CHAPTER 4

FUNDAMENTAL THEOREMS FOR NORMED AND BANACH SPACES

This chapter contains, roughly speaking, the basis of the more advanced theory of normed and Banach spaces without which the usefulness of these spaces and their applications would be rather limited. The four important theorems in the chapter are the Hahn-Banach theorem, the uniform boundedness theorem, the open mapping theorem, and the closed graph theorem. These are the cornerstones of the theory of Banach spaces. (The first theorem holds for any normed space.)

Brief orientation about main content

1. *Hahn-Banach theorem* 4.2-1 (variants 4.3-1, 4.3-2). This is an extension theorem for linear functionals on vector spaces. It guarantees that a normed space is richly supplied with linear functionals, so that one obtains an adequate theory of dual spaces as well as a satisfactory theory of adjoint operators (Secs. 4.5, 4.6).

2. *Uniform boundedness theorem* 4.7-3 by Banach and Steinhaus. This theorem gives conditions sufficient for $(\|T_n\|)$ to be bounded, where the T_n's are bounded linear operators from a Banach into a normed space. It has various (simple and deeper) applications in analysis, for instance in connection with Fourier series (cf. 4.7-5), weak convergence (Secs. 4.8, 4.9), summability of sequences (Sec. 4.10), numerical integration (Sec. 4.11), etc.

3. *Open mapping theorem* 4.12-2. This theorem states that a bounded linear operator T from a Banach space onto a Banach space is an open mapping, that is, maps open sets onto open sets. Hence if T is bijective, T^{-1} is continuous ("*bounded inverse theorem*").

4. *Closed graph theorem* 4.13-2. This theorem gives conditions under which a closed linear operator (cf. 4.13-1) is bounded. Closed linear operators are of importance in physical and other applications.

4.1 Zorn's Lemma

We shall need Zorn's lemma in the proof of the fundamental Hahn-Banach theorem, which is an extension theorem for linear functionals and is important for reasons which we shall state when we formulate the theorem. Zorn's lemma has various applications. Two of them will be shown later in this section. The setting for the lemma is a partially ordered set:

4.1-1 Definition (Partially ordered set, chain). A *partially ordered set* is a set M on which there is defined a *partial ordering*, that is, a binary relation which is written $\leqq$ and satisfies the conditions

(PO1) $a \leqq a$ for every $a \in M$. (*Reflexivity*)

(PO2) If $a \leqq b$ and $b \leqq a$, then $a = b$. (*Antisymmetry*)

(PO3) If $a \leqq b$ and $b \leqq c$, then $a \leqq c$. (*Transitivity*)

"Partially"emphasizes that M may contain elements a and b for which neither $a \leqq b$ nor $b \leqq a$ holds. Then a and b are called *incomparable elements.* In contrast, two elements a and b are called *comparable elements* if they satisfy $a \leqq b$ or $b \leqq a$ (or both).

A *totally ordered set* or *chain* is a partially ordered set such that every two elements of the set are comparable. In other words, a chain is a partially ordered set that has no incomparable elements.

An *upper bound* of a subset W of a partially ordered set M is an element $u \in M$ such that

$$x \leqq u \qquad \text{for every } x \in W.$$

(Depending on M and W, such a u may or may not exist.) A *maximal element* of M is an $m \in M$ such that

$$m \leqq x \qquad \text{implies} \qquad m = x.$$

(Again, M may or may not have maximal elements. Note further that a maximal element need not be an upper bound.) ∎

Examples

4.1-2 Real numbers. Let M be the set of all real numbers and let $x \leqq y$ have its usual meaning. M is totally ordered. M has no maximal elements.

4.1-3 Power set. Let $\mathcal{P}(X)$ be the *power set* (set of all subsets) of a given set X and let $A \leqq B$ mean $A \subset B$, that is, A is a subset of B. Then $\mathcal{P}(X)$ is partially ordered. The only maximal element of $\mathcal{P}(X)$ is X.

4.1-4 n-tuples of numbers. Let M be the set of all ordered n-tuples $x = (\xi_1, \cdots, \xi_n)$, $y = (\eta_1, \cdots, \eta_n)$, $\cdots$ of real numbers and let $x \leqq y$ mean $\xi_j \leqq \eta_j$ for every $j = 1, \cdots, n$, where $\xi_j \leqq \eta_j$ has its usual meaning. This defines a partial ordering on M.

4.1-5 Positive integers. Let $M = \mathbf{N}$, the set of all positive integers. Let $m \leqq n$ mean that m divides n. This defines a partial ordering on $\mathbf{N}$.

Some further examples are given in the problem set. See also G. Birkhoff (1967).

Using the concepts defined in 4.1-1, we can now formulate Zorn's lemma, which we regard as an axiom.[1]

4.1-6 Zorn's lemma. *Let $M \neq \emptyset$ be a partially ordered set. Suppose that every chain $C \subset M$ has an upper bound. Then M has at least one maximal element.*

Applications

4.1-7 Hamel basis. *Every vector space $X \neq \{0\}$ has a Hamel basis.* (Cf. Sec. 2.1.)

Proof. Let M be the set of all linearly independent subsets of X. Since $X \neq \{0\}$, it has an element $x \neq 0$ and $\{x\} \in M$, so that $M \neq \emptyset$. Set inclusion defines a partial ordering on M; cf. 4.1-3. Every chain $C \subset M$ has an upper bound, namely, the union of all subsets of X which are elements of C. By Zorn's lemma, M has a maximal element B. We show that B is a Hamel basis for X. Let $Y = \operatorname{span} B$. Then Y is a subspace of X, and $Y = X$ since otherwise $B \cup \{z\}$, $z \in X$, $z \notin Y$, would be a linearly independent set containing B as a proper subset, contrary to the maximality of B. ∎

[1] The name "lemma" is for historical reasons. Zorn's lemma can be derived from the **axiom of choice,** which states that for any given set E, there exists a mapping c ("*choice function*") from the power set $\mathcal{P}(E)$ into E such that if $B \subset E$, $B \neq \emptyset$, then $c(B) \in B$. Conversely, this axiom follows from Zorn's lemma, so that Zorn's lemma and the axiom of choice can be regarded as equivalent axioms.

4.1-8 Total orthonormal set. *In every Hilbert space $H \neq \{0\}$ there exists a total orthonormal set.* (Cf. Sec. 3.6.)

Proof. Let M be the set of all orthonormal subsets of H. Since $H \neq \{0\}$, it has an element $x \neq 0$, and an orthonormal subset of H is $\{y\}$, where $y = \|x\|^{-1}x$. Hence $M \neq \emptyset$. Set inclusion defines a partial ordering on M. Every chain $C \subset M$ has an upper bound, namely, the union of all subsets of X which are elements of C. By Zorn's lemma, M has a maximal element F. We prove that F is total in H. Suppose that this is false. Then by Theorem 3.6-2 there exists a nonzero $z \in H$ such that $z \perp F$. Hence $F_1 = F \cup \{e\}$, where $e = \|z\|^{-1}z$, is orthonormal, and F is a proper subset of F_1. This contradicts the maximality of F. ∎

Problems

1. Verify the statements in Example 4.1-3.

2. Let X be the set of all real-valued functions x on the interval $[0, 1]$, and let $x \leqq y$ mean that $x(t) \leqq y(t)$ for all $t \in [0, 1]$. Show that this defines a partial ordering. Is it a total ordering? Does X have maximal elements?

3. Show that the set of all complex numbers $z = x + iy$, $w = u + iv, \cdots$ can be partially ordered by defining $z \leqq w$ to mean $x \leqq u$ and $y \leqq v$, where for real numbers, $\leqq$ has its usual meaning.

4. Find all maximal elements of M with respect to the partial ordering in Example 4.1-5, where M is (*a*) $\{2, 3, 4, 8\}$, (*b*) the set of all prime numbers.

5. Prove that a finite partially ordered set A has at least one maximal element.

6. **(Least element, greatest element)** Show that a partially ordered set M can have at most one element a such that $a \leqq x$ for all $x \in M$ and at most one element b such that $x \leqq b$ for all $x \in M$. [If such an a (or b) exists, it is called the *least element* (*greatest element*, respectively) of M.]

7. **(Lower bound)** A *lower bound* of a subset $A \neq \emptyset$ of a partially ordered set M is an $x \in M$ such that $x \leqq y$ for all $y \in A$. Find upper and lower bounds of the subset $A = \{4, 6\}$ in Example 4.1-5.

8. A *greatest lower bound* of a subset $A \neq \emptyset$ of a partially ordered set M is a lower bound x of A such that $l \leqq x$ for any lower bound l of A; we write $x = \text{g.l.b.}\, A = \inf A$. Similarly, a *least upper bound* y of A, written $y = \text{l.u.b.}\, A = \sup A$, is an upper bound y of A such that $y \leqq u$ for any upper bound u of A. (*a*) If A has a g.l.b., show that it is unique. (*b*) What are g.l.b. $\{A, B\}$ and l.u.b. $\{A, B\}$ in Example 4.1-3?

9. **(Lattice)** A *lattice* is a partially ordered set M such that any two elements x, y of M have a g.l.b. (written $x \wedge y$) and a l.u.b. (written $x \vee y$). Show that the partially ordered set in Example 4.1-3 is a lattice, where $A \wedge B = A \cap B$ and $A \vee B = A \cup B$.

10. A *minimal element* of a partially ordered set M is an $x \in M$ such that $y \leqq x$ implies $y = x$. Find all minimal elements in Prob. 4(*a*).

4.2 Hahn-Banach Theorem

The Hahn-Banach theorem is an extension theorem for linear functionals. We shall see in the next section that the theorem guarantees that a normed space is richly supplied with bounded linear functionals and makes possible an adequate theory of dual spaces, which is an essential part of the general theory of normed spaces. In this way the Hahn-Banach theorem becomes one of the most important theorems in connection with bounded linear operators. Furthermore, our discussion will show that the theorem also characterizes the extent to which values of a linear functional can be preassigned. The theorem was discovered by H. Hahn (1927), rediscovered in its present more general form (Theorem 4.2-1) by S. Banach (1929) and generalized to complex vector spaces (Theorem 4.3-1) by H. F. Bohnenblust and A. Sobczyk (1938); cf. the references in Appendix 3.

Generally speaking, in an *extension problem* one considers a mathematical object (for example, a mapping) defined on a subset Z of a given set X and one wants to extend the object from Z to the entire set X in such a way that certain basic properties of the object continue to hold for the extended object.

In the Hahn-Banach theorem, the object to be extended is a linear functional f which is defined on a subspace Z of a vector space X and has a certain boundedness property which will be formulated in terms of a **sublinear functional.** By definition, this is a real-valued functional

p on a vector space X which is **subadditive,** that is,

(1) $$p(x+y) \leqq p(x)+p(y) \qquad \text{for all } x, y \in X,$$

and **positive-homogeneous,** that is,

(2) $$p(\alpha x) = \alpha p(x) \qquad \text{for all } \alpha \geqq 0 \text{ in } \mathbf{R} \text{ and } x \in X.$$

(Note that the *norm* on a normed space is such a functional.)

We shall assume that the functional f to be extended is majorized on Z by such a functional p defined on X, and we shall extend f from Z to X without losing the linearity and the majorization, so that the extended functional $\tilde{f}$ on X is still linear and still majorized by p. This is the crux of the theorem. X will be real; a generalization of the theorem that includes complex vector spaces follows in the next section.

4.2-1 Hahn-Banach Theorem (Extension of linear functionals). *Let X be a real vector space and p a sublinear functional on X. Furthermore, let f be a linear functional which is defined on a subspace Z of X and satisfies*

(3) $$f(x) \leqq p(x) \qquad \text{for all } x \in Z.$$

Then f has a linear extension $\tilde{f}$ from Z to X satisfying

(3*) $$\tilde{f}(x) \leqq p(x) \qquad \text{for all } x \in X,$$

that is, $\tilde{f}$ is a linear functional on X, satisfies (3) on X and $\tilde{f}(x) = f(x)$ for every $x \in Z$.*

Proof. Proceeding stepwise, we shall prove:

(*a*) The set E of all linear extensions g of f satisfying $g(x) \leqq p(x)$ on their domain $\mathcal{D}(g)$ can be partially ordered and Zorn's lemma yields a maximal element $\tilde{f}$ of E.

(*b*) $\tilde{f}$ is defined on the entire space X.

(*c*) An auxiliary relation which was used in (*b*).

We start with part

(a) Let E be the set of all linear extensions g of f which satisfy the condition

$$g(x) \leqq p(x) \qquad \text{for all } x \in \mathcal{D}(g)$$

Clearly, $E \neq \emptyset$ since $f \in E$. On E we can define a partial ordering by

$$g \leqq h \qquad \text{meaning} \qquad h \text{ is an extension of } g,$$

that is, by definition, $\mathfrak{D}(h) \supset \mathfrak{D}(g)$ and $h(x) = g(x)$ for every $x \in \mathfrak{D}(g)$.

For any chain $C \subset E$ we now define $\hat{g}$ by

$$\hat{g}(x) = g(x) \qquad \text{if } x \in \mathfrak{D}(g) \qquad (g \in C).$$

$\hat{g}$ is a linear functional, the domain being

$$\mathfrak{D}(\hat{g}) = \bigcup_{g \in C} \mathfrak{D}(g),$$

which is a vector space since C is a chain. The definition of $\hat{g}$ is unambiguous. Indeed, for an $x \in \mathfrak{D}(g_1) \cap \mathfrak{D}(g_2)$ with $g_1, g_2 \in C$ we have $g_1(x) = g_2(x)$ since C is a chain, so that $g_1 \leqq g_2$ or $g_2 \leqq g_1$. Clearly, $g \leqq \hat{g}$ for all $g \in C$. Hence $\hat{g}$ is an upper bound of C. Since $C \subset E$ was arbitrary, Zorn's lemma thus implies that E has a maximal element $\tilde{f}$. By the definition of E, this is a linear extension of f which satisfies

$$\tilde{f}(x) \leqq p(x) \qquad x \in \mathfrak{D}(\tilde{f}). \tag{4}$$

(b) We now show that $\mathfrak{D}(\tilde{f})$ is all of X. Suppose that this is false. Then we can choose a $y_1 \in X - \mathfrak{D}(\tilde{f})$ and consider the subspace Y_1 of X spanned by $\mathfrak{D}(\tilde{f})$ and y_1. Note that $y_1 \neq 0$ since $0 \in \mathfrak{D}(\tilde{f})$. Any $x \in Y_1$ can be written

$$x = y + \alpha y_1 \qquad y \in \mathfrak{D}(\tilde{f}).$$

This representation is unique. In fact, $y + \alpha y_1 = \tilde{y} + \beta y_1$ with $\tilde{y} \in \mathfrak{D}(\tilde{f})$ implies $y - \tilde{y} = (\beta - \alpha) y_1$, where $y - \tilde{y} \in \mathfrak{D}(\tilde{f})$ whereas $y_1 \notin \mathfrak{D}(\tilde{f})$, so that the only solution is $y - \tilde{y} = 0$ and $\beta - \alpha = 0$. This means uniqueness.

A functional g_1 on Y_1 is defined by

$$g_1(y + \alpha y_1) = \tilde{f}(y) + \alpha c \tag{5}$$

where c is any real constant. It is not difficult to see that g_1 is linear. Furthermore, for $\alpha = 0$ we have $g_1(y) = \tilde{f}(y)$. Hence g_1 is a *proper extension* of $\tilde{f}$, that is, an extension such that $\mathfrak{D}(\tilde{f})$ is a proper subset of

$\mathfrak{D}(g_1)$. Consequently, if we can prove that $g_1 \in E$ by showing that

$$g_1(x) \leqq p(x) \qquad \text{for all } x \in \mathfrak{D}(g_1), \tag{6}$$

this will contradict the maximality of $\tilde{f}$, so that $\mathfrak{D}(\tilde{f}) \neq X$ is false and $\mathfrak{D}(\tilde{f}) = X$ is true.

(c) Accordingly, we must finally show that g_1 with a suitable c in (5) satisfies (6).

We consider any y and z in $\mathfrak{D}(\tilde{f})$. From (4) and (1) we obtain

$$\begin{aligned} \tilde{f}(y) - \tilde{f}(z) = \tilde{f}(y-z) &\leqq p(y-z) \\ &= p(y + y_1 - y_1 - z) \\ &\leqq p(y + y_1) + p(-y_1 - z). \end{aligned}$$

Taking the last term to the left and the term $\tilde{f}(y)$ to the right, we have

$$-p(-y_1 - z) - \tilde{f}(z) \leqq p(y + y_1) - \tilde{f}(y), \tag{7}$$

where y_1 is fixed. Since y does not appear on the left and z not on the right, the inequality continues to hold if we take the supremum over $z \in \mathfrak{D}(\tilde{f})$ on the left (call it m_0) and the infimum over $y \in \mathfrak{D}(\tilde{f})$ on the right, call it m_1. Then $m_0 \leqq m_1$ and for a c with $m_0 \leqq c \leqq m_1$ we have from (7)

$$-p(-y_1 - z) - \tilde{f}(z) \leqq c \qquad \text{for all } z \in \mathfrak{D}(\tilde{f}) \tag{8a}$$

$$c \leqq p(y + y_1) - \tilde{f}(y) \qquad \text{for all } y \in \mathfrak{D}(\tilde{f}). \tag{8b}$$

We prove (6) first for negative α in (5) and then for positive α. For $\alpha < 0$ we use (8a) with z replaced by $\alpha^{-1}y$, that is,

$$-p\left(-y_1 - \frac{1}{\alpha}y\right) - \tilde{f}\left(\frac{1}{\alpha}y\right) \leqq c.$$

Multiplication by $-\alpha > 0$ gives

$$\alpha p\left(-y_1 - \frac{1}{\alpha}y\right) + \tilde{f}(y) \leqq -\alpha c.$$

From this and (5), using $y+\alpha y_1 = x$ (see above), we obtain the desired inequality

$$g_1(x) = \tilde{f}(y) + \alpha c \leqq -\alpha p\left(-y_1 - \frac{1}{\alpha} y\right) = p(\alpha y_1 + y) = p(x).$$

For $\alpha = 0$ we have $x \in \mathfrak{D}(\tilde{f})$ and nothing to prove. For $\alpha > 0$ we use (8b) with y replaced by $\alpha^{-1}y$ to get

$$c \leqq p\left(\frac{1}{\alpha} y + y_1\right) - \tilde{f}\left(\frac{1}{\alpha} y\right).$$

Multiplication by $\alpha > 0$ gives

$$\alpha c \leqq \alpha p\left(\frac{1}{\alpha} y + y_1\right) - \tilde{f}(y) = p(x) - \tilde{f}(y).$$

From this and (5),

$$g_1(x) = \tilde{f}(y) + \alpha c \leqq p(x). \qquad \blacksquare$$

Could we get away without Zorn's lemma? This question is of interest, in particular since the lemma does not give a method of construction. If in (5) we take f instead of $\tilde{f}$, we obtain for each real c a linear extension g_1 of f to the subspace Z_1 spanned by $\mathfrak{D}(f) \cup \{y_1\}$, and we can choose c so that $g_1(x) \leqq p(x)$ for all $x \in Z_1$, as may be seen from part (c) of the proof with $\tilde{f}$ replaced by f. If $X = Z_1$, we are done. If $X \neq Z_1$, we may take a $y_2 \in X - Z_1$ and repeat the process to extend f to Z_2 spanned by Z_1 and y_2, etc. This gives a sequence of subspaces Z_j each containing the preceding, and such that f can be extended linearly from one to the next and the extension g_j satisfies $g_j(x) \leqq p(x)$ for all $x \in Z_j$. If

$$X = \bigcup_{j=1}^{n} Z_j,$$

we are done after n steps, and if

$$X = \bigcup_{j=1}^{\infty} Z_j,$$

we can use ordinary induction. However, if X has no such representation, we do need Zorn's lemma in the proof presented here.

