boldsymbol{M}\| = \sup_{\boldsymbol{u}_j \neq \boldsymbol{0}\,\forall j} \frac{\|\boldsymbol{M}[\boldsymbol{u}_1,\ldots,\boldsymbol{u}_k]\|}{\|\boldsymbol{u}_1\|\cdots\|\boldsymbol{u}_k\|} = \sup_{\|\boldsymbol{u}_j\|=1\,\forall j} \|\boldsymbol{M}[\boldsymbol{u}_1,\ldots,\boldsymbol{u}_k]\|$$

on $\mathsf{L}^k(\mathsf{R}^m,\mathsf{R}^n)$ induced by the Euclidean norms on R^m and R^n is finite. For example, the norm of the k-linear form (2.8) is $\prod_{j=1}^k \|\boldsymbol{v}_j\|$. This value is attained by choosing $\boldsymbol{u}_j = \|\boldsymbol{v}_j\|^{-1}\boldsymbol{v}_j$ and serves as an absolute upper bound on the k-linear form on unit vectors by virtue of the Cauchy-Schwarz inequality. The inequality

$$\|\boldsymbol{M}[\boldsymbol{u}_1,\ldots,\boldsymbol{u}_k]\| \leq \|\boldsymbol{M}\|\|\boldsymbol{u}_1\|\cdots\|\boldsymbol{u}_k\| \tag{2.10}$$

is an immediate consequence of the definition of $\|\boldsymbol{M}\|$. Problem 33 asks the reader to verify that the map $(\boldsymbol{M},\boldsymbol{u}_1,\ldots,\boldsymbol{u}_k) \mapsto \boldsymbol{M}[\boldsymbol{u}_1,\ldots,\boldsymbol{u}_k]$ is jointly continuous in its $k+1$ variables. ■

2.6 Semicontinuity

For real-valued functions, the notions of lower and upper semicontinuity are often useful substitutes for continuity. A real-valued function $f(\boldsymbol{x})$ with domain $T \subset \mathsf{R}^m$ is lower semicontinuous if the set $\{\boldsymbol{x} \in T : f(\boldsymbol{x}) \leq c\}$ is closed in T for every constant c. Given the duality of closed and open sets, an equivalent condition is that $\{\boldsymbol{x} \in T : f(\boldsymbol{x}) > c\}$ is open in T for every constant c. A real-valued function $g(\boldsymbol{x})$ is said to be upper semicontinuous if and only if $f(\boldsymbol{x}) = -g(\boldsymbol{x})$ is lower semicontinuous. Owing to this simple relationship, we will confine our attention to lower semicontinuous functions. The next proposition gives two alternative definitions.

Proposition 2.6.1 *A necessary and sufficient condition for $f(\boldsymbol{x})$ to be lower semicontinuous is that*

$$f(\boldsymbol{x}) \leq \liminf_{n\to\infty} f(\boldsymbol{x}_n) \tag{2.11}$$

whenever $\lim_{n\to\infty} \boldsymbol{x}_n = \boldsymbol{x}$ in T. Another necessary and sufficient condition is that the epigraph $\{(\boldsymbol{x},y) \in T \times \mathsf{R} : f(\boldsymbol{x}) \leq y\}$ is a closed set.

Proof: Suppose $f(\boldsymbol{x})$ is lower semicontinuous and $\lim_{n\to\infty} \boldsymbol{x}_n = \boldsymbol{x}$. For any $\epsilon > 0$, the point $\boldsymbol{x}$ lies in the open set $\{\boldsymbol{y} \in T : f(\boldsymbol{y}) > f(\boldsymbol{x}) - \epsilon\}$. Hence, $f(\boldsymbol{x}_n) > f(\boldsymbol{x}) - \epsilon$ for all sufficient large n. But this implies inequality (2.11). Similarly, if $\boldsymbol{x}_n$ converges to $\boldsymbol{x}$ and $y_n \geq f(\boldsymbol{x}_n)$ converges to y, then the inequality $y \geq \liminf_{n\to\infty} f(\boldsymbol{x}_n) \geq f(\boldsymbol{x})$ follows, and the epigraph is closed. Thus, both stated conditions are necessary.