Of course, for special spaces the whole situation may become simpler. Hilbert spaces are of this type, because of the Riesz representation 3.8-1. We shall discuss this fact in the next section.

Problems

1. Show that the absolute value of a linear functional has the properties expressed in (1) and (2).

2. Show that a norm on a vector space X is a sublinear functional on X.

3. Show that $p(x) = \overline{\lim_{n\to\infty}} \xi_n$, where $x = (\xi_n) \in l^\infty$, ξ_n real, defines a sublinear functional on l^∞.

4. Show that a sublinear functional p satisfies $p(0) = 0$ and $p(-x) \geqq -p(x)$.

5. **(Convex set)** If p is a sublinear functional on a vector space X, show that $M = \{x \mid p(x) \leqq \gamma,\ \gamma > 0 \text{ fixed}\}$, is a convex set. (Cf. Sec. 3.3.)

6. If a subadditive functional p on a normed space X is continuous at 0 and $p(0) = 0$, show that p is continuous for all $x \in X$.

7. If p_1 and p_2 are sublinear functionals on a vector space X and c_1 and c_2 are positive constants, show that $p = c_1 p_1 + c_2 p_2$ is sublinear on X.

8. If a subadditive functional defined on a normed space X is nonnegative outside a sphere $\{x \mid \|x\| = r\}$, show that it is nonnegative for all $x \in X$.

9. Let p be a sublinear functional on a real vector space X. Let f be defined on $Z = \{x \in X \mid x = \alpha x_0,\ \alpha \in \mathbf{R}\}$ by $f(x) = \alpha p(x_0)$ with fixed $x_0 \in X$. Show that f is a linear functional on Z satisfying $f(x) \leqq p(x)$.

10. If p is a sublinear functional on a real vector space X, show that there exists a linear functional $\tilde{f}$ on X such that $-p(-x) \leqq \tilde{f}(x) \leqq p(x)$.

4.3 Hahn-Banach Theorem for Complex Vector Spaces and Normed Spaces

The Hahn-Banach theorem 4.2-1 concerns *real* vector spaces. A generalization that includes complex vector spaces was obtained by H. F. Bohnenblust and A. Sobczyk (1938):

4.3-1 Hahn-Banach Theorem (Generalized). *Let X be a real or complex vector space and p a real-valued functional on X which is subadditive, that is, for all $x, y \in X$,*

$$p(x+y) \leqq p(x)+p(y) \tag{1}$$

(as in Theorem 4.2-1*), and for every scalar α satisfies*

$$p(\alpha x) = |\alpha|\, p(x). \tag{2}$$

Furthermore, let f be a linear functional which is defined on a subspace Z of X and satisfies

$$|f(x)| \leqq p(x) \qquad \text{for all } x \in Z. \tag{3}$$

Then f has a linear extension $\tilde{f}$ from Z to X satisfying

$$|\tilde{f}(x)| \leqq p(x) \qquad \text{for all } x \in X. \tag{3*}$$

Proof. **(a)** *Real vector space.* If X is real, the situation is simple. Then (3) implies $f(x) \leqq p(x)$ for all $x \in Z$. Hence by the Hahn-Banach theorem 4.2-1 there is a linear extension $\tilde{f}$ from Z to X such that

$$\tilde{f}(x) \leqq p(x) \qquad \text{for all } x \in X. \tag{4}$$

From this and (2) we obtain

$$-\tilde{f}(x) = \tilde{f}(-x) \leqq p(-x) = |-1|\, p(x) = p(x),$$

that is, $\tilde{f}(x) \geqq -p(x)$. Together with (4) this proves (3*).

(b) *Complex vector space.* Let X be complex. Then Z is a complex vector space, too. Hence f is complex-valued, and we can write

$$f(x) = f_1(x) + if_2(x) \qquad x \in Z$$

where f_1 and f_2 are real-valued. For a moment we regard X and Z as real vector spaces and denote them by X_r and Z_r, respectively; this simply means that we restrict multiplication by scalars to real numbers (instead of complex numbers). Since f is linear on Z and f_1 and f_2 are

real-valued, f_1 and f_2 are linear functionals on Z_r. Also $f_1(x) \leqq |f(x)|$ because the real part of a complex number cannot exceed the absolute value. Hence by (3),

$$f_1(x) \leqq p(x) \qquad \text{for all } x \in Z_r.$$

By the Hahn-Banach theorem 4.2-1 there is a linear extension $\tilde{f}_1$ of f_1 from Z_r to X_r such that

$$\tilde{f}_1(x) \leqq p(x) \qquad \text{for all } x \in X_r. \tag{5}$$

This takes care of f_1 and we now turn to f_2. Returning to Z and using $f = f_1 + if_2$, we have for every $x \in Z$

$$i[f_1(x) + if_2(x)] = if(x) = f(ix) = f_1(ix) + if_2(ix).$$

The real parts on both sides must be equal:

$$f_2(x) = -f_1(ix) \qquad x \in Z. \tag{6}$$

Hence if for all $x \in X$ we set

$$\tilde{f}(x) = \tilde{f}_1(x) - i\tilde{f}_1(ix) \qquad x \in X, \tag{7}$$

we see from (6) that $\tilde{f}(x) = f(x)$ on Z. This shows that $\tilde{f}$ is an extension of f from Z to X. Our remaining task is to prove that

(*i*) $\tilde{f}$ is a linear functional on the *complex* vector space X,

(*ii*) $\tilde{f}$ satisfies (3*) on X.

That (*i*) holds can be seen from the following calculation which uses (7) and the linearity of $\tilde{f}_1$ on the real vector space X_r; here $a + ib$ with real a and b is any complex scalar:

$$\begin{aligned}
\tilde{f}((a+ib)x) &= \tilde{f}_1(ax + ibx) - i\tilde{f}_1(iax - bx) \\
&= a\tilde{f}_1(x) + b\tilde{f}_1(ix) - i[a\tilde{f}_1(ix) - b\tilde{f}_1(x)] \\
&= (a+ib)[\tilde{f}_1(x) - i\tilde{f}_1(ix)] \\
&= (a+ib)\tilde{f}(x).
\end{aligned}$$

We prove (*ii*). For any x such that $\tilde{f}(x) = 0$ this holds since $p(x) \geqq 0$ by (1) and (2); cf. also Prob. 1. Let x be such that $\tilde{f}(x) \neq 0$,

Then we can write, using the polar form of complex quantities,

$$\tilde{f}(x) = |\tilde{f}(x)|e^{i\theta}, \qquad \text{thus} \qquad |\tilde{f}(x)| = \tilde{f}(x)e^{-i\theta} = \tilde{f}(e^{-i\theta}x).$$

Since $|\tilde{f}(x)|$ is real, the last expression is real and thus equal to its real part. Hence by (2),

$$|\tilde{f}(x)| = \tilde{f}(e^{-i\theta}x) = \tilde{f}_1(e^{-i\theta}x) \leqq p(e^{-i\theta}x) = |e^{-i\theta}|\,p(x) = p(x).$$

This completes the proof. ∎

Although the Hahn-Banach theorem says nothing directly about continuity, a principal application of the theorem deals with bounded linear functionals. This brings us back to normed spaces, which is our main concern. In fact, Theorem 4.3-1 implies the basic

4.3-2 Hahn-Banach Theorem (Normed spaces). *Let f be a bounded linear functional on a subspace Z of a normed space X. Then there exists a bounded linear functional $\tilde{f}$ on X which is an extension of f to X and has the same norm,*

$$\|\tilde{f}\|_X = \|f\|_Z \tag{8}$$

where

$$\|\tilde{f}\|_X = \sup_{\substack{x \in X \\ \|x\|=1}} |\tilde{f}(x)|, \qquad \|f\|_Z = \sup_{\substack{x \in Z \\ \|x\|=1}} |f(x)|$$

(*and* $\|f\|_Z = 0$ *in the trivial case* $Z = \{0\}$).

Proof. If $Z = \{0\}$, then $f = 0$, and the extension is $\tilde{f} = 0$. Let $Z \neq \{0\}$. We want to use Theorem 4.3-1. Hence we must first discover a suitable p. For all $x \in Z$ we have

$$|f(x)| \leqq \|f\|_Z \|x\|.$$

This is of the form (3), where

$$p(x) = \|f\|_Z \|x\|. \tag{9}$$

We see that p is defined on all of X. Furthermore, p satisfies (1) on X since by the triangle inequality,

$$p(x+y) = \|f\|_Z \|x+y\| \leqq \|f\|_Z (\|x\| + \|y\|) = p(x) + p(y).$$

p also satisfies (2) on X because

$$p(\alpha x) = \|f\|_Z \|\alpha x\| = |\alpha| \, \|f\|_Z \|x\| = |\alpha| \, p(x).$$

Hence we can now apply Theorem 4.3-1 and conclude that there exists a linear functional $\tilde{f}$ on X which is an extension of f and satisfies

$$|\tilde{f}(x)| \leqq p(x) = \|f\|_Z \|x\| \qquad x \in X.$$

Taking the supremum over all $x \in X$ of norm 1, we obtain the inequality

$$\|\tilde{f}\|_X = \sup_{\substack{x \in X \\ \|x\|=1}} |\tilde{f}(x)| \leqq \|f\|_Z.$$

Since under an extension the norm cannot decrease, we also have $\|\tilde{f}\|_X \geqq \|f\|_Z$. Together we obtain (8) and the theorem is proved. ∎

In special cases the situation may become very simple. Hilbert spaces are of this type. Indeed, if Z is a closed subspace of a Hilbert space $X = H$, then f has a Riesz representation 3.8-1, say,

$$f(x) = \langle x, z \rangle \qquad z \in Z$$

where $\|z\| = \|f\|$. Of course, since the inner product is defined on all of H, this gives at once a linear extension $\tilde{f}$ of f from Z to H, and $\tilde{f}$ has the same norm as f because $\|\tilde{f}\| = \|z\| = \|f\|$ by Theorem 3.8-1. Hence in this case the extension is immediate.

From Theorem 4.3-2 we shall now derive another useful result which, roughly speaking, shows that the dual space X' of a normed space X consists of sufficiently many bounded linear functionals to distinguish between the points of X. This will become essential in connection with adjoint operators (Sec. 4.5) and so-called weak convergence (Sec. 4.8).

4.3-3 Theorem (Bounded linear functionals). *Let X be a normed space and let $x_0 \neq 0$ be any element of X. Then there exists a bounded linear functional $\tilde{f}$ on X such that*

$$\|\tilde{f}\| = 1, \qquad \tilde{f}(x_0) = \|x_0\|.$$

Proof. We consider the subspace Z of X consisting of all elements $x = \alpha x_0$ where α is a scalar. On Z we define a linear functional f by

$$f(x) = f(\alpha x_0) = \alpha \|x_0\|. \tag{10}$$

f is bounded and has norm $\|f\| = 1$ because

$$|f(x)| = |f(\alpha x_0)| = |\alpha| \, \|x_0\| = \|\alpha x_0\| = \|x\|.$$

Theorem 4.3-2 implies that f has a linear extension $\tilde{f}$ from Z to X, of norm $\|\tilde{f}\| = \|f\| = 1$. From (10) we see that $\tilde{f}(x_0) = f(x_0) = \|x_0\|$. ∎

4.3-4 Corollary (Norm, zero vector). *For every x in a normed space X we have*

$$\|x\| = \sup_{\substack{f \in X' \\ f \neq 0}} \frac{|f(x)|}{\|f\|}. \tag{11}$$

Hence if x_0 is such that $f(x_0) = 0$ for all $f \in X'$, then $x_0 = 0$.

Proof. From Theorem 4.3-3 we have, writing x for x_0,

$$\sup_{\substack{f \in X' \\ f \neq 0}} \frac{|f(x)|}{\|f\|} \geqq \frac{|\tilde{f}(x)|}{\|\tilde{f}\|} = \frac{\|x\|}{1} = \|x\|,$$

and from $|f(x)| \leqq \|f\| \, \|x\|$ we obtain

$$\sup_{\substack{f \in X' \\ f \neq 0}} \frac{|f(x)|}{\|f\|} \leqq \|x\|.$$

∎

Problems

1. **(Seminorm)** Show that (1) and (2) imply $p(0) = 0$ and $p(x) \geqq 0$, so that p is a seminorm (cf. Prob. 12, Sec. 2.3).

2. Show that (1) and (2) imply $|p(x) - p(y)| \leqq p(x - y)$.

3. It was shown that $\tilde{f}$ defined by (7) is a linear functional on the *complex* vector space X. Show that for this purpose it suffices to prove that $\tilde{f}(ix) = i\tilde{f}(x)$.

4. Let p be defined on a vector space X and satisfy (1) and (2). Show that for any given $x_0 \in X$ there is a linear functional $\tilde{f}$ on X such that $\tilde{f}(x_0) = p(x_0)$ and $|\tilde{f}(x)| \leqq p(x)$ for all $x \in X$.

5. If X in Theorem 4.3-1 is a normed space and $p(x) \leqq k\,\|x\|$ for some $k > 0$, show that $\|\tilde{f}\| \leqq k$.

6. To illustrate Theorem 4.3-2, consider a functional f on the Euclidean plane $\mathbf{R}^2$ defined by $f(x) = \alpha_1\xi_1 + \alpha_2\xi_2$, $x = (\xi_1, \xi_2)$, its linear extensions $\tilde{f}$ to $\mathbf{R}^3$ and the corresponding norms.

7. Give another proof of Theorem 4.3-3 in the case of a Hilbert space.

8. Let X be a normed space and X' its dual space. If $X \neq \{0\}$, show that X' cannot be $\{0\}$.

9. Show that for a separable normed space X, Theorem 4.3-2 can be proved directly, without the use of Zorn's lemma (which was used indirectly, namely, in the proof of Theorem 4.2-1).

10. Obtain the second statement in 4.3-4 directly from 4.3-3.

11. If $f(x) = f(y)$ for every bounded linear functional f on a normed space X, show that $x = y$.

12. To illustrate Theorem 4.3-3, let X be the Euclidean plane $\mathbf{R}^2$ and find the functional $\tilde{f}$.

13. Show that under the assumptions of Theorem 4.3-3 there is a bounded linear functional $\hat{f}$ on X such that $\|\hat{f}\| = \|x_0\|^{-1}$ and $\hat{f}(x_0) = 1$.

14. **(Hyperplane)** Show that for any sphere $S(0; r)$ in a normed space X and any point $x_0 \in S(0; r)$ there is a hyperplane $H_0 \ni x_0$ such that the ball $\tilde{B}(0; r)$ lies entirely in one of the two half spaces determined by

H_0. (Cf. Probs. 12, 15, Sec. 2.8.) A simple illustration is shown in Fig. 39.

15. If x_0 in a normed space X is such that $|f(x_0)| \leqq c$ for all $f \in X'$ of norm 1, show that $\|x_0\| \leqq c$.

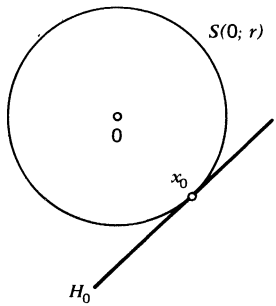


Fig. 39. Illustration of Prob. 14 in the case of the Euclidean plane $\mathbf{R}^2$

4.4 Application to Bounded Linear Functionals on $C[a, b]$

The Hahn-Banach theorem 4.3-2 has many important applications. One of them was considered in the preceding section. Another one will be presented in this section.[2] In fact, we shall use Theorem 4.3-2 for obtaining a general representation formula for bounded linear functionals on $C[a, b]$, where $[a, b]$ is a fixed compact interval. The significance of such general representations of functionals on special spaces was explained at the end of Sec. 2.10. In the present case the representation will be in terms of a Riemann-Stieltjes integral. So let us recall the definition and a few properties of this integral, which is a generalization of the familiar Riemann integral. We begin with the following concept.

A function w defined on $[a, b]$ is said to be of **bounded variation** on $[a, b]$ if its *total variation* $\mathrm{Var}(w)$ on $[a, b]$ is finite, where

$$\mathrm{Var}(w) = \sup \sum_{j=1}^{n} |w(t_j) - w(t_{j-1})|, \tag{1}$$

[2] This section is optional. It will be needed only once (namely, in Sec. 9.9).

the supremum being taken over all *partitions*

$$a = t_0 < t_1 < \cdots < t_n = b \tag{2}$$

of the interval $[a, b]$; here, $n \in \mathbf{N}$ is arbitrary and so is the choice of values $t_1, \ldots, t_{n-1}$ in $[a, b]$ which, however, must satisfy (2).

Obviously, all functions of bounded variation on $[a, b]$ form a vector space. A norm on this space is given by

$$\|w\| = |w(a)| + \mathrm{Var}(w). \tag{3}$$

The normed space thus defined is denoted by $BV[a, b]$, where BV suggests "bounded variation."