For sufficiency, suppose inequality (2.11) holds. Consider a sequence $\boldsymbol{x}_n$ in the set $\{\boldsymbol{y} \in T : f(\boldsymbol{y}) \leq c\}$ with limit $\boldsymbol{x}$. It is clear that $f(\boldsymbol{x}) \leq c$ as well. Thus, $\{\boldsymbol{y} \in T : f(\boldsymbol{y}) \leq c\}$ is closed. To deal with the second sufficient condition, suppose the epigraph is closed, but $f(\boldsymbol{x})$ is not lower semicontinuous. Then there exists a sequence $\boldsymbol{x}_n$ converging to $\boldsymbol{x}$ in T and an $\epsilon > 0$ such that $f(\boldsymbol{x}) - \epsilon > \liminf_{n\to\infty} f(\boldsymbol{x}_n)$. It follows that the pair $(\boldsymbol{x}_n, f(\boldsymbol{x}) - \epsilon)$ is in the epigraph for infinitely many n. Because the epigraph is closed, this forces the contradiction that $(\boldsymbol{x}, f(\boldsymbol{x}) - \epsilon)$ belongs to the epigraph. ■

Part of the motivation for defining semicontinuity is to generalize Proposition 2.5.4. The result stated there for global maxima holds for upper semicontinuous functions, and the result for global minima holds for lower semicontinuous functions. The proof carries over almost word for word. It is also obvious that any continuous function is lower semicontinuous, and any function that is both lower and upper semicontinuous is continuous. Fortunately, the closure properties of lower semicontinuous functions are quite flexible.

Proposition 2.6.2 *The collection of lower semicontinuous functions with common domain $T \subset \mathsf{R}^m$ satisfies the following rules:*

(a) *If $f_k(\boldsymbol{x})$ is a family of lower semicontinuous functions, then* $\sup_k f_k(\boldsymbol{x})$ *is lower semicontinuous.*

(b) *If $f_k(\boldsymbol{x})$ is a finite family of lower semicontinuous functions, then* $\min_k f_k(\boldsymbol{x})$ *is lower semicontinuous.*

(c) *If $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$ are lower semicontinuous, then $f(\boldsymbol{x}) + g(\boldsymbol{x})$ is lower semicontinuous.*

(d) *If $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$ are both positive and lower semicontinuous, then $f(\boldsymbol{x})g(\boldsymbol{x})$ is lower semicontinuous.*

(e) *If $f(\boldsymbol{x})$ is lower semicontinuous and $g(\boldsymbol{x})$ is continuous with range U contained in T, then $f \circ g(\boldsymbol{x})$ is lower semicontinuous.*

Proof: These rules follow from the set identities

$$\begin{aligned}
\{\boldsymbol{x} : \sup_k f_k(\boldsymbol{x}) > c\} &= \cup_k \{\boldsymbol{x} : f_k(\boldsymbol{x}) > c\} \\
\{\boldsymbol{x} : \min_k f_k(\boldsymbol{x}) > c\} &= \cap_k \{\boldsymbol{x} : f_k(\boldsymbol{x}) > c\}
\end{aligned}$$

$$\begin{aligned}
\{\boldsymbol{x} : f(\boldsymbol{x}) + g(\boldsymbol{x}) > c\} &= \cup_d(\{\boldsymbol{x} : f(\boldsymbol{x}) > c - d\} \cap \{\boldsymbol{x} : g(\boldsymbol{x}) > d\}) \\
\{\boldsymbol{x} : f(\boldsymbol{x})g(\boldsymbol{x}) > c\} &= \cup_{d>0}(\{\boldsymbol{x} : f(\boldsymbol{x}) > d^{-1}c\} \cap \{\boldsymbol{x} : g(\boldsymbol{x}) > d\}) \\
\{\boldsymbol{y} : f \circ g(\boldsymbol{y}) > c\} &= g^{-1}[\{\boldsymbol{x} : f(\boldsymbol{x}) > c\}]
\end{aligned}$$

and the properties of open sets and continuous functions summarized in Propositions 2.4.2 and 2.5.2. ■

Example 2.6.1 *Row and Column Rank*

Every $m \times n$ matrix $\boldsymbol{A}$ has a well defined nullity and rank. Although these are not continuous functions of $\boldsymbol{A}$, the former function is upper semicontinuous, and the latter function is lower semicontinuous. In view of the dimension identity $\text{nullity}(\boldsymbol{A}) = n - \text{rank}(\boldsymbol{A})$, to validate both claims it suffices to show that $\text{rank}(\boldsymbol{A})$ is lower semicontinuous. Consider an arbitrary constant c and an arbitrary matrix $\boldsymbol{A} = (a_{ij})$ with $\text{rank}(\boldsymbol{A}) > c$. If we abbreviate $\text{rank}(\boldsymbol{A}) = r$, then there exist row indices $1 \le i_1 < \cdots < i_r \le m$ and column indices $1 \le j_1 < \cdots < j_r \le n$ such that the submatrix

$$\begin{pmatrix} a_{i_1 j_1} & \cdots & a_{i_1 j_r} \\ \vdots & \ddots & \vdots \\ a_{i_r j_1} & \cdots & a_{i_r j_r} \end{pmatrix}$$

has nonzero determinant. Because the determinant function is continuous, the same submatrix has nonvanishing determinant for all $m \times n$ matrices $\boldsymbol{B}$ close to $\boldsymbol{A}$. It follows that $\{\boldsymbol{A} : \text{rank}(\boldsymbol{A}) > c\}$ is an open set and therefore that $\text{rank}(\boldsymbol{A})$ is lower semicontinuous. ■

2.7 Connectedness

Roughly speaking, a set is disconnected if it can be split into two pieces sharing no boundary. A set is connected if it is not disconnected. One way of making this vague distinction precise is to consider a set S disconnected if there exists a real-valued continuous function $\phi(\boldsymbol{x})$ defined on S and having range $\{0, 1\}$. The nonempty subsets $A = \phi^{-1}(0)$ and $B = \phi^{-1}(1)$ then constitute the two disconnected pieces of S. According to part (b) of Proposition 2.5.2, both A and B are closed. Because one is the complement of the other, both are also open.

Arcwise connectedness is a variation on the theme of connectedness. A set is said to be arcwise connected if for any pair of points $\boldsymbol{x}$ and $\boldsymbol{y}$ of the set there is a continuous function $f(t)$ from the interval $[0, 1]$ into the set satisfying $f(0) = \boldsymbol{x}$ and $f(1) = \boldsymbol{y}$. We will see shortly that arcwise connectedness implies connectedness. On open sets, the two notions coincide.

Can we identify the connected subsets of the real line? Intuition suggests that the only connected subsets are intervals. Here a single point x is viewed as the interval $[x, x]$. Suppose S is a connected subset of R, and let a and b be two points of S. In order for S to be an interval, every point $c \in (a, b)$ should be in S. If S fails to contain an intermediate point c, then we can define a continuous function $\phi(x)$ disconnecting S by taking $\phi(x) = 0$ for $x < c$ and $\phi(x) = 1$ for $x > c$. Thus, every connected subset must be an interval.

To prove the converse, suppose a disconnecting function $\phi(x)$ lives on an interval. Select points a and b of the interval with $\phi(a) = 0$ and $\phi(b) = 1$. Without loss of generality we can take $a < b$. On $[a, b]$ we now carry out the bisection strategy of Example 2.3.1, selecting the right or left subinterval at each stage so that the values of $\phi(x)$ at the endpoints of the selected subinterval disagree. Eventually, bisection leads to a subinterval contradicting the uniform continuity of $\phi(x)$ on $[a, b]$. Indeed, there is a number δ such that $|\phi(y) - \phi(x)| < 1$ whenever $|y - x| < \delta$; at some stage, the length of the subinterval containing points with both values of $\phi(x)$ falls below δ.

This result is the first of four characterizing connected sets.

Proposition 2.7.1 *Connected subsets of* R^n *have the following properties:*

(a) *A subset of the real line is connected if and only if it is an interval.*

(b) *The image of a connected set under a continuous function is connected.*

(c) *The union* $S = \cup_\alpha S_\alpha$ *of an arbitrary collection of connected subsets is connected if one of the sets* S_β *has a nonempty intersection* $S_\beta \cap S_\alpha$ *with every other set* S_α*.*

(d) *Every arcwise connected set* S *is connected.*

Proof: To prove part (b) let $f(\boldsymbol{x})$ be a continuous map from a connected set $S \subset \mathsf{R}^m$ into R^n. If the image $f(S)$ is disconnected, then there is a continuous function $\phi(\boldsymbol{x})$ disconnecting it. The composition $\phi \circ f(\boldsymbol{x})$ is continuous by part (g) of Proposition 2.5.1 and serves to disconnect S, contradicting the connectedness of S. To prove (c) suppose that the continuous function $\phi(\boldsymbol{x})$ disconnects the union S. Then there exists $\boldsymbol{y} \in S_{\alpha_1}$ and $\boldsymbol{z} \in S_{\alpha_2}$ with $\phi(\boldsymbol{y}) = 0$ and $\phi(\boldsymbol{z}) = 1$. Choose $\boldsymbol{u} \in S_\beta \cap S_{\alpha_1}$ and $\boldsymbol{v} \in S_\beta \cap S_{\alpha_2}$. If $\phi(\boldsymbol{u}) \neq \phi(\boldsymbol{v})$, then $\phi(\boldsymbol{x})$ disconnects S_β. If $\phi(\boldsymbol{u}) = \phi(\boldsymbol{v})$, then $\phi(\boldsymbol{y}) \neq \phi(\boldsymbol{u})$ or $\phi(\boldsymbol{z}) \neq \phi(\boldsymbol{v})$. In the former case $\phi(\boldsymbol{x})$ disconnects S_{α_1}, and in the latter case $\phi(\boldsymbol{x})$ disconnects S_{α_2}. Finally, to prove part (d), suppose the arcwise connected set S fails to be connected. Then there exists a continuous disconnecting function $\phi(\boldsymbol{x})$ with $\phi(\boldsymbol{y}) = 0$ and $\phi(\boldsymbol{z}) = 1$. Let $f(t)$ be an arc in S connecting $\boldsymbol{y}$ and $\boldsymbol{z}$. The continuous function $\phi \circ f(t)$ then serves to disconnect $[0, 1]$. ■