We now obtain the concept of a Riemann-Stieltjes integral as follows. Let $x \in C[a, b]$ and $w \in BV[a, b]$. Let P_n be any partition of $[a, b]$ given by (2) and denote by $\eta(P_n)$ the length of a largest interval $[t_{j-1}, t_j]$, that is,

$$\eta(P_n) = \max(t_1 - t_0, \cdots, t_n - t_{n-1}).$$

For every partition P_n of $[a, b]$ we consider the sum

$$s(P_n) = \sum_{j=1}^{n} x(t_j)[w(t_j) - w(t_{j-1})]. \tag{4}$$

There exists a number $\mathcal{J}$ with the property that for every $\varepsilon > 0$ there is a $\delta > 0$ such that

$$\eta(P_n) < \delta \tag{5}$$

implies

$$|\mathcal{J} - s(P_n)| < \varepsilon. \tag{6}$$

$\mathcal{J}$ is called the **Riemann-Stieltjes integral** of x over $[a, b]$ with respect to w and is denoted by

$$\int_a^b x(t)\, dw(t). \tag{7}$$

Hence we can obtain (7) as the limit of the sums (4) for a sequence (P_n) of partitions of $[a, b]$ satisfying $\eta(P_n) \longrightarrow 0$ as $n \longrightarrow \infty$; cf. (5).

Note that for $w(t)=t$, the integral (7) is the familiar Riemann integral of x over $[a, b]$.

Also, if x is continuous on $[a, b]$ and w has a derivative which is integrable on $[a, b]$, then

$$\int_a^b x(t)\,dw(t)=\int_a^b x(t)w'(t)\,dt \tag{8}$$

where the prime denotes differentiation with respect to t.

The integral (7) depends linearly on $x\in C[a, b]$, that is, for all $x_1, x_2\in C[a, b]$ and scalars α and β we have

$$\int_a^b [\alpha x_1(t)+\beta x_2(t)]\,dw(t)=\alpha\int_a^b x_1(t)\,dw(t)+\beta\int_a^b x_2(t)\,dw(t).$$

The integral also depends linearly on $w\in BV[a, b]$; that is, for all $w_1, w_2\in BV[a, b]$ and scalars γ and δ we have

$$\int_a^b x(t)\,d(\gamma w_1+\delta w_2)(t)=\gamma\int_a^b x(t)\,dw_1(t)+\delta\int_a^b x(t)\,dw_2(t).$$

We shall also need the inequality

$$\left|\int_a^b x(t)\,dw(t)\right|\leqq\max_{t\in J}|x(t)|\,\mathrm{Var}(w), \tag{9}$$

where $J=[a, b]$. We note that this generalizes a familiar formula from calculus. In fact, if $w(t)=t$, then $\mathrm{Var}(w)=b-a$ and (9) takes the form

$$\left|\int_a^b x(t)\,dt\right|\leqq\max_{t\in J}|x(t)|\,(b-a).$$

The representation theorem for bounded linear functionals on $C[a, b]$ by F. Riesz (1909) can now be stated as follows.

4.4-1 Riesz's Theorem (Functionals on $C[a, b]$). *Every bounded linear functional f on $C[a, b]$ can be represented by a Riemann-Stieltjes integral*

$$f(x)=\int_a^b x(t)\,dw(t) \tag{10}$$

where w is of bounded variation on $[a, b]$ and has the total variation

(11) $$\text{Var}(w) = \|f\|.$$

Proof. From the Hahn-Banach theorem 4.3-2 for normed spaces we see that f has an extension $\tilde{f}$ from $C[a, b]$ to the normed space $B[a, b]$ consisting of all bounded functions on $[a, b]$ with norm defined by

$$\|x\| = \sup_{t \in J} |x(t)| \qquad J = [a, b].$$

Furthermore, by that theorem, the linear functional $\tilde{f}$ is bounded and has the same norm as f, that is,

$$\|\tilde{f}\| = \|f\|.$$

We define the function w needed in (10). For this purpose we consider the function x_t shown in Fig. 40. This function is defined on $[a, b]$ and, by definition, is 1 on $[a, t]$ and 0 otherwise. Clearly, $x_t \in B[a, b]$. We mention that x_t is called the *characteristic function* of the interval $[a, t]$. Using x_t and the functional $\tilde{f}$, we define w on $[a, b]$ by

$$w(a) = 0 \qquad w(t) = \tilde{f}(x_t), \qquad t \in (a, b].$$

We show that this function w is of bounded variation and $\text{Var}(w) \leqq \|f\|$.

For a complex quantity we can use the polar form. In fact, setting $\theta = \arg \zeta$, we may write

$$\zeta = |\zeta| \, e(\zeta) \qquad \text{where} \qquad e(\zeta) = \begin{cases} 1 & \text{if } \zeta = 0 \\ e^{i\theta} & \text{if } \zeta \neq 0. \end{cases}$$

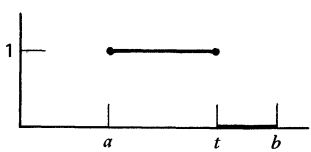


Fig. 40. The function x_t

We see that if $\zeta \neq 0$, then $|\zeta| = \zeta/e^{i\theta} = \zeta e^{-i\theta}$. Hence for any ζ, zero or not, we have

$$|\zeta| = \zeta\,\overline{e(\zeta)}, \tag{12}$$

where the bar indicates complex conjugation, as usual. For simplifying our subsequent formulas we also write

$$\varepsilon_j = \overline{e(w(t_j) - w(t_{j-1}))}$$

and $x_{t_j} = x_j$. In this way we avoid subscripts of subscripts. Then, by (12), for any partition (2) we obtain

$$\begin{aligned}
\sum_{j=1}^{n} |w(t_j) - w(t_{j-1})| &= |\tilde{f}(x_1)| + \sum_{j=2}^{n} |\tilde{f}(x_j) - \tilde{f}(x_{j-1})| \\
&= \varepsilon_1 \tilde{f}(x_1) + \sum_{j=2}^{n} \varepsilon_j [\tilde{f}(x_j) - \tilde{f}(x_{j-1})] \\
&= \tilde{f}\left(\varepsilon_1 x_1 + \sum_{j=2}^{n} \varepsilon_j [x_j - x_{j-1}]\right) \\
&\leqq \|\tilde{f}\| \left\| \varepsilon_1 x_1 + \sum_{j=2}^{n} \varepsilon_j [x_j - x_{j-1}] \right\|.
\end{aligned}$$

On the right, $\|\tilde{f}\| = \|f\|$ (see before) and the other factor $\|\cdots\|$ equals 1 because $|\varepsilon_j| = 1$ and from the definition of the x_j's we see that for each $t \in [a, b]$ only one of the terms x_1, $x_2 - x_1$, $\cdots$ is not zero (and its norm is 1). On the left we can now take the supremum over all partitions of $[a, b]$. Then we have

$$\operatorname{Var}(w) \leqq \|f\|. \tag{13}$$

Hence w is of bounded variation on $[a, b]$.

We prove (10), where $x \in C[a, b]$. For every partition P_n of the form (2) we define a function, which we denote simply by z_n [instead of $z(P_n)$ or z_{P_n}, say], keeping in mind that z_n depends on P_n, not merely on n. The defining formula is

$$z_n = x(t_0)x_1 + \sum_{j=2}^{n} x(t_{j-1})[x_j - x_{j-1}]. \tag{14}$$

Then $z_n \in B[a, b]$. By the definition of w,

$$\begin{aligned} \tilde{f}(z_n) &= x(t_0)\tilde{f}(x_1) + \sum_{j=2}^{n} x(t_{j-1})[\tilde{f}(x_j) - \tilde{f}(x_{j-1})] \\ &= x(t_0)w(t_1) + \sum_{j=2}^{n} x(t_{j-1})[w(t_j) - w(t_{j-1})] \\ &= \sum_{j=1}^{n} x(t_{j-1})[w(t_j) - w(t_{j-1})], \end{aligned} \tag{15}$$

where the last equality follows from $w(t_0) = w(a) = 0$. We now choose any sequence (P_n) of partitions of $[a, b]$ such that $\eta(P_n) \longrightarrow 0$; cf. (5). (Note that the t_j in (15) depend on P_n, a fact which we keep in mind without expressing it by a bulkier notation such as $t_{j,n}$.) As $n \longrightarrow \infty$, the sum on the right-hand side of (15) approaches the integral in (10), and (10) follows, provided $\tilde{f}(z_n) \longrightarrow \tilde{f}(x)$, which equals $f(x)$ since $x \in C[a, b]$.

We prove that $\tilde{f}(z_n) \longrightarrow \tilde{f}(x)$. Remembering the definition of x_t (see Fig. 40), we see that (14) yields $z_n(a) = x(a) \cdot 1$ since the sum in (14) is zero at $t = a$. Hence $z_n(a) - x(a) = 0$. Furthermore, by (14), if $t_{j-1} < t \leqq t_j$, then we obtain $z_n(t) = x(t_{j-1}) \cdot 1$; see Fig. 40. It follows that for those t,

$$|z_n(t) - x(t)| = |x(t_{j-1}) - x(t)|.$$

Consequently, if $\eta(P_n) \longrightarrow 0$, then $\|z_n - x\| \longrightarrow 0$ because x is continuous on $[a, b]$, hence uniformly continuous on $[a, b]$, since $[a, b]$ is compact. The continuity of $\tilde{f}$ now implies that $\tilde{f}(z_n) \longrightarrow \tilde{f}(x)$, and $\tilde{f}(x) = f(x)$, so that (10) is established.

We finally prove (11). From (10) and (9) we have

$$|f(x)| \leqq \max_{t \in J} |x(t)| \operatorname{Var}(w) = \|x\| \operatorname{Var}(w).$$

Taking the supremum over all $x \in C[a, b]$ of norm one, we obtain $\|f\| \leqq \operatorname{Var}(w)$. Together with (13) this yields (11). ∎

We note that w in the theorem is not unique, but can be made unique by imposing the normalizing conditions that w be zero at a and

continuous from the right:

$$w(a) = 0, \qquad w(t+0) = w(t) \qquad (a < t < b).$$

For details, see A. E. Taylor (1958), pp. 197–200. Cf. also F. Riesz and B. Sz.-Nagy (1955), p. 111.

It is interesting that Riesz's theorem also served later as a starting point of the modern theory of integration. For further historical remarks, see N. Bourbaki (1955), p. 169.

4.5 Adjoint Operator

With a bounded linear operator T: $X \longrightarrow Y$ on a normed space X we can associate the so-called adjoint operator $T^\times$ of T. A motivation for $T^\times$ comes from its usefulness in the solution of equations involving operators, as we shall see in Sec. 8.5; such equations arise, for instance, in physics and other applications. In the present section we define the adjoint operator $T^\times$ and consider some of its properties, including its relation to the Hilbert-adjoint[3] operator T^* defined in Sec. 3.9. It is important to note that our present discussion depends on the Hahn-Banach theorem (via Theorem 4.3-3), and we would not get very far without it.

We consider a bounded linear operator T: $X \longrightarrow Y$, where X and Y are normed spaces, and want to define the adjoint operator $T^\times$ of T. For this purpose we start from any bounded linear functional g on Y. Clearly, g is defined for all $y \in Y$. Setting $y = Tx$, we obtain a functional on X, call it f:

$$\textbf{(1)} \qquad f(x) = g(Tx) \qquad x \in X.$$

f is linear since g and T are linear. f is bounded because

$$|f(x)| = |g(Tx)| \leqq \|g\| \, \|Tx\| \leqq \|g\| \, \|T\| \, \|x\|.$$

[3] In the case of Hilbert spaces the adjoint operator $T^\times$ is *not* identical with the Hilbert-adjoint operator T^* of T (although $T^\times$ and T^* are then related as explained later in this section). The asterisk for the Hilbert-adjoint operator is almost standard. Hence one should *not* denote the adjoint operator by T^*, because it is troublesome to have a notation mean one thing in a Hilbert space and another thing in the theory of general normed spaces. We use $T^\times$ for the adjoint operator. We prefer this over the less perspicuous T' which is also used in the literature.

Taking the supremum over all $x \in X$ of norm one, we obtain the inequality

$$\|f\| \leqq \|g\| \, \|T\|. \tag{2}$$

This shows that $f \in X'$, where X' is the dual space of X defined in 2.10-3. By assumption, $g \in Y'$. Consequently, for *variable* $g \in Y'$, formula (1) defines an operator from Y' into X', which is called the *adjoint operator* of T and is denoted by $T^\times$. Thus we have

$$\begin{array}{c} X \xrightarrow{\quad T \quad} Y \\ \\ X' \xleftarrow{\quad T^\times \quad} Y' \end{array} \tag{3}$$

Note carefully that $T^\times$ is an operator defined on Y' whereas the given operator T is defined on X. We summarize:

4.5-1 Definition (Adjoint operator $T^\times$). Let $T: X \longrightarrow Y$ be a bounded linear operator, where X and Y are normed spaces. Then the *adjoint operator* $T^\times: Y' \longrightarrow X'$ of T is defined by

$$f(x) = (T^\times g)(x) = g(Tx) \qquad (g \in Y') \tag{4}$$

where X' and Y' are the dual spaces of X and Y, respectively.

Our first goal is to prove that the adjoint operator has the same norm as the operator itself. This property is basic, as we shall see later. In the proof we shall need Theorem 4.3-3, which resulted from the Hahn-Banach theorem. In this way the Hahn-Banach theorem is vital for establishing a satisfactory theory of adjoint operators, which in turn is an essential part of the general theory of linear operators.

4.5-2 Theorem (Norm of the adjoint operator). *The adjoint operator $T^\times$ in Def.* 4.5-1 *is linear and bounded, and*

$$\|T^\times\| = \|T\|. \tag{5}$$

Proof. The operator $T^\times$ is linear since its domain Y' is a vector space and we readily obtain

$$\begin{aligned}(T^\times(\alpha g_1+\beta g_2))(x) &= (\alpha g_1+\beta g_2)(Tx)\\ &= \alpha g_1(Tx)+\beta g_2(Tx)\\ &= \alpha(T^\times g_1)(x)+\beta(T^\times g_2)(x).\end{aligned}$$

We prove (5). From (4) we have $f = T^\times g$, and by (2) it follows that

$$\|T^\times g\| = \|f\| \leqq \|g\|\,\|T\|.$$

Taking the supremum over all $g \in Y'$ of norm one, we obtain the inequality

$$\|T^\times\| \leqq \|T\|. \tag{6}$$

Hence to get (5), we must now prove $\|T^\times\| \geqq \|T\|$. Theorem 4.3-3 implies that for every nonzero $x_0 \in X$ there is a $g_0 \in Y'$ such that

$$\|g_0\| = 1 \qquad \text{and} \qquad g_0(Tx_0) = \|Tx_0\|.$$

Here, $g_0(Tx_0) = (T^\times g_0)(x_0)$ by the definition of the adjoint operator $T^\times$ Writing $f_0 = T^\times g_0$, we thus obtain

$$\begin{aligned}\|Tx_0\| = g_0(Tx_0) &= f_0(x_0)\\ &\leqq \|f_0\|\,\|x_0\|\\ &= \|T^\times g_0\|\,\|x_0\|\\ &\leqq \|T^\times\|\,\|g_0\|\,\|x_0\|.\end{aligned}$$

Since $\|g_0\| = 1$, we thus have for every $x_0 \in X$

$$\|Tx_0\| \leqq \|T^\times\|\,\|x_0\|.$$

(This includes $x_0 = 0$ since $T0 = 0$.) But always

$$\|Tx_0\| \leqq \|T\|\,\|x_0\|,$$

and here $c = \|T\|$ is the *smallest* constant c such that $\|Tx_0\| \leqq c\|x_0\|$ holds for all $x_0 \in X$. Hence $\|T^\times\|$ cannot be smaller than $\|T\|$, that is, we must have $\|T^\times\| \geqq \|T\|$. This and (6) imply (5). ∎

Let us illustrate the present discussion by matrices representing operators. This will also help the reader in setting up examples of his own.

4.5-3 Example (Matrix). In n-dimensional Euclidean space $\mathbf{R}^n$ a linear operator T: $\mathbf{R}^n \longrightarrow \mathbf{R}^n$ can be represented by matrices, (cf. Sec. 2.9) where such a matrix $T_E = (\tau_{jk})$ depends on the choice of a basis $E = \{e_1, \cdots, e_n\}$ for $\mathbf{R}^n$, whose elements are arranged in some order which is kept fixed. We choose a basis E, regard $x = (\xi_1, \cdots, \xi_n)$, $y = (\eta_1, \cdots, \eta_n)$ as column vectors and employ the usual notation for matrix multiplication. Then

$$(7) \qquad y = T_E x, \qquad \text{in components} \qquad \eta_j = \sum_{k=1}^{n} \tau_{jk}\xi_k,$$

where $j = 1, \cdots, n$. Let $F = \{f_1, \cdots, f_n\}$ be the dual basis of E (cf. Sec. 2.9). This is a basis for $\mathbf{R}^{n\prime}$ (which is also Euclidean n-space, by 2.10-5). Then every $g \in \mathbf{R}^{n\prime}$ has a representation

$$g = \alpha_1 f_1 + \cdots + \alpha_n f_n.$$

Now by the definition of the dual basis we have $f_j(y) = f_j(\sum \eta_k e_k) = \eta_j$. Hence by (7) we obtain

$$g(y) = g(T_E x) = \sum_{j=1}^{n} \alpha_j \eta_j = \sum_{j=1}^{n} \sum_{k=1}^{n} \alpha_j \tau_{jk} \xi_k.$$

Interchanging the order of summation, we can write this in the form

$$(8) \qquad g(T_E x) = \sum_{k=1}^{n} \beta_k \xi_k \qquad \text{where} \qquad \beta_k = \sum_{j=1}^{n} \tau_{jk} \alpha_j.$$

We may regard this as the definition of a functional f on X in terms of g, that is,

$$f(x) = g(T_E x) = \sum_{k=1}^{n} \beta_k \xi_k.$$

Remembering the definition of the adjoint operator, we can write this

$$f = T_E{}^{\times} g, \qquad \text{in components,} \qquad \beta_k = \sum_{j=1}^{n} \tau_{jk}\alpha_j.$$

Noting that in β_k we sum with respect to the first subscript (so that we sum over all elements of a *column* of T_E), we have the following result.

If T is represented by a matrix T_E, then the adjoint operator $T^{\times}$ is represented by the transpose of T_E.