Example 2.7.1 *The Intermediate Value Property*

Consider a continuous function $f(x)$ from an interval $[a, b]$ to the real line. The intermediate value theorem asserts that the image $f([a, b])$ coincides with the interval $[\min f(x), \max f(x)]$. This theorem, which is a consequence of properties (a) and (b) of Proposition 2.7.1, has many applications. For example, suppose $g(x)$ is a continuous function from $[0,1]$ into $[0,1]$. If $f(x) = g(x) - x$, then it is obvious that $f(0) \geq 0$ and $f(1) \leq 0$. It follows that $f(x) = 0$ for some x. In other words, $g(x)$ has a fixed point satisfying $g(x) = x$. ■

Example 2.7.2 *Connectedness of Spheres*

The set $S(\boldsymbol{x}, r)$ in R^n is the image of the continuous map $\boldsymbol{y} \mapsto \boldsymbol{x} + r\boldsymbol{y}/\|\boldsymbol{y}\|$ of the domain $T = \mathsf{R}^n \setminus \mathbf{0}$. Hence, to prove connectedness when $n > 1$, it suffices to prove that T is connected. To achieve this, we argue that T is arcwise connected. Consider two points $\boldsymbol{u}$ and $\boldsymbol{v}$ in T. If $\mathbf{0}$ does not lie on the line segment between $\boldsymbol{u}$ and $\boldsymbol{v}$, then we can use the function $f(t) = \boldsymbol{u} + t(\boldsymbol{v} - \boldsymbol{u})$ to connect $\boldsymbol{u}$ and $\boldsymbol{v}$. If $\mathbf{0}$ lies on the line segment, choose any $\boldsymbol{w}$ not on the line determined by $\boldsymbol{u}$ and $\boldsymbol{v}$. Now the continuous function

$$f(t) = \begin{cases} \boldsymbol{u} + 2t(\boldsymbol{w} - \boldsymbol{u}) & t \in [0, \frac{1}{2}] \\ \boldsymbol{w} + (2t-1)(\boldsymbol{v} - \boldsymbol{w}) & t \in [\frac{1}{2}, 1] \end{cases}$$

connects $\boldsymbol{u}$ and $\boldsymbol{v}$. The sphere $S(x, r)$ in R reduces to the two points $x - r$ and $x + r$ and is disconnected. ■

2.8 Uniform Convergence

Many delicate issues of analysis revolve around the question of whether a given property of a sequence of functions $f_m(\boldsymbol{x})$ is preserved under a passage to a limit. As a simple example, consider the sequence $f_m(x) = x^m$ of continuous functions defined on the unit interval $[0,1]$. It is clear that $f_m(x)$ converges pointwise to the discontinuous function

$$f(x) = \begin{cases} 0 & 0 \leq x < 1 \\ 1 & x = 1\,. \end{cases}$$

The failure of $f(x)$ to be continuous suggests that an additional hypothesis must be imposed. The key hypothesis is uniform convergence. This requires for each $\epsilon > 0$ that there exists an integer k such that $|f_m(\boldsymbol{x}) - f(\boldsymbol{x})| < \epsilon$ for all $m \geq k$ and all $\boldsymbol{x}$. Here the adjective "uniform" refers to the assumption that the same k works for all $\boldsymbol{x}$. Of course, k is allowed to depend on ϵ.

Proposition 2.8.1 *Suppose the sequence of continuous functions $f_m(\boldsymbol{x})$ maps a domain $D \subset \mathsf{R}^p$ into R^q. If $f_m(\boldsymbol{x})$ converges uniformly to $f(\boldsymbol{x})$ on D, then $f(\boldsymbol{x})$ is also continuous.*

Proof: Choose $\boldsymbol{y} \in D$ and $\epsilon > 0$, and take k so that $\|f_m(\boldsymbol{x}) - f(\boldsymbol{x})\| < \frac{\epsilon}{3}$ for all $m \geq k$ and $\boldsymbol{x}$. By virtue of the continuity of $f_k(\boldsymbol{x})$, there is a $\delta > 0$

such that $\|f_k(\boldsymbol{x}) - f_k(\boldsymbol{y})\| < \frac{\epsilon}{3}$ whenever $\|\boldsymbol{x} - \boldsymbol{y}\| < \delta$. Assuming that $\boldsymbol{y}$ is fixed and $\|\boldsymbol{x} - \boldsymbol{y}\| < \delta$, we have

$$\begin{aligned}
\|f(\boldsymbol{x}) - f(\boldsymbol{y})\| &\leq \|f(\boldsymbol{x}) - f_k(\boldsymbol{x})\| + \|f_k(\boldsymbol{x}) - f_k(\boldsymbol{y})\| + \|f_k(\boldsymbol{y}) - f(\boldsymbol{y})\| \\
&< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} \\
&= \epsilon.
\end{aligned}$$