We mention that this also holds if T is a linear operator from $\mathbf{C}^n$ into $\mathbf{C}^n$. ∎

In working with the adjoint operator, the subsequent formulas (9) to (12) are helpful; the corresponding proofs are left to the reader. Let $S, T \in B(X, Y)$; cf. Sec. 2.10. Then

$$(S+T)^{\times} = S^{\times} + T^{\times} \tag{9}$$

$$(\alpha T)^{\times} = \alpha T^{\times}. \tag{10}$$

Let X, Y, Z be normed spaces and $T \in B(X, Y)$ and $S \in B(Y, Z)$. Then for the adjoint operator of the product ST we have (see Fig. 41)

$$(ST)^{\times} = T^{\times} S^{\times}. \tag{11}$$

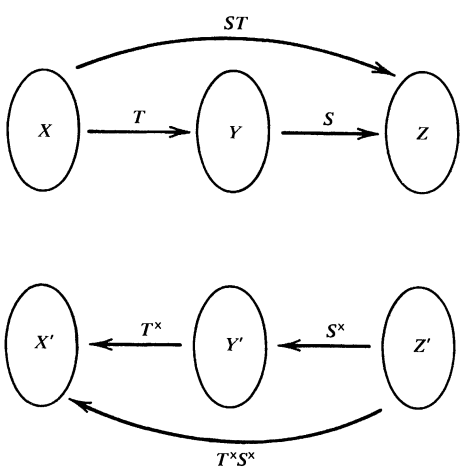


Fig. 41. Illustration of formula (11)

If $T \in B(X, Y)$ and T^{-1} exists and $T^{-1} \in B(Y, X)$, then $(T^\times)^{-1}$ also exists, $(T^\times)^{-1} \in B(X', Y')$ and

$$(T^\times)^{-1} = (T^{-1})^\times. \tag{12}$$

Relation between the adjoint operator $T^\times$ and the Hilbert-adjoint operator T^*. (Cf. Sec. 3.9.) We show that such a relation exists in the case of a bounded linear operator T: $X \longrightarrow Y$ if X and Y are Hilbert spaces, say $X = H_1$ and $Y = H_2$. In this case we first have (Fig. 42)

$$\begin{array}{ccc} H_1 & \xrightarrow{\ T\ } & H_2 \\ H_1' & \xleftarrow{\ T^\times\ } & H_2' \end{array} \tag{13}$$

where, as before, the adjoint operator $T^\times$ of the given operator T is defined by

$$\begin{array}{lll} \text{(a)} & T^\times g = f & \\ & & (f \in H_1',\ g \in H_2'). \\ \text{(b)} & g(Tx) = f(x) & \end{array} \tag{14}$$

The new feature is that since f and g are functionals on Hilbert spaces, they have Riesz representations (cf. 3.8-1), say,

$$\begin{array}{lll} \text{(a)} & f(x) = \langle x, x_0 \rangle & (x_0 \in H_1) \\ \text{(b)} & g(y) = \langle y, y_0 \rangle & (y_0 \in H_2), \end{array} \tag{15}$$

and from Theorem 3.8-1 we also know that x_0 and y_0 are uniquely determined by f and g, respectively. This defines operators

$$A_1\colon\ H_1' \longrightarrow H_1 \qquad \text{by} \qquad A_1 f = x_0,$$

$$A_2\colon\ H_2' \longrightarrow H_2 \qquad \text{by} \qquad A_2 g = y_0.$$

From Theorem 3.8-1 we see that A_1 and A_2 are bijective and isometric since $\|A_1 f\| = \|x_0\| = \|f\|$, and similarly for A_2. Furthermore, the operators A_1 and A_2 are conjugate linear (cf. Sec. 3.1). In fact, if

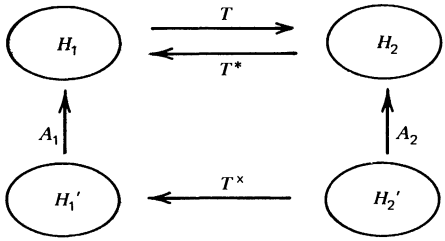


Fig. 42. Operators in formulas (13) and (17)

we write $f_1(x) = \langle x, x_1 \rangle$ and $f_2(x) = \langle x, x_2 \rangle$, we have for all x and scalars α, β

$$(\alpha f_1 + \beta f_2)(x) = \alpha f_1(x) + \beta f_2(x)$$

(16)
$$= \alpha \langle x, x_1 \rangle + \beta \langle x, x_2 \rangle$$
$$= \langle x, \bar{\alpha} x_1 + \bar{\beta} x_2 \rangle.$$

By the definition of A_1 this shows conjugate linearity

$$A_1(\alpha f_1 + \beta f_2) = \bar{\alpha} A_1 f_1 + \bar{\beta} A_1 f_2.$$

For A_2 the proof is similar.

Composition gives the operator (see Fig. 42)

(17) $T^* = A_1 T^\times A_2^{-1} \colon H_2 \longrightarrow H_1$ defined by $T^* y_0 = x_0$.

T^* is linear since it involves *two* conjugate linear mappings, in addition to the linear operator $T^\times$. We prove that T^* is indeed the Hilbert-adjoint operator of T. This is simple since from (14) to (16) we immediately have

$$\langle Tx, y_0 \rangle = g(Tx) = f(x) = \langle x, x_0 \rangle = \langle x, T^* y_0 \rangle,$$

which is (1) in Sec. 3.9, except for the notation. Our result is:

Formula (17) *represents the Hilbert-adjoint operator* T^* *of a linear operator* T *on a Hilbert space in terms of the adjoint operator* $T^\times$ *of* T.

Note further that $\|T^*\| = \|T\|$ (Theorem 3.9-2) now follows immediately from (5) and the isometry of A_1 and A_2. ∎

To complete this discussion, we should also list some of the main differences between the adjoint operator $T^\times$ of T: $X \longrightarrow Y$ and the Hilbert-adjoint operator T^* of T: $H_1 \longrightarrow H_2$, where X, Y are normed spaces and H_1, H_2 are Hilbert spaces.

$T^\times$ is defined on the dual of the space which contains the range of T, whereas T^* is defined directly on the space which contains the range of T. This property of T^* enabled us to define important classes of operators by the use of their Hilbert-adjoint operators (cf. 3.10-1).

For $T^\times$ we have by (10)

$$(\alpha T)^\times = \alpha T^\times$$

but for T^* we have by 3.9-4

$$(\alpha T)^* = \bar{\alpha} T^*.$$

In the finite dimensional case, $T^\times$ is represented by the transpose of the matrix representing T, whereas T^* is represented by the complex conjugate transpose of that matrix (for details, see 4.5-3 and 3.10-2).

Problems

1. Show that the functional defined by (1) is linear.

2. What are the adjoints of a zero operator 0 and an identity operator I?

3. Prove (9).

4. Prove (10).

5. Prove (11).

6. Show that $(T^n)^\times = (T^\times)^n$.

7. What formula for matrices do we obtain by combining (11) and Example 4.5-3?

8. Prove (12).

9. **(Annihilator)** Let X and Y be normed spaces, T: $X \longrightarrow Y$ a bounded linear operator and $M = \overline{\mathcal{R}(T)}$, the closure of the range of T. Show that (cf. Prob. 13, Sec. 2.10)

$$M^a = \mathcal{N}(T^\times).$$

10. (Annihilator) Let B be a subset of the dual space X' of a normed space X. The *annihilator* ${}^aB$ of B is defined to be

$$ {}^aB = \{x \in X \mid f(x) = 0 \text{ for all } f \in B\}. $$

Show that in Prob. 9,

$$ \mathfrak{R}(T) \subset {}^a\mathfrak{N}(T^\times). $$

What does this mean with respect to the task of solving an equation $Tx = y$?

4.6 Reflexive Spaces

Algebraic reflexivity of *vector* spaces was discussed in Sec. 2.8. *Reflexivity* of *normed* spaces will be the topic of the present section. But let us first recall what we did in Sec. 2.8. We remember that a vector space X is said to be *algebraically reflexive* if the canonical mapping C: $X \longrightarrow X^{**}$ is surjective. Here $X^{**} = (X^*)^*$ is the second algebraic dual space of X and the mapping C is defined by $x \longmapsto g_x$ where

$$ g_x(f) = f(x) \qquad (f \in X^* \text{ variable}); \tag{1} $$

that is, for any $x \in X$ the image is the linear functional g_x defined by (1). If X is finite dimensional, then X is algebraically reflexive. This was shown in Theorem 2.9-3.

Let us now turn to our actual task. We consider a normed space X, its dual space X' as defined in 2.10-3 and, moreover, the dual space $(X')'$ of X'. This space is denoted by X'' and is called the **second dual space** of X (or *bidual space* of X).

We define a functional g_x on X' by choosing a fixed $x \in X$ and setting

$$ g_x(f) = f(x) \qquad (f \in X' \text{ variable}). \tag{2} $$

This looks like (1), but note that now f is bounded. And g_x turns out to be bounded, too, since we have the basic

4.6-1 Lemma (Norm of g_x). *For every fixed x in a normed space X, the functional g_x defined by (2) is a bounded linear functional on X', so that $g_x \in X''$, and has the norm*

$$\|g_x\| = \|x\|. \tag{3}$$

Proof. Linearity of g_x is known from Sec. 2.8, and (3) follows from (2) and Corollary 4.3-4:

$$\|g_x\| = \sup_{\substack{f \in X' \\ f \neq 0}} \frac{|g_x(f)|}{\|f\|} = \sup_{\substack{f \in X' \\ f \neq 0}} \frac{|f(x)|}{\|f\|} = \|x\|. \tag{4}$$

∎

To every $x \in X$ there corresponds a unique bounded linear functional $g_x \in X''$ given by (2). This defines a mapping

$$\begin{aligned} C:\ X &\longrightarrow X'' \\ x &\longmapsto g_x. \end{aligned} \tag{5}$$

C is called the **canonical mapping** of X into X''. We show that C is linear and injective and preserves the norm. This can be expressed in terms of an isomorphism of normed spaces as defined in Sec. 2.10:

4.6-2 Lemma (Canonical mapping). *The canonical mapping C given by (5) is an isomorphism of the normed space X onto the normed space $\mathcal{R}(C)$, the range of C.*

Proof. Linearity of C is seen as in Sec. 2.8 because

$$g_{\alpha x+\beta y}(f) = f(\alpha x + \beta y) = \alpha f(x) + \beta f(y) = \alpha g_x(f) + \beta g_y(f).$$

In particular, $g_x - g_y = g_{x-y}$. Hence by (3) we obtain

$$\|g_x - g_y\| = \|g_{x-y}\| = \|x - y\|.$$

This shows that C is isometric; it preserves the norm. Isometry implies injectivity. We can also see this directly from our formula. Indeed, if $x \neq y$, then $g_x \neq g_y$ by axiom (N2) in Sec. 2.2. Hence C is bijective, regarded as a mapping onto its range. ∎

X is said to be **embeddable** in a normed space Z if X is isomorphic with a subspace of Z. This is similar to Sec. 2.8, but note that here we are dealing with isomorphisms of normed spaces, that is, vector space isomorphisms which preserve norm (cf. Sec. 2.10). Lemma 4.6-2 shows that X is embeddable in X'', and C is also called the *canonical embedding* of X into X''.

In general, C will not be surjective, so that the range $\mathcal{R}(C)$ will be a *proper* subspace of X''. The surjective case when $\mathcal{R}(C)$ is all of X'' is important enough to give it a name:

4.6-3 Definition (Reflexivity). A normed space X is said to be *reflexive* if

$$\mathcal{R}(C) = X''$$

where C: $X \longrightarrow X''$ is the canonical mapping given by (5) and (2). ∎

This concept was introduced by H. Hahn (1927) and called "reflexivity" by E. R. Lorch (1939). Hahn recognized the importance of reflexivity in his study of linear equations in normed spaces which was motivated by integral equations and also contains the Hahn-Banach theorem as well as the earliest investigation of dual spaces.

If X is reflexive, it is isomorphic (hence isometric) with X'', by Lemma 4.6-2. It is interesting that the converse does not generally hold, as R. C. James (1950, 1951) has shown.

Furthermore, completeness does not imply reflexivity, but conversely we have

4.6-4 Theorem (Completeness). *If a normed space X is reflexive, it is complete (hence a Banach space).*

Proof. Since X'' is the dual space of X', it is complete by Theorem 2.10-4. Reflexivity of X means that $\mathcal{R}(C) = X''$. Completeness of X now follows from that of X'' by Lemma 4.6-2. ∎

$\mathbf{R}^n$ is reflexive. This follows directly from 2.10-5. It is typical of any finite dimensional normed space X. Indeed, if $\dim X < \infty$, then every linear functional on X is bounded (cf. 2.7-8), so that $X' = X^*$ and algebraic reflexivity of X (cf. 2.9-3) thus implies

4.6-5 Theorem (Finite dimension). *Every finite dimensional normed space is reflexive.*

l^p with $1 < p < +\infty$ is reflexive. This follows from 2.10-7. Similarly, $L^p[a, b]$ with $1 < p < +\infty$ is reflexive, as can be shown. It can also be proved that nonreflexive spaces are $C[a, b]$ (cf. 2.2-5), l^1 (proof below), $L^1[a, b]$, l^∞ (cf. 2.2-4) and the subspaces c and c_0 of l^∞, where c is the space of all convergent sequences of scalars and c_0 is the space of all sequences of scalars converging to zero.

4.6-6 Theorem (Hilbert space). *Every Hilbert space H is reflexive.*

Proof. We shall prove surjectivity of the canonical mapping $C: H \longrightarrow H''$ by showing that for every $g \in H''$ there is an $x \in H$ such that $g = Cx$. As a preparation we define $A: H' \longrightarrow H$ by $Af = z$, where z is given by the Riesz representation $f(x) = \langle x, z \rangle$ in 3.8-1. From 3.8-1 we know that A is bijective and isometric. A is conjugate linear, as we see from (16), Sec. 4.5. Now H' is complete by 2.10-4 and a Hilbert space with inner product defined by

$$\langle f_1, f_2 \rangle_1 = \langle Af_2, Af_1 \rangle.$$

Note the order of f_1, f_2 on both sides. (IP1) to (IP4) in Sec. 3.1 is readily verified. In particular, (IP2) follows from the conjugate linearity of A:

$$\langle \alpha f_1, f_2 \rangle_1 = \langle Af_2, A(\alpha f_1) \rangle = \langle Af_2, \bar{\alpha} Af_1 \rangle = \alpha \langle f_1, f_2 \rangle_1.$$

Let $g \in H''$ be arbitrary. Let its Riesz representation be

$$g(f) = \langle f, f_0 \rangle_1 = \langle Af_0, Af \rangle.$$

We now remember that $f(x) = \langle x, z \rangle$ where $z = Af$. Writing $Af_0 = x$, we thus have

$$\langle Af_0, Af \rangle = \langle x, z \rangle = f(x).$$

Together, $g(f) = f(x)$, that is, $g = Cx$ by the definition of C. Since $g \in H''$ was arbitrary, C is surjective, so that H is reflexive. ∎

Sometimes separability and nonseparability (cf. 1.3-5) can play a role in proofs that certain spaces are not reflexive. This connection between reflexivity and separability is interesting and quite simple. The key is Theorem 4.6-8 (below), which states that separability of X'

implies separability of X (the converse not being generally true). Hence if a normed space X is reflexive, X'' is isomorphic with X by 4.6-2, so that in this case, separability of X implies separability of X'' and, by 4.6-8, the space X' is also separable. From this we have the following result.

A separable normed space X with a nonseparable dual space X' cannot be reflexive.

Example. l^1 is not reflexive.

Proof. l^1 is separable by 1.3-10, but $l^{1\prime} = l^\infty$ is not; cf. 2.10-6 and 1.3-9.

The desired Theorem 4.6-8 will be obtained from the following lemma. A simple illustration of the lemma is shown in Fig. 43.

4.6-7 Lemma (Existence of a functional). *Let Y be a proper closed subspace of a normed space X. Let $x_0 \in X - Y$ be arbitrary and*

$$\delta = \inf_{\tilde{y} \in Y} \|\tilde{y} - x_0\| \tag{6}$$

the distance from x_0 to Y. Then there exists an $\tilde{f} \in X'$ such that

$$\|\tilde{f}\| = 1, \qquad \tilde{f}(y) = 0 \text{ for all } y \in Y, \qquad \tilde{f}(x_0) = \delta. \tag{7}$$

Proof. The idea of the proof is simple. We consider the subspace $Z \subset X$ spanned by Y and x_0, define on Z a bounded linear functional f by

$$f(z) = f(y + \alpha x_0) = \alpha\delta \qquad y \in Y, \tag{8}$$

show that f satisfies (7) and extend f to X by 4.3-2. The details are as follows.

Every $z \in Z = \operatorname{span}(Y \cup \{x_0\})$ has a unique representation

$$z = y + \alpha x_0 \qquad y \in Y.$$

This is used in (8). Linearity of f is readily seen. Also, since Y is closed, $\delta > 0$, so that $f \neq 0$. Now $\alpha = 0$ gives $f(y) = 0$ for all $y \in Y$. For

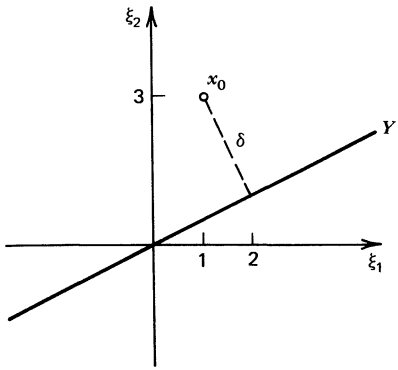


Fig. 43. Illustration of Lemma 4.6-7 for the Euclidean space $X = \mathbf{R}^3$, where Y is represented by $\xi_2 = \xi_1/2$, $\xi_3 = 0$ and $x_0 = (1, 3, 0)$, so that $\delta = \sqrt{5}$, $Z = \text{span}\,(Y \cup \{x_0\})$ is the $\xi_1\xi_2$-plane and $f(z) = (-\xi_1 + 2\xi_2)/\sqrt{5}$.

$\alpha = 1$ and $y = 0$ we have $f(x_0) = \delta$.

We show that f is bounded. $\alpha = 0$ gives $f(z) = 0$. Let $\alpha \neq 0$. Using (6) and noting that $-(1/\alpha)y \in Y$, we obtain

$$|f(z)| = |\alpha|\,\delta = |\alpha| \inf_{\tilde{y} \in Y} \|\tilde{y} - x_0\|$$

$$\leqq |\alpha| \left\| -\frac{1}{\alpha} y - x_0 \right\|$$

$$= \|y + \alpha x_0\|,$$

that is, $|f(z)| \leqq \|z\|$. Hence f is bounded and $\|f\| \leqq 1$.

We show that $\|f\| \geqq 1$. By the definition of an infimum, Y contains a sequence (y_n) such that $\|y_n - x_0\| \longrightarrow \delta$. Let $z_n = y_n - x_0$. Then we have $f(z_n) = -\delta$ by (8) with $\alpha = -1$. Also

$$\|f\| = \sup_{\substack{z \in Z \\ z \neq 0}} \frac{|f(z)|}{\|z\|} \geqq \frac{|f(z_n)|}{\|z_n\|} = \frac{\delta}{\|z_n\|} \qquad \longrightarrow \qquad \frac{\delta}{\delta} = 1$$

as $n \longrightarrow \infty$. Hence $\|f\| \geqq 1$, so that $\|f\| = 1$. By the Hahn-Banach theorem 4.3-2 for normed spaces we can extend f to X without increasing the norm. ∎

Using this lemma, we shall now obtain the desired

4.6-8 Theorem (Separability). *If the dual space X' of a normed space X is separable, then X itself is separable.*

Proof. We assume that X' is separable. Then the unit sphere $U' = \{f \mid \|f\| = 1\} \subset X'$ also contains a countable dense subset, say, (f_n). Since $f_n \in U'$, we have

$$\|f_n\| = \sup_{\|x\|=1} |f_n(x)| = 1.$$

By the definition of a supremum we can find points $x_n \in X$ of norm 1 such that

$$|f_n(x_n)| \geqq \frac{1}{2}.$$

Let Y be the closure of span (x_n). Then Y is separable because Y has a countable dense subset, namely, the set of all linear combinations of the x_n's with coefficients whose real and imaginary parts are rational.

We show that $Y = X$. Suppose $Y \neq X$. Then, since Y is closed, by Lemma 4.6-7 there exists an $\tilde{f} \in X'$ with $\|\tilde{f}\| = 1$ and $\tilde{f}(y) = 0$ for all $y \in Y$. Since $x_n \in Y$, we have $\tilde{f}(x_n) = 0$ and for all n,

$$\begin{aligned} \frac{1}{2} \leqq |f_n(x_n)| &= |f_n(x_n) - \tilde{f}(x_n)| \\ &= |(f_n - \tilde{f})(x_n)| \\ &\leqq \|f_n - \tilde{f}\| \, \|x_n\|, \end{aligned}$$

where $\|x_n\| = 1$. Hence $\|f_n - \tilde{f}\| \geqq \frac{1}{2}$, but this contradicts the assumption that (f_n) is dense in U' because $\tilde{f}$ is itself in U'; in fact, $\|\tilde{f}\| = 1$. ∎

Problems

1. What are the functionals f and g_x in (2) if $X = \mathbf{R}^n$?
2. Give a simpler proof of Lemma 4.6-7 for the case that X is a Hilbert space.
3. If a normed space X is reflexive, show that X' is reflexive.

4. Show that a Banach space X is reflexive if and only if its dual space X' is reflexive. (*Hint.* It can be shown that a closed subspace of a reflexive Banach space is reflexive. Use this fact, without proving it.)

5. Show that under the assumptions of Lemma 4.6-7 there exists a bounded linear functional h on X such that

$$\|h\| = 1/\delta, \qquad h(y) = 0 \text{ for all } y \in Y, \qquad h(x_0) = 1.$$

6. Show that different closed subspaces Y_1 and Y_2 of a normed space X have different annihilators. (Cf. Sec. 2.10, Prob. 13.)

7. Let Y be a closed subspace of a normed space X such that every $f \in X'$ which is zero everywhere on Y is zero everywhere on the whole space X. Show that then $Y = X$.

8. Let M be any subset of a normed space X. Show that an $x_0 \in X$ is an element of $A = \overline{\text{span } M}$ if and only if $f(x_0) = 0$ for every $f \in X'$ such that $f|_M = 0$.

9. **(Total set)** Show that a subset M of a normed space X is total in X if and only if every $f \in X'$ which is zero everywhere on M is zero everywhere on X.

10. Show that if a normed space X has a linearly independent subset of n elements, so does the dual space X'.

4.7 Category Theorem. Uniform Boundedness Theorem

The uniform boundedness theorem (or *uniform boundedness principle*) by S. Banach and H. Steinhaus (1927) is of great importance. In fact, throughout analysis there are many instances of results related to this theorem, the earliest being an investigation by H. Lebesgue (1909). The uniform boundedness theorem is often regarded as one of the corner stones of functional analysis in normed spaces, the others being the Hahn–Banach theorem (Secs. 4.2, 4.3), the open mapping theorem (Sec. 4.12) and the closed graph theorem (Sec. 4.13). Unlike the Hahn–Banach theorem, the other three of these four theorems require completeness. Indeed, they characterize some of the most important

properties of Banach spaces which normed spaces in general may not have.

It is quite interesting to note that we shall obtain all three theorems from a common source. More precisely, we shall prove the so-called Baire's category theorem and derive from it the uniform boundedness theorem (in this section) as well as the open mapping theorem (in Sec. 4.12). The latter will then readily entail the closed graph theorem (in Sec. 4.13).

Baire's category theorem has various other applications in functional analysis and is the main reason why category enters into numerous proofs; cf., for instance, the more advanced books by R. E. Edwards (1965) and J. L. Kelley and I. Namioka (1963).

In Def. 4.7-1 we state the concepts needed for Baire's theorem 4.7-2. Each concept has two names, a new name and an old one given in parentheses. The latter is on the way out because "category" is now being used for an entirely different mathematical purpose (which will not occur in this book).

4.7-1 Definition (Category). A subset M of a metric space X is said to be

(a) *rare* (or *nowhere dense*) *in* X if its closure $\bar{M}$ has no interior points (cf. Sec. 1.3),

(b) *meager* (or *of the first category*) *in* X if M is the union of countably many sets each of which is rare in X,

(c) *nonmeager* (or *of the second category*) *in* X if M is not meager in X. ∎

4.7-2 Baire's Category Theorem (Complete metric spaces). *If a metric space* $X \neq \emptyset$ *is complete, it is nonmeager in itself.*

Hence if $X \neq \emptyset$ *is complete and*

$$X = \bigcup_{k=1}^{\infty} A_k \qquad (A_k \text{ closed}) \tag{1}$$

then at least one A_k *contains a nonempty open subset.*

Proof. The idea of the proof is simple. Suppose the complete metric space $X \neq \emptyset$ were meager in itself. Then

$$X = \bigcup_{k=1}^{\infty} M_k \tag{1*}$$

with each M_k rare in X. We shall construct a Cauchy sequence (p_k) whose limit p (which exists by completeness) is in no M_k, thereby contradicting the representation (1*).

By assumption, M_1 is rare in X, so that, by definition, $\bar{M}_1$ does not contain a nonempty open set. But X does (for instance, X itself). This implies $\bar{M}_1 \neq X$. Hence the complement $\bar{M}_1^{\mathsf{C}} = X - \bar{M}_1$ of $\bar{M}_1$ is not empty and open. We may thus choose a point p_1 in $\bar{M}_1^{\mathsf{C}}$ and an open ball about it, say,

$$B_1 = B(p_1; \varepsilon_1) \subset \bar{M}_1^{\mathsf{C}} \qquad \varepsilon_1 < \tfrac{1}{2}.$$

By assumption, M_2 is rare in X, so that $\bar{M}_2$ does not contain a nonempty open set. Hence it does not contain the open ball $B(p_1; \frac{1}{2}\varepsilon_1)$. This implies that $\bar{M}_2^{\mathsf{C}} \cap B(p_1; \frac{1}{2}\varepsilon_1)$ is not empty and open, so that we may choose an open ball in this set, say,

$$B_2 = B(p_2; \varepsilon_2) \subset \bar{M}_2^{\mathsf{C}} \cap B(p_1; \tfrac{1}{2}\varepsilon_1) \qquad \varepsilon_2 < \tfrac{1}{2}\varepsilon_1.$$

By induction we thus obtain a sequence of balls

$$B_k = B(p_k; \varepsilon_k) \qquad \varepsilon_k < 2^{-k}$$

such that $B_k \cap M_k = \varnothing$ and

$$B_{k+1} \subset B(p_k; \tfrac{1}{2}\varepsilon_k) \subset B_k \qquad k = 1, 2, \cdots.$$

Since $\varepsilon_k < 2^{-k}$, the sequence (p_k) of the centers is Cauchy and converges, say, $p_k \longrightarrow p \in X$ because X is complete by assumption. Also, for every m and $n > m$ we have $B_n \subset B(p_m; \frac{1}{2}\varepsilon_m)$, so that

$$\begin{aligned} d(p_m, p) &\leqq d(p_m, p_n) + d(p_n, p) \\ &< \tfrac{1}{2}\varepsilon_m + d(p_n, p) \qquad \longrightarrow \qquad \tfrac{1}{2}\varepsilon_m \end{aligned}$$

as $n \longrightarrow \infty$. Hence $p \in B_m$ for every m. Since $B_m \subset \bar{M}_m^{\mathsf{C}}$, we now see that $p \notin M_m$ for every m, so that $p \notin \bigcup M_m = X$. This contradicts $p \in X$. Baire's theorem is proved. ∎

We note that the converse of Baire's theorem is not generally true. An example of an incomplete normed space which is nonmeager in itself is given in N. Bourbaki (1955), Ex. 6, pp. 3–4.

From Baire's theorem we shall now readily obtain the desired uniform boundedness theorem. This theorem states that if X is a Banach space and a sequence of operators $T_n \in B(X, Y)$ is bounded at every point $x \in X$, then the sequence is uniformly bounded. In other words, pointwise boundedness implies boundedness in some stronger sense, namely, uniform boundedness. (The real number c_x in (2), below, will vary in general with x, a fact which we indicate by the subscript x; the essential point is that c_x does not depend on n.)

4.7-3 Uniform Boundedness Theorem. *Let (T_n) be a sequence of bounded linear operators $T_n: X \longrightarrow Y$ from a Banach space X into a normed space Y such that $(\|T_n x\|)$ is bounded for every $x \in X$, say,*

$$\|T_n x\| \leqq c_x \qquad n = 1, 2, \cdots, \tag{2}$$

where c_x is a real number. Then the sequence of the norms $\|T_n\|$ is bounded, that is, there is a c such that

$$\|T_n\| \leqq c \qquad n = 1, 2, \cdots. \tag{3}$$

Proof. For every $k \in \mathbf{N}$, let $A_k \subset X$ be the set of all x such that

$$\|T_n x\| \leqq k \qquad \text{for all } n.$$

A_k is closed. Indeed, for any $x \in \bar{A}_k$ there is a sequence (x_j) in A_k converging to x. This means that for every fixed n we have $\|T_n x_j\| \leqq k$ and obtain $\|T_n x\| \leqq k$ because T_n is continuous and so is the norm (cf. Sec. 2.2). Hence $x \in A_k$, and A_k is closed.

By (2), each $x \in X$ belongs to some A_k. Hence

$$X = \bigcup_{k=1}^{\infty} A_k.$$

Since X is complete, Baire's theorem implies that some A_k contains an open ball, say,

$$B_0 = B(x_0; r) \subset A_{k_0}. \tag{4}$$

Let $x \in X$ be arbitrary, not zero. We set

$$z = x_0 + \gamma x \qquad \gamma = \frac{r}{2\|x\|}. \tag{5}$$

Then $\|z - x_0\| < r$, so that $z \in B_0$. By (4) and from the definition of A_{k_0} we thus have $\|T_n z\| \leqq k_0$ for all n. Also $\|T_n x_0\| \leqq k_0$ since $x_0 \in B_0$. From (5) we obtain

$$x = \frac{1}{\gamma}(z - x_0).$$

This yields for all n

$$\|T_n x\| = \frac{1}{\gamma}\|T_n(z - x_0)\| \leqq \frac{1}{\gamma}(\|T_n z\| + \|T_n x_0\|) \leqq \frac{4}{r}\|x\| \, k_0.$$

Hence for all n,

$$\|T_n\| = \sup_{\|x\|=1} \|T_n x\| \leqq \frac{4}{r} k_0,$$

which is of the form (3) with $c = 4k_0/r$. ∎

Applications

4.7-4 Space of polynomials. *The normed space X of all polynomials with norm defined by*

$$\|x\| = \max_j |\alpha_j| \qquad (\alpha_0, \alpha_1, \cdots \text{ the coefficients of } x) \tag{6}$$

is not complete.

Proof. We construct a sequence of bounded linear operators on X which satisfies (2) but not (3), so that X cannot be complete.

We may write a polynomial $x \neq 0$ of degree N_x in the form

$$x(t) = \sum_{j=0}^{\infty} \alpha_j t^j \qquad (\alpha_j = 0 \text{ for } j > N_x).$$

(For $x = 0$ the degree is not defined in the usual discussion of degree, but this does not matter here.) As a sequence of operators on X we take the sequence of functionals $T_n = f_n$ defined by

$$T_n 0 = f_n(0) = 0, \qquad T_n x = f_n(x) = \alpha_0 + \alpha_1 + \cdots + \alpha_{n-1}. \tag{7}$$

f_n is linear. f_n is bounded since $|\alpha_j| \leqq \|x\|$ by (6), so that $|f_n(x)| \leqq n\,\|x\|$. Furthermore, for each fixed $x \in X$ the sequence $(|f_n(x)|)$ satisfies (2) because a polynomial x of degree N_x has $N_x + 1$ coefficients, so that by (7) we have

$$|f_n(x)| \leqq (N_x + 1) \max_j |\alpha_j| = c_x$$

which is of the form (2).

We now show that (f_n) does not satisfy (3), that is, there is no c such that $\|T_n\| = \|f_n\| \leqq c$ for all n. This we do by choosing particularly disadvantageous polynomials. For f_n we choose x defined by

$$x(t) = 1 + t + \cdots + t^n.$$

Then $\|x\| = 1$ by (6) and

$$f_n(x) = 1 + 1 + \cdots + 1 = n = n\,\|x\|.$$

Hence $\|f_n\| \geqq |f_n(x)|/\|x\| = n$, so that $(\|f_n\|)$ is unbounded. ∎

4.7-5 Fourier series. From 3.5-1 we remember that the *Fourier series* of a given periodic function x of period 2π is of the form

$$\frac{1}{2}a_0 + \sum_{m=1}^{\infty} (a_m \cos mt + b_m \sin mt) \tag{8}$$

with the Fourier coefficients of x given by the Euler formulas

$$a_m = \frac{1}{\pi}\int_0^{2\pi} x(t) \cos mt\, dt, \qquad b_m = \frac{1}{\pi}\int_0^{2\pi} x(t) \sin mt\, dt. \tag{9}$$

[We wrote $a_0/2$ in (8) to have only two formulas in (9), whereas in 3.5-1 we wrote a_0 and needed three Euler formulas.]

It is well-known that the series (8) may converge even at points where x is discontinuous. (Problem 15 gives a simple example.) This shows that continuity is not necessary for convergence. Surprising enough, continuity is not sufficient either.[4] Indeed, using the uniform boundedness theorem, we can show the following.

[4] Continuity and the existence of the right-hand and left-hand derivatives at a point t_0 is sufficient for convergence at t_0. Cf. W. Rogosinski (1959), p. 70.

There exist real-valued continuous functions whose Fourier series diverge at a given point t_0.

Proof. Let X be the normed space of all real-valued continuous functions of period 2π with norm defined by

$$\|x\| = \max |x(t)|. \tag{10}$$

X is a Banach space, as follows from 1.5-5 with $a = 0$ and $b = 2\pi$. We may take $t_0 = 0$, without restricting generality. To prove our statement, we shall apply the uniform boundedness theorem 4.7-3 to $T_n = f_n$ where $f_n(x)$ is the value at $t = 0$ of the nth partial sum of the Fourier series of x. Since for $t = 0$ the sine terms are zero and the cosine is one, we see from (8) and (9) that

$$\begin{aligned} f_n(x) &= \tfrac{1}{2}a_0 + \sum_{m=1}^{n} a_m \\ &= \frac{1}{\pi}\int_0^{2\pi} x(t)\left[\frac{1}{2} + \sum_{m=1}^{n} \cos mt\right] dt. \end{aligned}$$

We want to determine the function represented by the sum under the integral sign. For this purpose we calculate

$$\begin{aligned} 2\sin\tfrac{1}{2}t \sum_{m=1}^{n} \cos mt &= \sum_{m=1}^{n} 2\sin\tfrac{1}{2}t \cos mt \\ &= \sum_{m=1}^{n} [-\sin(m-\tfrac{1}{2})t + \sin(m+\tfrac{1}{2})t] \\ &= -\sin\tfrac{1}{2}t + \sin(n+\tfrac{1}{2})t, \end{aligned}$$

where the last expression follows by noting that most of the terms drop out in pairs. Dividing this by $\sin\frac{1}{2}t$ and adding 1 on both sides, we have

$$1 + 2\sum_{m=1}^{n} \cos mt = \frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}.$$

Consequently, the formula for $f_n(x)$ can be written in the simple form

$$f_n(x) = \frac{1}{2\pi}\int_0^{2\pi} x(t)q_n(t)\,dt, \qquad q_n(t) = \frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}. \tag{11}$$

Using this, we can show that the linear functional f_n is bounded. In fact, by (10) and (11),

$$|f_n(x)| \leqq \frac{1}{2\pi} \max |x(t)| \int_0^{2\pi} |q_n(t)|\, dt = \frac{\|x\|}{2\pi} \int_0^{2\pi} |q_n(t)|\, dt.$$

From this we see that f_n is bounded. Furthermore, by taking the supremum over all x of norm one we obtain

$$\|f_n\| \leqq \frac{1}{2\pi} \int_0^{2\pi} |q_n(t)|\, dt.$$

Actually, the equality sign holds, as we shall now prove. For this purpose let us first write

$$|q_n(t)| = y(t) q_n(t)$$

where $y(t) = +1$ at every t at which $q_n(t) \geqq 0$ and $y(t) = -1$ elsewhere. y is not continuous, but for any given $\varepsilon > 0$ it may be modified to a continuous x of norm 1 such that for this x we have

$$\frac{1}{2\pi} \left| \int_0^{2\pi} [x(t) - y(t)] q_n(t)\, dt \right| < \varepsilon.$$

Writing this as two integrals and using (11), we obtain

$$\frac{1}{2\pi} \left| \int_0^{2\pi} x(t) q_n(t)\, dt - \int_0^{2\pi} y(t) q_n(t)\, dt \right| = \left| f_n(x) - \frac{1}{2\pi} \int_0^{2\pi} |q_n(t)|\, dt \right| < \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary and $\|x\| = 1$, this proves the desired formula

$$\|f_n\| = \frac{1}{2\pi} \int_0^{2\pi} |q_n(t)|\, dt. \tag{12}$$

We finally show that the sequence $(\|f_n\|)$ is unbounded. Substituting into (12) the expression for q_n from (11), using the fact that

$|\sin \frac{1}{2}t| < \frac{1}{2}t$ for $t \in (0, 2\pi]$ and setting $(n+\frac{1}{2})t = v$, we obtain

$$\begin{aligned}
\|f_n\| &= \frac{1}{2\pi} \int_0^{2\pi} \left| \frac{\sin (n+\frac{1}{2})t}{\sin \frac{1}{2}t} \right| dt \\
&> \frac{1}{\pi} \int_0^{2\pi} \frac{|\sin (n+\frac{1}{2})t|}{t} \, dt \\
&= \frac{1}{\pi} \int_0^{(2n+1)\pi} \frac{|\sin v|}{v} \, dv \\
&= \frac{1}{\pi} \sum_{k=0}^{2n} \int_{k\pi}^{(k+1)\pi} \frac{|\sin v|}{v} \, dv \\
&\geqq \frac{1}{\pi} \sum_{k=0}^{2n} \frac{1}{(k+1)\pi} \int_{k\pi}^{(k+1)\pi} |\sin v| \, dv \\
&= \frac{2}{\pi^2} \sum_{k=0}^{2n} \frac{1}{k+1} \longrightarrow \infty \qquad \text{as } n \longrightarrow \infty
\end{aligned}$$

since the harmonic series diverges. Hence $(\|f_n\|)$ is unbounded, so that (3) (with $T_n = f_n$) does not hold. Since X is complete, this implies that (2) cannot hold for all x. Hence there must be an $x \in X$ such that $(|f_n(x)|)$ is unbounded. But by the definition of the f_n's this means that the Fourier series of that x diverges at $t = 0$. ∎

Note that our existence proof does not tell us how to find such a continuous function x whose Fourier series diverges at a t_0. Examples of such functions were given by L. Fejér (1910); one is reproduced in W. Rogosinski (1959), pp. 76–77.