This shows that $f(\boldsymbol{x})$ is continuous at $\boldsymbol{y}$. ■

Example 2.8.1 *Weierstrass M-Test*

Suppose the entries $g_k(\boldsymbol{x})$ of a sequence of continuous functions satisfy $\|g_k(\boldsymbol{x})\| \leq M_k$, where $\sum_{k=1}^{\infty} M_k < \infty$. Then Cauchy's criterion and Proposition 2.8.1 together imply that the partial sums $f_l(\boldsymbol{x}) = \sum_{k=1}^{l} g_k(\boldsymbol{x})$ converge uniformly to the continuous function $f(\boldsymbol{x}) = \sum_{k=1}^{\infty} g_k(\boldsymbol{x})$. ■

2.9 Problems

1. Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ be points in R^n. State and prove a necessary and sufficient condition under which the Euclidean norm equality

$$\|\boldsymbol{x}_1 + \cdots + \boldsymbol{x}_m\| = \|\boldsymbol{x}_1\| + \cdots + \|\boldsymbol{x}_m\|$$

holds. (Hints: Square and expand both sides. Use the necessary and sufficient conditions of the Cauchy-Schwarz inequality term by term.)

2. Show that it is possible to choose $n+1$ points $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ in R^n such that $\|\boldsymbol{x}_i\| = 1$ for all i and $\|\boldsymbol{x}_i - \boldsymbol{x}_j\| = \|\boldsymbol{x}_k - \boldsymbol{x}_l\|$ for all pairs $i \neq j$ and $k \neq l$. These points define a regular simplex with vertices on the unit sphere. (Hint: One possibility is to take $\boldsymbol{x}_0 = n^{-1/2}\mathbf{1}$ and $\boldsymbol{x}_i = a\mathbf{1} + b\boldsymbol{e}_i$ for $i \geq 1$, where

$$a = -\frac{1 + \sqrt{n+1}}{n^{3/2}}, \quad b = \sqrt{\frac{n+1}{n}}.$$

Any rotated version of these points also works.)

3. Show that

$$\|\boldsymbol{x}\|_q \leq \|\boldsymbol{x}\|_p \tag{2.12}$$
$$\|\boldsymbol{x}\|_p \leq n^{\frac{1}{p} - \frac{1}{q}} \|\boldsymbol{x}\|_q \tag{2.13}$$

when p and q are chosen from $\{1, 2, \infty\}$ and $p < q$. Here $\|\boldsymbol{x}\|_2$ is the Euclidean norm on R^n. These inequalities are sharp. Equality holds in inequality (2.12) when $\boldsymbol{x} = (1, 0, \ldots, 0)^*$, and equality holds in inequality (2.13) when $\boldsymbol{x} = (1, 1, \ldots, 1)^*$.

4. Show that $\|\boldsymbol{x}\|^2 \leq \|\boldsymbol{x}\|_\infty \|\boldsymbol{x}\|_1 \leq \sqrt{n}\|\boldsymbol{x}\|^2$ for any vector $\boldsymbol{x} \in \mathsf{R}^n$.

5. Prove that $1 \leq \|\boldsymbol{I}\|_\dagger$ and $\|\boldsymbol{M}\|_\dagger^{-1} \leq \|\boldsymbol{M}^{-1}\|_\dagger$ for any matrix norm on square matrices satisfying the defining properties (a) through (e) of Sect. 2.2.

6. Set $\|\boldsymbol{M}\|_{\max} = \max_{i,j} |m_{ij}|$ for $\boldsymbol{M} = (m_{ij})$. Show that this defines a vector norm but not a matrix norm on $n \times n$ matrices $\boldsymbol{M}$.

7. Let $\boldsymbol{M}$ be an $m \times n$ matrix. Prove that its spectral norm satisfies

$$\|\boldsymbol{M}\| \quad = \quad \sup_{\|\boldsymbol{v}\|=1} \|\boldsymbol{M}\boldsymbol{v}\| \quad = \quad \sup_{\|\boldsymbol{u}\|=1,\ \|\boldsymbol{v}\|=1} \boldsymbol{u}^*\boldsymbol{M}\boldsymbol{v}$$

for $\boldsymbol{u} \in \mathsf{R}^m$ and $\boldsymbol{v} \in \mathsf{R}^n$.