Problems

1. Of what category is the set of all rational numbers (*a*) in $\mathbf{R}$, (*b*) in itself (taken with the usual metric)?

2. Of what category is the set of all integers (*a*) in $\mathbf{R}$, (*b*) in itself (taken with the metric induced from $\mathbf{R}$)?

3. Find all rare sets in a discrete metric space X. (Cf. 1.1-8.)

4. Find a meager dense subset in $\mathbf{R}^2$.

5. Show that a subset M of a metric space X is rare in X if and only if $(\bar{M})^c$ is dense in X.

6. Show that the complement M^c of a meager subset M of a complete metric space X is nonmeager.

7. **(Resonance)** Let X be a Banach space, Y a normed space and $T_n \in B(X, Y)$, $n = 1, 2, \cdots$, such that $\sup_n \|T_n\| = +\infty$. Show that there is an $x_0 \in X$ such that $\sup_n \|T_n x_0\| = +\infty$. [The point x_0 is often called a *point of resonance*, and our problem motivates the term *resonance theorem* for the uniform boundedness theorem.]

8. Show that completeness of X is essential in Theorem 4.7-3 and cannot be omitted. [Consider the subspace $X \subset l^\infty$ consisting of all $x = (\xi_j)$ such that $\xi_j = 0$ for $j \geqq J \in \mathbf{N}$, where J depends on x, and let T_n be defined by $T_n x = f_n(x) = n\xi_n$.]

9. Let $T_n = S^n$, where the operator S: $l^2 \longrightarrow l^2$ is defined by $(\xi_1, \xi_2, \xi_3, \cdots) \longmapsto (\xi_3, \xi_4, \xi_5, \cdots)$. Find a bound for $\|T_n x\|$; find $\lim_{n\to\infty} \|T_n x\|$, $\|T_n\|$ and $\lim_{n\to\infty} \|T_n\|$.

10. **(Space c_0)** Let $y = (\eta_j)$, $\eta_j \in \mathbf{C}$, be such that $\sum \xi_j \eta_j$ converges for every $x = (\xi_j) \in c_0$, where $c_0 \subset l^\infty$ is the subspace of all complex sequences converging to zero. Show that $\sum |\eta_j| < \infty$. (Use 4.7-3.)

11. Let X be a Banach space, Y a normed space and $T_n \in B(X, Y)$ such that $(T_n x)$ is Cauchy in Y for every $x \in X$. Show that $(\|T_n\|)$ is bounded.

12. If, in addition, Y in Prob. 11 is complete, show that $T_n x \longrightarrow Tx$, where $T \in B(X, Y)$.

13. If (x_n) in a Banach space X is such that $(f(x_n))$ is bounded for all $f \in X'$, show that $(\|x_n\|)$ is bounded.

14. If X and Y are Banach spaces and $T_n \in B(X, Y)$, $n = 1, 2, \cdots$, show that equivalent statements are:
(*a*) $(\|T_n\|)$ is bounded,
(*b*) $(\|T_n x\|)$ is bounded for all $x \in X$,
(*c*) $(|g(T_n x)|)$ is bounded for all $x \in X$ and all $g \in Y'$.

15. To illustrate that a Fourier series of a function x may converge even at a point where x is discontinuous, find the Fourier series of

$$x(t) = \begin{cases} 0 & \text{if } -\pi \leqq t < 0 \\ 1 & \text{if } 0 \leqq t < \pi \end{cases} \qquad \text{and} \qquad x(t + 2\pi) = x(t).$$

Graph x and the partial sums s_0, s_1, s_2, s_3, and compare with Fig. 44. Show that at $t = \pm n\pi$ the series has the value 1/2, the arithmetic mean of the right and left limits of x; this behavior is typical of Fourier series.

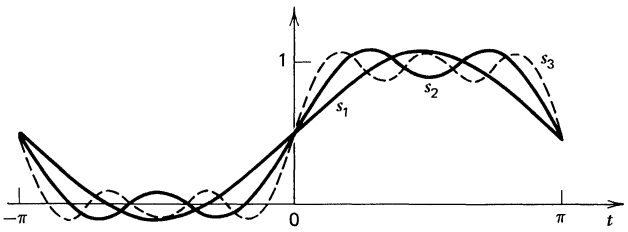


Fig. 44. First three partial sums s_1, s_2, s_3 in Prob. 15

4.8 Strong and Weak Convergence

We know that in calculus one defines different types of convergence (ordinary, conditional, absolute and uniform convergence). This yields greater flexibility in the theory and application of sequences and series. In functional analysis the situation is similar, and one has an even greater variety of possibilities that turn out to be of practical interest. In the present section we are primarily concerned with "weak convergence". This is a basic concept. We present it now since its theory makes essential use of the uniform boundedness theorem discussed in the previous section. In fact, this is one of the major applications of that theorem.

Convergence of sequences of elements in a normed space was defined in Sec. 2.3 and, from now on, will be called *strong convergence*, to distinguish it from "weak convergence" to be introduced shortly. Hence we first state

4.8-1 Definition (Strong convergence). A sequence (x_n) in a normed space X is said to be *strongly convergent* (or *convergent in the norm*) if there is an $x \in X$ such that

$$\lim_{n\to\infty} \|x_n - x\| = 0.$$

This is written

$$\lim_{n\to\infty} x_n = x$$

or simply

$$x_n \longrightarrow x.$$

x is called the *strong limit* of (x_n), and we say that (x_n) *converges strongly to* x. ∎

Weak convergence is defined in terms of bounded linear functionals on X as follows.

4.8-2 Definition (Weak convergence). A sequence (x_n) in a normed space X is said to be *weakly convergent* if there is an $x \in X$ such that for every $f \in X'$,

$$\lim_{n\to\infty} f(x_n) = f(x).$$

This is written

$$x_n \xrightarrow{w} x$$

or $x_n \longrightarrow x$. The element x is called the *weak limit* of (x_n), and we say that (x_n) *converges weakly to* x. ∎

Note that weak convergence means convergence of the sequence of numbers $a_n = f(x_n)$ for every $f \in X'$.

Weak convergence has various applications throughout analysis (for instance, in the calculus of variations and the general theory of differential equations). The concept illustrates a basic principle of functional analysis, namely, the fact that the investigation of spaces is often related to that of their dual spaces.

For applying weak convergence one needs to know certain basic properties, which we state in the following lemma. The reader will note that in the proof we use the Hahn–Banach theorem (via 4.3-4 as well as 4.6-1) and the uniform boundedness theorem. This demonstrates the importance of these theorems in connection with weak convergence.

4.8-3 Lemma (Weak convergence). *Let (x_n) be a weakly convergent sequence in a normed space X, say, $x_n \xrightarrow{w} x$. Then:*

(a) *The weak limit x of (x_n) is unique.*

(b) *Every subsequence of (x_n) converges weakly to x.*

(c) *The sequence $(\|x_n\|)$ is bounded.*

Proof. **(a)** Suppose that $x_n \xrightarrow{w} x$ as well as $x_n \xrightarrow{w} y$. Then $f(x_n) \longrightarrow f(x)$ as well as $f(x_n) \longrightarrow f(y)$. Since $(f(x_n))$ is a sequence of numbers, its limit is unique. Hence $f(x) = f(y)$, that is, for every $f \in X'$ we have

$$f(x) - f(y) = f(x - y) = 0.$$

This implies $x - y = 0$ by Corollary 4.3-4 and shows that the weak limit is unique.

(b) follows from the fact that $(f(x_n))$ is a convergent sequence of numbers, so that every subsequence of $(f(x_n))$ converges and has the same limit as the sequence.

(c) Since $(f(x_n))$ is a convergent sequence of numbers, it is bounded, say, $|f(x_n)| \leqq c_f$ for all n, where c_f is a constant depending on f but not on n. Using the canonical mapping $C: X \longrightarrow X''$ (Sec. 4.6), we can define $g_n \in X''$ by

$$g_n(f) = f(x_n) \qquad f \in X'.$$

(We write g_n instead of g_{x_n}, to avoid subscripts of subscripts.) Then for all n,

$$|g_n(f)| = |f(x_n)| \leqq c_f,$$

that is, the sequence $(|g_n(f)|)$ is bounded for every $f \in X'$. Since X' is complete by 2.10-4, the uniform boundedness theorem 4.7-3 is applicable and implies that $(\|g_n\|)$ is bounded. Now $\|g_n\| = \|x_n\|$ by 4.6-1, so that (c) is proved. ∎

The reader may perhaps wonder why weak convergence does not play a role in calculus. The simple reason is that in finite dimensional

normed spaces the distinction between strong and weak convergence disappears completely. Let us prove this fact and also justify the terms "strong" and "weak."

4.8-4 Theorem (Strong and weak convergence). *Let (x_n) be a sequence in a normed space X. Then:*

(a) *Strong convergence implies weak convergence with the same limit.*

(b) *The converse of (a) is not generally true.*

(c) *If $\dim X < \infty$, then weak convergence implies strong convergence.*

Proof. **(a)** By definition, $x_n \longrightarrow x$ means $\|x_n - x\| \longrightarrow 0$ and implies that for every $f \in X'$,

$$|f(x_n) - f(x)| = |f(x_n - x)| \leqq \|f\| \, \|x_n - x\| \longrightarrow 0.$$

This shows that $x_n \xrightarrow{w} x$.

(b) can be seen from an orthonormal sequence (e_n) in a Hilbert space H. In fact, every $f \in H'$ has a Riesz representation $f(x) = \langle x, z \rangle$. Hence $f(e_n) = \langle e_n, z \rangle$. Now the Bessel inequality is (cf. 3.4-6)

$$\sum_{n=1}^{\infty} |\langle e_n, z \rangle|^2 \leqq \|z\|^2.$$

Hence the series on the left converges, so that its terms must approach zero as $n \longrightarrow \infty$. This implies

$$f(e_n) = \langle e_n, z \rangle \longrightarrow 0.$$

Since $f \in H'$ was arbitrary, we see that $e_n \xrightarrow{w} 0$. However, (e_n) does not converge strongly because

$$\|e_m - e_n\|^2 = \langle e_m - e_n, e_m - e_n \rangle = 2 \qquad (m \neq n).$$

(c) Suppose that $x_n \xrightarrow{w} x$ and $\dim X = k$. Let $\{e_1, \cdots, e_k\}$ be any basis for X and, say,

$$x_n = \alpha_1^{(n)} e_1 + \cdots + \alpha_k^{(n)} e_k$$

and

$$x = \alpha_1 e_1 + \cdots + \alpha_k e_k.$$

By assumption, $f(x_n) \longrightarrow f(x)$ for every $f \in X'$. We take in particular $f_1, \cdots, f_k$ defined by

$$f_j(e_j) = 1, \qquad f_j(e_m) = 0 \qquad (m \neq j).$$

(We mention that this is the dual basis of $\{e_1, \cdots, e_k\}$; cf. Sec. 2.9.) Then

$$f_j(x_n) = \alpha_j^{(n)}, \qquad f_j(x) = \alpha_j.$$

Hence $f_j(x_n) \longrightarrow f_j(x)$ implies $\alpha_j^{(n)} \longrightarrow \alpha_j$. From this we readily obtain

$$\|x_n - x\| = \left\| \sum_{j=1}^{k} (\alpha_j^{(n)} - \alpha_j) e_j \right\|$$

$$\leqq \sum_{j=1}^{k} |\alpha_j^{(n)} - \alpha_j| \, \|e_j\| \longrightarrow 0$$

as $n \longrightarrow \infty$. This shows that (x_n) converges strongly to x. ∎

It is interesting to note that there also exist infinite dimensional spaces such that strong and weak convergence are equivalent concepts. An example is l^1, as was shown by I. Schur (1921).

In conclusion let us take a look at weak convergence in two particularly important types of spaces.

Examples

4.8-5 Hilbert space. *In a Hilbert space,* $x_n \xrightarrow{w} x$ *if and only if* $\langle x_n, z \rangle \longrightarrow \langle x, z \rangle$ *for all* z *in the space.*

Proof. Clear by 3.8-1.

4.8-6 Space l^p. *In the space l^p, where $1<p<+\infty$, we have $x_n \xrightarrow{w} x$ if and only if:*

(A) *The sequence $(\|x_n\|)$ is bounded.*

(B) *For every fixed j we have $\xi_j^{(n)} \longrightarrow \xi_j$ as $n \longrightarrow \infty$; here, $x_n=(\xi_j^{(n)})$ and $x=(\xi_j)$.*

Proof. The dual space of l^p is l^q; cf. 2.10-7. A Schauder basis of l^q is (e_n), where $e_n=(\delta_{nj})$ has 1 in the nth place and zeros elsewhere. Span (e_n) is dense in l^q, so that the conclusion results from the following lemma.

4.8-7 Lemma (Weak convergence). *In a normed space X we have $x_n \xrightarrow{w} x$ if and only if:*

(A) *The sequence $(\|x_n\|)$ is bounded.*

(B) *For every element f of a total subset $M \subset X'$ we have $f(x_n) \longrightarrow f(x)$.*

Proof. In the case of weak convergence, (A) follows from Lemma 4.8-3 and (B) is trivial.

Conversely, suppose that (A) and (B) hold. Let us consider any $f \in X'$ and show that $f(x_n) \longrightarrow f(x)$, which means weak convergence, by the definition.

By (A) we have $\|x_n\| \leqq c$ for all n and $\|x\| \leqq c$, where c is sufficiently large. Since M is total in X', for every $f \in X'$ there is a sequence (f_j) in span M such that $f_j \longrightarrow f$. Hence for any given $\varepsilon>0$ we can find a j such that

$$\|f_j-f\|<\frac{\varepsilon}{3c}.$$

Moreover, since $f_j \in \operatorname{span} M$, by assumption (B) there is an N such that for all $n>N$,

$$|f_j(x_n)-f_j(x)|<\frac{\varepsilon}{3}.$$

Using these two inequalities and applying the triangle inequality, we

obtain for $n > N$

$$|f(x_n) - f(x)| \leqq |f(x_n) - f_j(x_n)| + |f_j(x_n) - f_j(x)| + |f_j(x) - f(x)|$$

$$< \|f - f_j\| \, \|x_n\| + \frac{\varepsilon}{3} + \|f_j - f\| \, \|x\|$$

$$< \frac{\varepsilon}{3c} c + \frac{\varepsilon}{3} + \frac{\varepsilon}{3c} c = \varepsilon.$$

Since $f \in X'$ was arbitrary, this shows that the sequence (x_n) converges weakly to x. ∎

Problems

1. **(Pointwise convergence)** If $x_n \in C[a, b]$ and $x_n \xrightarrow{w} x \in C[a, b]$, show that (x_n) is *pointwise convergent* on $[a, b]$, that is, $(x_n(t))$ converges for every $t \in [a, b]$.

2. Let X and Y be normed spaces, $T \in B(X, Y)$ and (x_n) a sequence in X. If $x_n \xrightarrow{w} x_0$, show that $Tx_n \xrightarrow{w} Tx_0$.

3. If (x_n) and (y_n) are sequences in the same normed space X, show that $x_n \xrightarrow{w} x$ and $y_n \xrightarrow{w} y$ implies $x_n + y_n \xrightarrow{w} x + y$ as well as $\alpha x_n \xrightarrow{w} \alpha x$, where α is any scalar.

4. Show that $x_n \xrightarrow{w} x_0$ implies $\lim_{n \to \infty} \|x_n\| \geqq \|x_0\|$. (Use Theorem 4.3-3.)

5. If $x_n \xrightarrow{w} x_0$ in a normed space X, show that $x_0 \in \bar{Y}$, where $Y = \operatorname{span}(x_n)$. (Use Lemma 4.6-7.)

6. If (x_n) is a weakly convergent sequence in a normed space X, say, $x_n \xrightarrow{w} x_0$, show that there is a sequence (y_m) of linear combinations of elements of (x_n) which converges strongly to x_0.

7. Show that any closed subspace Y of a normed space X contains the limits of all weakly convergent sequences of its elements.

8. **(Weak Cauchy sequence)** A *weak Cauchy sequence* in a real or complex normed space X is a sequence (x_n) in X such that for every $f \in X'$ the sequence $(f(x_n))$ is Cauchy in $\mathbf{R}$ or $\mathbf{C}$, respectively. [Note that then $\lim_{n \to \infty} f(x_n)$ exists.] Show that a weak Cauchy sequence is bounded.

9. Let A be a set in a normed space X such that every nonempty subset of A contains a weak Cauchy sequence. Show that A is bounded.

10. (Weak completeness) A normed space X is said to be *weakly complete* if each weak Cauchy sequence in X converges weakly in X. If X is reflexive, show that X is weakly complete.