8. Demonstrate that the spectral norm on $m \times n$ matrices satisfies $\|\boldsymbol{U}\boldsymbol{M}\| = \|\boldsymbol{M}\| = \|\boldsymbol{M}\boldsymbol{V}\|$ for all orthogonal matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ of the right dimensions. Show that the Frobenius norm satisfies the same orthogonal invariance principle.

9. Show that an $m \times n$ matrix $\boldsymbol{M} = (m_{ij})$ has the matrix norms

$$\begin{aligned}\|\boldsymbol{M}\|_1 &= \max_{1 \leq j \leq n} \sum_{i=1}^m |m_{ij}| \\ \|\boldsymbol{M}\|_\infty &= \max_{1 \leq i \leq m} \sum_{j=1}^n |m_{ij}|\end{aligned}$$

induced by the vector norms $\|\boldsymbol{x}\|_1$ and $\|\boldsymbol{x}\|_\infty$.

10. Let $\boldsymbol{M}$ be an $m \times n$ matrix of full column rank n. Prove that there exists a positive constant c such that $\|\boldsymbol{M}\boldsymbol{y}\| \geq c\|\boldsymbol{y}\|$ for all $\boldsymbol{y} \in \mathsf{R}^n$.

11. Demonstrate properties (2.4) and (2.5) of the limit superior and limit inferior. Also check that the sequence x_n has a limit if and only if equality holds in inequality (2.5).

12. Let $l = \limsup_{n\to\infty} x_n$. Show that:

    (a) $l = -\infty$ if and only if $\lim_{n\to\infty} x_n = -\infty$.

    (b) $l = +\infty$ if and only if for every positive integer m and real r there exists an $n \geq m$ with $x_n > r$.

    (c) l is finite if and only if (a) for every $\epsilon > 0$ there is an m such that $n \geq m$ implies $x_n < l + \epsilon$ and (b) for every $\epsilon > 0$ and every m there is an $n \geq m$ such that $x_n > l - \epsilon$.

    Similar properties hold for the limit inferior.

13. For any sequence of real numbers x_n, prove that

$$\inf_n x_n \quad \le \quad \liminf_{n\to\infty} x_n \quad \text{and} \quad \limsup_{n\to\infty} x_n \quad \le \quad \sup_n x_n .$$

If y_n is a second sequence of real numbers, then prove that

$$\begin{aligned}\liminf_{n\to\infty}\,(x_n+y_n) &\ \ge\ \liminf_{n\to\infty} x_n + \liminf_{n\to\infty} y_n \\ \limsup_{n\to\infty}\,(x_n+y_n) &\ \le\ \limsup_{n\to\infty} x_n + \limsup_{n\to\infty} y_n .\end{aligned}$$

Finally, if $x_n \le y_n$ for all n, then prove that

$$\liminf_{n\to\infty} x_n \quad \le \quad \liminf_{n\to\infty} y_n \quad \text{and} \quad \limsup_{n\to\infty} x_n \quad \le \quad \limsup_{n\to\infty} y_n .$$

14. Let x_n be a sequence of nonnegative real numbers with

$$x_{n+1} \quad \le \quad x_n + \frac{1}{n^2}$$

for all $n \ge 1$. Show that $\lim_{n\to\infty} x_n$ exists [69].

15. Let $\boldsymbol{x}_m$ be a convergent sequence in R^n with limit $\boldsymbol{x}$. Prove that the sequence $\boldsymbol{s}_m = (\boldsymbol{x}_1 + \cdots + \boldsymbol{x}_m)/m$ of arithmetic means converges to $\boldsymbol{x}$.

16. Show that

$$\lim_{x\to\infty} p(x)e^{-x} \quad = \quad 0$$

for every polynomial $p(x)$.

17. Prove that the set of invertible matrices is open and that the sets of symmetric and orthogonal matrices are closed in R^{n^2}.

18. A square matrix is nilpotent if $\boldsymbol{A}^k = \boldsymbol{0}$ for some positive integer k. If $\boldsymbol{A}$ and $\boldsymbol{B}$ are nilpotent, then show that $\boldsymbol{A}+\boldsymbol{B}$ need not be nilpotent. If we add the hypothesis that $\boldsymbol{A}$ and $\boldsymbol{B}$ commute, then show that $\boldsymbol{A}+\boldsymbol{B}$ is nilpotent. Use Example 2.3.3 to construct the inverses of the matrices $\boldsymbol{I}+\boldsymbol{A}$ and $\boldsymbol{I}-\boldsymbol{A}$ for $\boldsymbol{A}$ nilpotent [69].