4.9 Convergence of Sequences of Operators and Functionals

Sequences of bounded linear operators and functionals arise frequently in the abstract formulation of concrete situations, for instance in connection with convergence problems of Fourier series or sequences of interpolation polynomials or methods of numerical integration, to name just a few. In such cases one is usually concerned with the convergence of those sequences of operators or functionals, with boundedness of corresponding sequences of norms or with similar properties.

Experience shows that for sequences of *elements* in a normed space, strong and weak convergence as defined in the previous section are useful concepts. For sequences of *operators* $T_n \in B(X, Y)$ three types of convergence turn out to be of theoretical as well as practical value. These are

(1) Convergence in the norm on $B(X, Y)$,
(2) Strong convergence of $(T_n x)$ in Y,
(3) Weak convergence of $(T_n x)$ in Y.

The definitions and terminology are as follows; they were introduced by J. von Neumann (1929–30b).

4.9-1 Definition (Convergence of sequences of operators). Let X and Y be normed spaces. A sequence (T_n) of operators $T_n \in B(X, Y)$ is said to be:

(1) **uniformly operator convergent**[5] if (T_n) converges in the norm on $B(X, Y)$

[5] "Operator" is often omitted from each of the three terms. We retain it for clarity.

(2) **strongly operator convergent** if $(T_n x)$ converges strongly in Y for every $x \in X$,

(3) **weakly operator convergent** if $(T_n x)$ converges weakly in Y for every $x \in X$.

In formulas this means that there is an operator $T: X \longrightarrow Y$ such that

(1) $\|T_n - T\| \longrightarrow 0$

(2) $\|T_n x - Tx\| \longrightarrow 0$ for all $x \in X$

(3) $|f(T_n x) - f(Tx)| \longrightarrow 0$ for all $x \in X$ and all $f \in Y'$

respectively. T is called the *uniform*, *strong* and *weak operator limit* of (T_n), respectively. ∎

We pointed out in the previous section that even in calculus, in a much simpler situation, the use of several concepts of convergence gives greater flexibility. Nevertheless the reader may still be bewildered by the many concepts of convergence we have just introduced. He may ask why it is necessary to have three kinds of convergence for sequences of operators. The answer is that many of the operators that appear in practical problems are given as some sort of limit of simpler operators. And it is important to know what is meant by "some sort" and to know what properties of the limiting operator are implied by the properties of the sequence. Also, at the beginning of an investigation, one does not always know in what sense limits will exist; hence it is useful to have a variety of possibilities. Perhaps in a specific problem one is at first able to establish convergence only in a very "mild" sense, so that one has at least something to start from, and then later develop tools for proving convergence in a stronger sense, which guarantees "better" properties of the limit operator. This is a typical situation, for example in partial differential equations.

It is not difficult to show that

$$(1) \quad \Longrightarrow \quad (2) \quad \Longrightarrow \quad (3)$$

(the limit being the same), but the converse is not generally true, as can be seen from the following examples.

Examples

4.9-2 (Space l^2). In the space l^2 we consider a sequence (T_n), where $T_n: l^2 \longrightarrow l^2$ is defined by

$$T_n x = (\underbrace{0, 0, \cdots, 0}_{(n \text{ zeros})}, \xi_{n+1}, \xi_{n+2}, \xi_{n+3}, \cdots);$$

here, $x = (\xi_1, \xi_2, \cdots) \in l^2$. This operator T_n is linear and bounded. Clearly, (T_n) is strongly operator convergent to 0 since $T_n x \longrightarrow 0 = 0x$. However, (T_n) is not uniformly operator convergent since we have $\|T_n - 0\| = \|T_n\| = 1$.

4.9-3 (Space l^2). Another sequence (T_n) of operators $T_n: l^2 \longrightarrow l^2$ is defined by

$$T_n x = (\underbrace{0, 0, \cdots, 0}_{(n \text{ zeros})}, \xi_1, \xi_2, \xi_3, \cdots)$$

where $x = (\xi_1, \xi_2, \cdots) \in l^2$. This operator T_n is linear and bounded. We show that (T_n) is weakly operator convergent to 0 but not strongly.

Every bounded linear functional f on l^2 has a Riesz representation 3.8-1, that is, by 3.1-6,

$$f(x) = \langle x, z \rangle = \sum_{j=1}^{\infty} \xi_j \bar{\zeta}_j$$

where $z = (\zeta_j) \in l^2$. Hence, setting $j = n + k$ and using the definition of T_n, we have

$$f(T_n x) = \langle T_n x, z \rangle = \sum_{j=n+1}^{\infty} \xi_{j-n} \bar{\zeta}_j = \sum_{k=1}^{\infty} \xi_k \bar{\zeta}_{n+k}.$$

By the Cauchy–Schwarz inequality in 1.2-3,

$$|f(T_n x)|^2 = |\langle T_n x, z \rangle|^2 \leqq \sum_{k=1}^{\infty} |\xi_k|^2 \sum_{m=n+1}^{\infty} |\zeta_m|^2.$$

The last series is the remainder of a convergent series. Hence the right-hand side approaches 0 as $n \longrightarrow \infty$; thus $f(T_n x) \longrightarrow 0 = f(0x)$. Consequently, (T_n) is weakly operator convergent to 0.

However, (T_n) is not strongly operator convergent because for $x = (1, 0, 0, \cdots)$ we have

$$\|T_m x - T_n x\| = \sqrt{1^2 + 1^2} = \sqrt{2} \qquad (m \neq n). \quad \blacksquare$$

Linear functionals are linear operators (with range in the scalar field $\mathbf{R}$ or $\mathbf{C}$), so that (1), (2) and (3) apply immediately. However, (2) and (3) now become equivalent, for the following reason. We had $T_n x \in Y$, but we now have $f_n(x) \in \mathbf{R}$ (or $\mathbf{C}$). Hence convergence in (2) and (3) now takes place in the finite dimensional (one-dimensional) space $\mathbf{R}$ (or $\mathbf{C}$) and equivalence of (2) and (3) follows from Theorem 4.8-4(*c*). The two remaining concepts are called *strong* and *weak* convergence* (read "weak star convergence"):

4.9-4 Definition (Strong and weak* convergence of a sequence of functionals). Let (f_n) be a sequence of bounded linear functionals on a normed space X. Then:

(a) *Strong convergence* of (f_n) means that there is an $f \in X'$ such that $\|f_n - f\| \longrightarrow 0$. This is written

$$f_n \longrightarrow f.$$

(b) *Weak* convergence* of (f_n) means that there is an $f \in X'$ such that $f_n(x) \longrightarrow f(x)$ for all $x \in X$. This is written[6]

$$f_n \xrightarrow{w^*} f.$$

f in (*a*) and (*b*) is called the *strong limit* and *weak* limit* of (f_n), respectively. $\blacksquare$

Returning to operators $T_n \in B(X, Y)$, we ask what can be said about the limit operator T: $X \longrightarrow Y$ in (1), (2) and (3).

If the convergence is uniform, $T \in B(X, Y)$; otherwise $\|T_n - T\|$ would not make sense. If the convergence is strong or weak, T is still linear but may be unbounded if X is not complete.

[6] This concept is somewhat more important than *weak convergence* of (f_n), which, by 4.8-2, means that $g(f_n) \longrightarrow g(f)$ for all $g \in X''$. Weak convergence implies weak* convergence, as can be seen by the use of the canonical mapping defined in Sec. 4.6. (Cf. Prob. 4.)

Example. The space X of all sequences $x=(\xi_j)$ in l^2 with only finitely many nonzero terms, taken with the metric on l^2, is not complete. A sequence of bounded linear operators T_n on X is defined by

$$T_n x=(\xi_1, 2\xi_2, 3\xi_3, \cdots, n\xi_n, \xi_{n+1}, \xi_{n+2}, \cdots),$$

so that $T_n x$ has terms $j\xi_j$ if $j\leqq n$ and ξ_j if $j>n$. This sequence (T_n) converges strongly to the unbounded linear operator T defined by $Tx=(\eta_j)$, where $\eta_j=j\xi_j$.

However, if X is complete, the situation illustrated by this example cannot occur since then we have the basic

4.9-5 Lemma (Strong operator convergence). *Let $T_n\in B(X, Y)$, where X is a Banach space and Y a normed space. If (T_n) is strongly operator convergent with limit T, then $T\in B(X, Y)$.*

Proof. Linearity of T follows readily from that of T_n. Since $T_n x\longrightarrow Tx$ for every $x\in X$, the sequence $(T_n x)$ is bounded for every x; cf. 1.4-2. Since X is complete, $(\|T_n\|)$ is bounded by the uniform boundedness theorem, say, $\|T_n\|\leqq c$ for all n. From this, it follows that $\|T_n x\|\leqq\|T_n\|\,\|x\|\leqq c\,\|x\|$. This implies $\|Tx\|\leqq c\,\|x\|$. ∎

A useful criterion for strong operator convergence is

4.9-6 Theorem (Strong operator convergence). *A sequence (T_n) of operators $T_n\in B(X, Y)$, where X and Y are Banach spaces, is strongly operator convergent if and only if:*

(*A*) *The sequence $(\|T_n\|)$ is bounded.*
(*B*) *The sequence $(T_n x)$ is Cauchy in Y for every x in a total subset M of X.*

Proof. If $T_n x\longrightarrow Tx$ for every $x\in X$, then (*A*) follows from the uniform boundedness theorem (since X is complete) and (*B*) is trivial.

Conversely, suppose that (*A*) and (*B*) hold, so that, say, $\|T_n\|\leqq c$ for all n. We consider any $x\in X$ and show that $(T_n x)$ converges strongly in Y. Let $\varepsilon>0$ be given. Since span M is dense in X, there is a $y\in$ span M such that

$$\|x-y\|<\frac{\varepsilon}{3c}.$$

Since $y \in \operatorname{span} M$, the sequence $(T_n y)$ is Cauchy by (B). Hence there is an N such that

$$\|T_n y - T_m y\| < \frac{\varepsilon}{3} \qquad (m, n > N).$$

Using these two inequalities and applying the triangle inequality, we readily see that $(T_n x)$ is Cauchy in Y because for $m, n > N$ we obtain

$$\|T_n x - T_m x\| \leqq \|T_n x - T_n y\| + \|T_n y - T_m y\| + \|T_m y - T_m x\|$$

$$< \|T_n\| \, \|x - y\| + \frac{\varepsilon}{3} + \|T_m\| \, \|x - y\|$$

$$< c\frac{\varepsilon}{3c} + \frac{\varepsilon}{3} + c\frac{\varepsilon}{3c} = \varepsilon.$$

Since Y is complete, $(T_n x)$ converges in Y. Since $x \in X$ was arbitrary, this proves strong operator convergence of (T_n). ∎

4.9-7 Corollary (Functionals). *A sequence (f_n) of bounded linear functionals on a Banach space X is weak* convergent, the limit being a bounded linear functional on X, if and only if:*

(A) The sequence $(\|f_n\|)$ is bounded.
(B) The sequence $(f_n(x))$ is Cauchy for every x in a total subset M of X.

This has interesting applications. Two of them will be discussed in the next sections.

Problems

1. Show that uniform operator convergence $T_n \longrightarrow T$, $T_n \in B(X, Y)$, implies strong operator convergence with the same limit T.

2. If $S_n, T_n \in B(X, Y)$, and (S_n) and (T_n) are strongly operator convergent with limits S and T, show that $(S_n + T_n)$ is strongly operator convergent with the limit $S + T$.

3. Show that strong operator convergence in $B(X, Y)$ implies weak operator convergence with the same limit.

4. Show that weak convergence in footnote 6 implies weak* convergence. Show that the converse holds if X is reflexive.

5. Strong operator convergence does not imply uniform operator convergence. Illustrate this by considering $T_n = f_n\colon l^1 \longrightarrow \mathbf{R}$, where $f_n(x) = \xi_n$ and $x = (\xi_n)$.

6. Let $T_n \in B(X, Y)$, where $n = 1, 2, \cdots$. To motivate the term "uniform" in Def. 4.9-1, show that $T_n \longrightarrow T$ if and only if for every $\varepsilon > 0$ there is an N, *depending only on* ε, such that for all $n > N$ and all $x \in X$ of norm 1 we have

$$\|T_n x - Tx\| < \varepsilon.$$

7. Let $T_n \in B(X, Y)$, where X is a Banach space. If (T_n) is strongly operator convergent, show that $(\|T_n\|)$ is bounded.

8. Let $T_n \longrightarrow T$, where $T_n \in B(X, Y)$. Show that for every $\varepsilon > 0$ and every closed ball $K \subset X$ there is an N such that $\|T_n x - Tx\| < \varepsilon$ for all $n > N$ and all $x \in K$.

9. Show that $\|T\| \leqq \varliminf_{n\to\infty} \|T_n\|$ in Lemma 4.9-5.

10. Let X be a separable Banach space and $M \subset X'$ a bounded set. Show that every sequence of elements of M contains a subsequence which is weak* convergent to an element of X'.

4.10 Application to Summability of Sequences

Weak* convergence has important applications in the theory of divergent sequences (and series). A divergent sequence has no limit in the usual sense. In that theory, one aims at associating with certain divergent sequences a "limit" in a generalized sense. A procedure for that purpose is called a *summability method.*

For instance, a divergent sequence $x = (\xi_k)$ being given, we may calculate the sequence $y = (\eta_n)$ of the arithmetic means

$$\eta_1 = \xi_1, \qquad \eta_2 = \frac{1}{2}(\xi_1 + \xi_2), \quad \cdots, \quad \eta_n = \frac{1}{n}(\xi_1 + \cdots + \xi_n), \quad \cdots.$$

This is an example of a summability method. If y converges with limit η (in the usual sense), we say that x is *summable* by the present method and has the *generalized limit* η. For instance, if

$$x = (0, 1, 0, 1, 0, \cdots) \qquad \text{then} \qquad y = (0, \tfrac{1}{2}, \tfrac{1}{3}, \tfrac{1}{2}, \tfrac{2}{5}, \cdots)$$

and x has the generalized limit $\frac{1}{2}$.

A summability method is called a *matrix method* if it can be represented in the form

$$y = Ax$$

where $x = (\xi_k)$ and $y = (\eta_n)$ are written as infinite column vectors and $A = (\alpha_{nk})$ is an infinite matrix; here, n, $k = 1, 2, \cdots$. In the formula $y = Ax$ we used matrix multiplication, that is, y has the terms

$$\eta_n = \sum_{k=1}^{\infty} \alpha_{nk}\xi_k \qquad n = 1, 2, \cdots. \tag{1}$$

The above example illustrates a matrix method. (What is the matrix?)

Relevant terms are as follows. The method given by (1) is briefly called an *A-method* because the corresponding matrix is denoted by A. If the series in (1) all converge and $y = (\eta_n)$ converges in the usual sense, its limit is called the *A-limit* of x, and x is said to be *A-summable.* The set of all A-summable sequences is called the *range* of the A-method.

An A-method is said to be **regular** (or *permanent*) if its range includes all convergent sequences and if for every such sequence the A-limit equals the usual limit, that is, if

$$\xi_k \longrightarrow \xi \qquad \text{implies} \qquad \eta_n \longrightarrow \xi.$$

Obviously, regularity is a rather natural requirement. In fact, a method which is not applicable to certain convergent sequences or alters their limit would be of no practical use. A basic criterion for regularity is as follows.

4.10-1 Toeplitz Limit Theorem (Regular summability methods). *An A-summability method with matrix $A = (\alpha_{nk})$ is regular if and*

only if

$$\lim_{n\to\infty} \alpha_{nk} = 0 \qquad \text{for } k = 1, 2, \cdots \tag{2}$$

$$\lim_{n\to\infty} \sum_{k=1}^{\infty} \alpha_{nk} = 1 \tag{3}$$

$$\sum_{k=1}^{\infty} |\alpha_{nk}| \leqq \gamma \qquad \text{for } n = 1, 2, \cdots \tag{4}$$

where γ is a constant which does not depend on n.

Proof. We show that
(*a*) (2) to (4) are necessary for regularity,
(*b*) (2) to (4) are sufficient for regularity.
The details are as follows.

(a) Suppose that the A-method is regular. Let x_k have 1 as the kth term and all other terms zero. For x_k we have $\eta_n = \alpha_{nk}$ in (1). Since x_k is convergent and has the limit 0, this shows that (2) must hold.

Furthermore, $x = (1, 1, 1, \cdots)$ has the limit 1. And from (1) we see that η_n now equals the series in (3). Consequently, (3) must hold.

We prove that (4) is necessary for regularity. Let c be the Banach space of all convergent sequences with norm defined by

$$\|x\| = \sup_j |\xi_j|,$$

cf. 1.5-3. Linear functionals f_{nm} on c are defined by

$$f_{nm}(x) = \sum_{k=1}^{m} \alpha_{nk}\xi_k \qquad m, n = 1, 2, \cdots \tag{5}$$

Each f_{nm} is bounded since

$$|f_{nm}(x)| \leqq \sup_j |\xi_j| \sum_{k=1}^{m} |\alpha_{nk}| = \left(\sum_{k=1}^{m} |\alpha_{nk}|\right)\|x\|.$$

Regularity implies the convergence of the series in (1) for all $x \in c$.

Hence (1) defines linear functionals $f_1, f_2, \cdots$ on c given by

$$\eta_n = f_n(x) = \sum_{k=1}^{\infty} \alpha_{nk}\xi_k \qquad n = 1, 2, \cdots. \tag{6}$$

From (5) we see that $f_{nm}(x) \longrightarrow f_n(x)$ as $m \longrightarrow \infty$ for all $x \in c$. This is weak* convergence, and f_n is bounded by Lemma 4.9-5 (with $T = f_n$). Also, $(f_n(x))$ converges for all $x \in c$, and $(\|f_n\|)$ is bounded by Corollary 4.9-7, say,

$$\|f_n\| \leqq \gamma \qquad \text{for all } n. \tag{7}$$

For an arbitrary fixed $m \in \mathbf{N}$ define

$$\xi_k^{(n,m)} = \begin{cases} |\alpha_{nk}|/\alpha_{nk} & \text{if } k \leqq m \text{ and } \alpha_{nk} \neq 0 \\ 0 & \text{if } k > m \text{ or } \alpha_{nk} = 0. \end{cases}$$

Then we have $x_{nm} = (\xi_k^{(n,m)}) \in c$. Also $\|x_{nm}\| = 1$ if $x_{nm} \neq 0$ and $\|x_{nm}\| = 0$ if $x_{nm} = 0$. Furthermore,

$$f_{nm}(x_{nm}) = \sum_{k=1}^{m} \alpha_{nk}\xi_k^{(n,m)} = \sum_{k=1}^{m} |\alpha_{nk}|$$

for all m. Hence

$$\text{(a)} \qquad \sum_{k=1}^{m} |\alpha_{nk}| = f_{nm}(x_{nm}) \leqq \|f_{nm}\| \tag{8}$$

$$\text{(b)} \qquad \sum_{k=1}^{\infty} |\alpha_{nk}| \leqq \|f_n\|.$$

This shows that the series in (4) converges, and (4) follows from (7).