19. Show that $e^{-\boldsymbol{M}}$ is the matrix inverse of $e^{\boldsymbol{M}}$. A skew symmetric matrix $\boldsymbol{M}$ satisfies $\boldsymbol{M}^* = -\boldsymbol{M}$. Show that $e^{\boldsymbol{M}}$ is orthogonal when $\boldsymbol{M}$ is skew symmetric.

20. Demonstrate that the matrix exponential function $\boldsymbol{M} \mapsto e^{\boldsymbol{M}}$ is continuous. (Hint: Apply the Weierstrass M-test.)

21. Demonstrate that the function

$$f(\boldsymbol{x}) \quad = \quad \begin{cases} \frac{x_1x_2}{x_1^2-x_2^2} & |x_1| \ne |x_2| \\ 0 & \text{otherwise} \end{cases}$$

on R^2 is discontinuous at $\boldsymbol{0}$.

22. Let $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$ be real-valued continuous functions defined on the same domain. Prove that $\max\{f(\boldsymbol{x}), g(\boldsymbol{x})\}$ and $\min\{f(\boldsymbol{x}), g(\boldsymbol{x})\}$ are continuous functions. Prove that the function $f(\boldsymbol{x}) = \max_i x_i$ is continuous on R^n.

23. Define the function

$$f(x) = \begin{cases} x & x \text{ is rational} \\ 1-x & x \text{ is irrational} \end{cases}$$

on $[0,1]$. At what points is $f(x)$ continuous? What is the image $f([0,1])$?

24. Give an example of a continuous function that does not map an open set to an open set. Give another example of a continuous function that does not map a closed set to a closed set.

25. Show that the set of $n \times n$ orthogonal matrices is compact. (Hint: Show that every orthogonal matrix $\boldsymbol{O}$ has norm $\|\boldsymbol{O}\| = 1$.)

26. Let $f(\boldsymbol{x})$ be a continuous function from a compact set $S \subset \mathsf{R}^m$ into R^n. If $f(\boldsymbol{x})$ is one-to-one, then demonstrate that the inverse function $f^{-1}(\boldsymbol{y})$ is continuous from $f(S)$ to S.

27. Let $C = A \times B$ be the Cartesian product of two subsets $A \subset \mathsf{R}^m$ and $B \subset \mathsf{R}^n$. Prove that:

    (a) C is closed in R^{m+n} if both A and B are closed.

    (b) C is open in R^{m+n} if both A and B are open.

    (c) C is compact in R^{m+n} if both A and B are compact.

    (d) C is connected in R^{m+n} if both A and B are connected.

28. Prove the converse of each of the assertions in Problem 27.

29. Without appeal to Proposition 2.5.5, show that every polynomial on R is uniformly continuous on a compact interval $[a,b]$.

30. Let $f(x)$ be uniformly continuous on R and satisfy $f(0) = 0$. Demonstrate that there exists a nonnegative constant c such that

$$|f(x)| \leq 1 + c|x|$$

for all x [69].

31. Suppose that $f(x)$ is continuous on $[0,\infty)$ and $\lim_{x\to\infty} f(x)$ exists and is finite. Prove that $f(x)$ is uniformly continuous on $[0,\infty)$ [69].

32. Characterize those maps $f(\boldsymbol{x})$ from R^n into itself that have the property $\|f(\boldsymbol{y}) - f(\boldsymbol{x})\| = c\|\boldsymbol{y} - \boldsymbol{x}\|$ for all $\boldsymbol{x}$ and $\boldsymbol{y}$. Here the constant c need not equal 1.

33. Prove that the multilinear map $(\boldsymbol{M}, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_k) \mapsto \boldsymbol{M}[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k]$ is jointly continuous in its $k+1$ variables. (Hint: Write

$$\begin{aligned}
\boldsymbol{M}[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k] - \boldsymbol{N}[\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k] &= (\boldsymbol{M} - \boldsymbol{N})[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k] \\
&\quad + \boldsymbol{N}[\boldsymbol{u}_1 - \boldsymbol{v}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k] \\
&\quad + \boldsymbol{N}[\boldsymbol{v}_1, \boldsymbol{u}_2 - \boldsymbol{v}_2, \ldots, \boldsymbol{u}_k] \\
&\quad\ \vdots \\
&\quad + \boldsymbol{N}[\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{u}_k - \boldsymbol{v}_k]
\end{aligned}$$

and take norms.)

34. Let $\boldsymbol{M}[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k]$ be a symmetric k-linear map. Demonstrate that

$$\boldsymbol{M}[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k] = \frac{1}{2^k k!} \sum \epsilon_1 \cdots \epsilon_k \boldsymbol{M}[(\epsilon_1 \boldsymbol{u}_1 + \cdots + \epsilon_k \boldsymbol{u}_k)^k],$$

where the sum ranges over all combinations of $\epsilon_1 = \pm 1, \ldots, \epsilon_k = \pm 1$. Hence, a symmetric k-linear map is determined by its values on the diagonal of its domain.

35. Continuing Problem 34, define the alternative norm

$$\|\boldsymbol{M}\|_{\text{sym}} = \sup_{\boldsymbol{u} \neq \boldsymbol{0}} \frac{\|\boldsymbol{M}[\boldsymbol{u}^k]\|}{\|\boldsymbol{u}\|^k} = \sup_{\|\boldsymbol{u}\|=1} \|\boldsymbol{M}[\boldsymbol{u}^k]\|.$$

Prove the inequalities

$$\|\boldsymbol{M}\|_{\text{sym}} \leq \|\boldsymbol{M}\| \leq \frac{k^k}{k!} \|\boldsymbol{M}\|_{\text{sym}}.$$

36. Show that the indicator function of an open set is lower semicontinuous and that the indicator function of a closed set is upper semicontinuous. Also show that floor function $f(x) = \lfloor x \rfloor$ is upper semicontinuous and that the ceiling function $f(x) = \lceil x \rceil$ is lower semicontinuous.

37. Suppose the n numbers $x_1, \ldots, x_n$ lie on $[0, 1]$. Prove that the function

$$f(x) = \frac{1}{n} \sum_{i=1}^{n} |x - x_i|$$

attains the value $\frac{1}{2}$ for some $x \in [0, 1]$. (Hint: Consider $f(0)$ and $f(1)$.)

38. Show that a hyperplane $\{\boldsymbol{x} : \boldsymbol{z}^* \boldsymbol{x} = c\}$ in R^n is connected but that its complement is disconnected.

39. If the real-valued function $f(x)$ on $[a, b]$ is continuous and one-to-one, then prove that $f(x)$ is either strictly increasing or strictly decreasing.

40. Demonstrate that a polynomial of odd degree possesses at least one real root.

41. Prove that the closure of a connected set is connected.

42. Suppose T is a connected set in R^n. Define

$$\begin{aligned} U_\epsilon &= \{\boldsymbol{y} \in \mathsf{R}^n : \operatorname{dist}(\boldsymbol{y}, T) < \epsilon\} \\ V_\epsilon &= \{\boldsymbol{y} \in \mathsf{R}^n : \operatorname{dist}(\boldsymbol{y}, T) \leq \epsilon\} \end{aligned}$$

for $\epsilon > 0$ and $\operatorname{dist}(\boldsymbol{y}, T) = \inf_{\boldsymbol{x} \in T} \|\boldsymbol{y} - \boldsymbol{x}\|$. Demonstrate that U_ϵ and V_ϵ are connected. (Hints: V_ϵ is the closure of U_ϵ. For U_ϵ argue by contradiction using the definition of a connected set.)

43. On what domains do the sequences of functions

    (a) $f_n(x) = (1 + x/n)^n$
    (b) $f_n(x) = nx/(1 + n^2x^2)$
    (c) $f_n(x) = n^x$
    (d) $f_n(x) = x^{-1} \sin(nx)$
    (e) $f_n(x) = xe^{-nx}$
    (f) $f_n(x) = x^{2n}/(1 + x^{2n})$

    converge [68]? On what domains do they converge uniformly?

44. Suppose that $f(x)$ is a function from the real line to itself satisfying $f(x + y) = f(x) + f(y)$ for all x and y. If $f(x)$ is continuous at a single point, then show that $f(x) = cx$ for some constant c. (Hints: Prove that $f(x)$ is continuous everywhere and that $f(q) = f(1)q$ for all rational numbers q.)

45. Suppose that $g(x)$ is a function from the real line to itself satisfying $g(x + y) = g(x)g(y)$ for all x and y. If $g(x)$ is continuous at a single point, then prove that either $g(x)$ is identically 0 or that there exists a positive constant d with $g(x) = d^x$. (Hint: Show that either $g(x)$ is identically 0 or that $g(x)$ is positive for all x. In the latter case, take logarithms and reduce to the previous problem.)

46. Suppose the real-valued function $f(\boldsymbol{x}, \boldsymbol{y})$ is jointly continuous in its two vector arguments and C is a compact set. Show that the functions

$$g(\boldsymbol{x}) = \inf_{\boldsymbol{y} \in C} f(\boldsymbol{x}, \boldsymbol{y}) \quad \text{and} \quad h(\boldsymbol{x}) = \sup_{\boldsymbol{y} \in C} f(\boldsymbol{x}, \boldsymbol{y})$$

are continuous.

http://www.springer.com/978-1-4614-5837-1

Optimization
Lange, K.
2013, XVII, 529 p., Hardcover
ISBN: 978-1-4614-5837-1