(b) We prove that (2) to (4) is sufficient for regularity. We define a linear functional f on c by

$$f(x) = \xi = \lim_{k \to \infty} \xi_k .$$

where $x = (\xi_k) \in c$. Boundedness of f can be seen from

$$|f(x)| = |\xi| \leqq \sup_j |\xi_j| = \|x\|.$$

Let $M \subset c$ be the set of all sequences whose terms are equal from some term on, say, $x = (\xi_k)$ where

$$\xi_j = \xi_{j+1} = \xi_{j+2} = \cdots = \xi,$$

and j depends on x. Then $f(x) = \xi$, as above, and in (1) and (6) we obtain

$$\eta_n = f_n(x) = \sum_{k=1}^{j-1} \alpha_{nk}\xi_k + \xi \sum_{k=j}^{\infty} \alpha_{nk}$$

$$= \sum_{k=1}^{j-1} \alpha_{nk}(\xi_k - \xi) + \xi \sum_{k=1}^{\infty} \alpha_{nk}.$$

Hence by (2) and (3),

$$\eta_n = f_n(x) \longrightarrow 0 + \xi \cdot 1 = \xi = f(x) \tag{9}$$

for every $x \in M$.

We want to use Corollary 4.9-7 again. Hence we show next that the set M on which we have the convergence expressed in (9) is dense in c. Let $x = (\xi_k) \in c$ with $\xi_k \longrightarrow \xi$. Then for every $\varepsilon > 0$ there is an N such that

$$|\xi_k - \xi| < \varepsilon \qquad \text{for } k \geqq N.$$

Clearly,

$$\tilde{x} = (\xi_1, \cdots, \xi_{N-1}, \xi, \xi, \xi, \cdots) \in M$$

and

$$x - \tilde{x} = (0, \cdots, 0, \xi_N - \xi, \xi_{N+1} - \xi, \cdots).$$

It follows that $\|x - \tilde{x}\| \leqq \varepsilon$. Since $x \in c$ was arbitrary, this shows that M is dense in c.

Finally, by (4),

$$|f_n(x)| \leqq \|x\| \sum_{k=1}^{\infty} |\alpha_{nk}| \leqq \gamma \|x\|$$

for every $x \in c$ and all n. Hence $\|f_n\| \leqq \gamma$, that is, $(\|f_n\|)$ is bounded. Furthermore, (9) means convergence $f_n(x) \longrightarrow f(x)$ for all x in the dense set M. By Corollary 4.9-7 this implies weak* convergence $f_n \xrightarrow{w^*} f$. Thus we have shown that if $\xi = \lim \xi_k$ exists, it follows that $\eta_n \longrightarrow \xi$. By definition, this means regularity and the theorem is proved. ∎

Problems

1. **Cesàro's summability method C_1** is defined by

$$\eta_n = \frac{1}{n}(\xi_1 + \cdots + \xi_n) \qquad n = 1, 2, \cdots,$$

that is, one takes arithmetic means. Find the corresponding matrix A.

2. Apply the method C_1 in Prob. 1 to the sequences

$$(1, 0, 1, 0, 1, 0, \cdots) \quad \text{and} \quad \left(1, 0, -\frac{1}{4}, -\frac{2}{8}, -\frac{3}{16}, -\frac{4}{32}, \cdots\right).$$

3. In Prob. 1, express (ξ_n) in terms of (η_n). Find (ξ_n) such that $(\eta_n) = (1/n)$.

4. Use the formula in Prob. 3 for obtaining a sequence which is not C_1-summable.

5. **Hölder's summability methods H_p** are defined as follows. H_1 is identical with C_1 in Prob. 1. The method H_2 consists of two successive applications of H_1, that is, we first take the arithmetic means and then again the arithmetic means of those means. H_3 consists of three successive applications of H_1, etc. Apply H_1 and H_2 to the sequence $(1, -3, 5, -7, 9, -11, \cdots)$. Comment.

6. **(Series)** An infinite series is said to be *A-summable* if the sequence of its partial sums is A-summable, and the A-limit of that sequence is called the *A-sum* of the series. Show that $1 + z + z^2 + \cdots$ is C_1-summable for $|z| = 1$, $z \neq 1$, and the C_1-sum is $1/(1-z)$.

7. (Cesàro's C_k-method) Given (ξ_n), let $\sigma_n^{(0)} = \xi_n$ and

$$\sigma_n^{(k)} = \sigma_0^{(k-1)} + \sigma_1^{(k-1)} + \cdots + \sigma_n^{(k-1)} \qquad (k \geqq 1,\ n = 0, 1, 2, \cdots).$$

If for a fixed $k \in \mathbf{N}$ we have $\eta_n^{(k)} = \sigma_n^{(k)} / \binom{n+k}{k} \longrightarrow \eta$, we say that (ξ_n) is *C_k-summable* and has the *C_k-limit* η. Show that the method has the advantage that $\sigma_n^{(k)}$ can be represented in terms of the ξ_j's in a very simple fashion, namely

$$\sigma_n^{(k)} = \sum_{\nu=0}^{n} \binom{n+k-1-\nu}{k-1} \xi_\nu.$$

8. Euler's method for series associates with a given series

$$\sum_{j=0}^{\infty} (-1)^j a_j \qquad \text{the transformed series} \qquad \sum_{n=0}^{\infty} \frac{\Delta^n a_0}{2^{n+1}}$$

where

$$\Delta^0 a_j = a_j, \qquad \Delta^n a_j = \Delta^{n-1} a_j - \Delta^{n-1} a_{j+1}, \qquad j = 1, 2, \cdots,$$

and $(-1)^j$ is written for convenience (hence the a_j need not be positive). It can be shown that the method is regular, so that the convergence of the given series implies that of the transformed series, the sum being the same. Show that the method gives

$$\ln 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + - \cdots = \frac{1}{1 \cdot 2^1} + \frac{1}{2 \cdot 2^2} + \frac{1}{3 \cdot 2^3} + \frac{1}{4 \cdot 2^4} + \cdots.$$

9. Show that Euler's method in Prob. 8 yields

$$\frac{\pi}{4} = \arctan 1 = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + - \cdots = \frac{1}{2}\left(1 + \frac{1}{3} + \frac{1 \cdot 2}{3 \cdot 5} + \frac{1 \cdot 2 \cdot 3}{3 \cdot 5 \cdot 7} + \cdots\right).$$

10. Show that Euler's method yields the following result. Comment.

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{4^n} = \frac{1}{2} \sum_{n=0}^{\infty} \left(\frac{3}{8}\right)^n.$$

4.11 Numerical Integration and Weak* Convergence

Weak* convergence has useful applications to numerical integration, differentiation and interpolation. In this section we consider numerical integration, that is, the problem of obtaining approximate values for a given integral

$$\int_a^b x(t)\,dt.$$

Since the problem is important in applications, various methods have been developed for that purpose, for example the trapezoidal rule, Simpson's rule and more complicated formulas by Newton-Cotes and Gauss. (For a review of some elementary facts, see the problem set at the end of the section.)

The common feature of those and other methods is that we first choose points in $[a, b]$, called *nodes*, and then approximate the unknown value of the integral by a linear combination of the values of x at the nodes. The nodes and the coefficients of that linear combination depend on the method but not on the integrand x. Of course, the usefulness of a method is largely determined by its accuracy, and one may want the accuracy to increase as the number of nodes gets larger.

In this section we shall see that functional analysis can offer help in that respect. In fact, we shall describe a general setting for those methods and consider the problem of convergence as the number of nodes increases.

We shall be concerned with continuous functions. This suggests introducing the Banach space $X = C[a, b]$ of all continuous real-valued functions on $J = [a, b]$, with norm defined by

$$\|x\| = \max_{t \in J} |x(t)|.$$

Then the above definite integral defines a linear functional f on X by means of

$$f(x) = \int_a^b x(t)\,dt. \tag{1}$$

To obtain a formula for numerical integration, we may proceed as in those aforementioned methods. Thus, for each positive integer n we

choose $n+1$ real numbers

$$t_0^{(n)}, \cdots, t_n^{(n)} \qquad \text{(called } \textbf{nodes}\text{)}$$

such that

$$a \leqq t_0^{(n)} < \cdots < t_n^{(n)} \leqq b. \tag{2}$$

Then we choose $n+1$ real numbers

$$\alpha_0^{(n)}, \cdots, \alpha_n^{(n)} \qquad \text{(called } \textbf{coefficients}\text{)}$$

and define linear functionals f_n on X by setting

$$f_n(x) = \sum_{k=0}^{n} \alpha_k^{(n)} x(t_k^{(n)}) \qquad n = 1, 2, \cdots. \tag{3}$$

This defines a numerical process of integration, the value $f_n(x)$ being an approximation to $f(x)$, where x is given. To find out about the accuracy of the process, we consider the f_n's as follows.

Each f_n is bounded since $|x(t_k^{(n)})| \leqq \|x\|$ by the definition of the norm. Consequently,

$$|f_n(x)| \leqq \sum_{k=0}^{n} |\alpha_k^{(n)}| |x(t_k^{(n)})| \leqq \left(\sum_{k=0}^{n} |\alpha_k^{(n)}| \right) \|x\|. \tag{4}$$

For later use we show that f_n has the norm

$$\|f_n\| = \sum_{k=0}^{n} |\alpha_k^{(n)}|. \tag{5}$$

Indeed, (4) shows that $\|f_n\|$ cannot exceed the right-hand side of (5), and equality follows if we take an $x_0 \in X$ such that $|x_0(t)| \leqq 1$ on J and

$$x_0(t_k^{(n)}) = \operatorname{sgn} \alpha_k^{(n)} = \begin{cases} 1 & \text{if } \alpha_k^{(n)} \geqq 0 \\ -1 & \text{if } \alpha_k^{(n)} < 0 \end{cases}$$

since then $\|x_0\| = 1$ and

$$f_n(x_0) = \sum_{k=0}^{n} \alpha_k^{(n)} \operatorname{sgn} \alpha_k^{(n)} = \sum_{k=0}^{n} |\alpha_k^{(n)}|.$$

For a given $x \in X$, formula (3) yields an approximate value $f_n(x)$ of $f(x)$ in (1). Of course, we are interested in the accuracy, as was mentioned before, and we want it to increase with increasing n. This suggests the following concept.

4.11-1 Definition (Convergence). The numerical process of integration defined by (3) is said to be *convergent* for an $x \in X$ if for that x,

$$f_n(x) \longrightarrow f(x) \qquad (n \longrightarrow \infty), \tag{6}$$

where f is defined by (1). ∎

Furthermore, since exact integration of polynomials is easy, it is natural to make the following

4.11-2 Requirement. *For every n, if x is a polynomial of degree not exceeding n, then*

$$f_n(x) = f(x). \tag{7}$$ ∎

Since the f_n's are linear, it suffices to require (7) for the $n+1$ powers defined by

$$x_0(t) = 1, \qquad x_1(t) = t, \qquad \cdots, \qquad x_n(t) = t^n.$$

In fact, then for a polynomial of degree n given by $x(t) = \sum \beta_j t^j$ we obtain

$$f_n(x) = \sum_{j=0}^{n} \beta_j f_n(x_j) = \sum_{j=0}^{n} \beta_j f(x_j) = f(x).$$

We see that we thus have the $n+1$ conditions

$$f_n(x_j) = f(x_j) \qquad j = 0, \cdots, n. \tag{8}$$

We show that these conditions can be fulfilled. $2n+2$ parameters are available, namely, $n+1$ nodes and $n+1$ coefficients. Hence we can choose some of them in an arbitrary fashion. Let us choose the nodes $t_k^{(n)}$, and let us prove that we can then determine those coefficients uniquely.

In (8) we now have $x_j(t_k^{(n)}) = (t_k^{(n)})^j$ so that (8) takes the form

$$\sum_{k=0}^{n} \alpha_k^{(n)} (t_k^{(n)})^j = \int_a^b t^j \, dt = \frac{1}{j+1} (b^{j+1} - a^{j+1}) \tag{9}$$

where $j = 0, \cdots, n$. For each fixed n this is a nonhomogeneous system of $n+1$ linear equations in the $n+1$ unknowns $\alpha_0^{(n)}, \cdots, \alpha_n^{(n)}$. A unique solution exists if the corresponding homogeneous system

$$\sum_{k=0}^{n} (t_k^{(n)})^j \gamma_k = 0 \qquad (j = 0, \cdots, n)$$

has only the trivial solution $\gamma_0 = 0, \cdots, \gamma_n = 0$ or, equally well, if the same holds for the system

$$\sum_{j=0}^{n} (t_k^{(n)})^j \gamma_j = 0 \qquad (k = 0, \cdots, n) \tag{10}$$

whose coefficient matrix is the transpose of the coefficient matrix of the previous system. This holds, since (10) means that the polynomial

$$\sum_{j=0}^{n} \gamma_j t^j$$

which is of degree n, is zero at the $n+1$ nodes, hence it must be identically zero, that is, all its coefficients γ_j are zero. ∎

Our result is that for every choice of nodes satisfying (2) there are uniquely determined coefficients such that 4.11-2 holds; hence the corresponding process is convergent for all polynomials. And we ask what additional conditions we should impose in order that the process is convergent for all real-valued continuous functions on $[a, b]$. A corresponding criterion was given by G. Pólya (1933):

4.11-3 Pólya Convergence Theorem (Numerical integration). *A process of numerical integration* (3) *which satisfies* 4.11-2 *converges for all real-valued continuous functions on* $[a, b]$ *if and only if there is a number* c *such that*

$$\sum_{k=0}^{n} |\alpha_k^{(n)}| \leqq c \qquad \text{for all } n. \tag{11}$$

Proof. The set W of all polynomials with real coefficients is dense in the real space $X = C[a, b]$, by the Weierstrass approximation theorem (proof below), and for every $x \in W$ we have convergence by 4.11-2. From (5) we see that $(\|f_n\|)$ is bounded if and only if (11) holds for some real number c. The theorem now follows from Corollary 4.9-7, since convergence $f_n(x) \longrightarrow f(x)$ for all $x \in X$ is weak* convergence $f_n \xrightarrow{w^*} f$. ∎

It is trivial that in this theorem we may replace the polynomials by any other set which is dense in the real space $C[a, b]$.

Furthermore, in most integration methods the coefficients are all nonnegative. Taking $x = 1$, we then have by 4.11-2

$$f_n(1) = \sum_{k=0}^{n} \alpha_k^{(n)} = \sum_{k=0}^{n} |\alpha_k^{(n)}| = f(1) = \int_a^b dt = b - a,$$

so that (11) holds. This proves

4.11-4 Steklov's Theorem (Numerical integration). *A process of numerical integration* (3) *which satisfies* 4.11-2 *and has nonnegative coefficients* $\alpha_k^{(n)}$ *converges for every continuous function.*

In the proof of 4.11-3 we used the

4.11-5 Weierstrass Approximation Theorem (Polynomials). *The set W of all polynomials with real coefficients is dense in the real space $C[a, b]$.*

Hence for every $x \in C[a, b]$ and given $\varepsilon > 0$ there exists a polynomial p such that $|x(t) - p(t)| < \varepsilon$ for all $t \in [a, b]$.

Proof. Every $x \in C[a, b]$ is uniformly continuous on $J = [a, b]$ since J is compact. Hence for any $\varepsilon > 0$ there is a y whose graph is an arc of a polygon such that

$$\max_{t \in J} |x(t) - y(t)| < \frac{\varepsilon}{3}. \tag{12}$$

We first assume that $x(a) = x(b)$ and $y(a) = y(b)$. Since y is piecewise linear and continuous, its Fourier coefficients have bounds of the form

$|a_0| < k$, $|a_m| < k/m^2$, $|b_m| < k/m^2$. This can be seen by applying integration by parts to the formulas for a_m and b_m (cf. 3.5-1 where we have $[a, b] = [0, 2\pi]$). (Cf. also Prob. 10 at the end of this section.) Hence for the Fourier series of y (representing the periodic extension of y, of period $b-a$), we have, writing $\kappa = 2\pi/(b-a)$ for simplicity,

$$(13) \quad \left| a_0 + \sum_{m=1}^{\infty} (a_m \cos \kappa m t + b_m \sin \kappa m t) \right| \leqq 2k\left(1 + \sum_{m=1}^{\infty} \frac{1}{m^2}\right) = 2k\left(1 + \frac{1}{6}\pi^2\right).$$

This shows that the series converges uniformly on J. Consequently, for the nth partial sum s_n with sufficiently large n,

$$(14) \quad \max_{t \in J} |y(t) - s_n(t)| < \frac{\varepsilon}{3}.$$

The Taylor series of the cosine and sine functions in s_n also converge uniformly on J, so that there is a polynomial p (obtained, for instance, from suitable partial sums of those series) such that

$$\max_{t \in J} |s_n(t) - p(t)| < \frac{\varepsilon}{3}.$$

From this, (12), (14) and

$$|x(t) - p(t)| \leqq |x(t) - y(t)| + |y(t) - s_n(t)| + |s_n(t) - p(t)|$$

we have

$$(15) \quad \max_{t \in J} |x(t) - p(t)| < \varepsilon.$$

This takes care of every $x \in C[a, b]$ such that $x(a) = x(b)$. If $x(a) \neq x(b)$, take $u(t) = x(t) - \gamma(t-a)$ with γ such that $u(a) = u(b)$. Then for u there is a polynomial q satisfying $|u(t) - q(t)| < \varepsilon$ on J. Hence $p(t) = q(t) + \gamma(t-a)$ satisfies (15) because $x - p = u - q$. Since $\varepsilon > 0$ was arbitrary, we have shown that W is dense in $C[a, b]$. ∎

The first proof of the theorem was given by K. Weierstrass (1885), and there are many other proofs, for instance, one by S. N. Bernstein (1912), which yields a uniformly convergent sequence of polynomials ("Bernstein polynomials") explicitly in terms of x. Bernstein's proof can be found in K. Yosida (1971), pp. 8–9.

Problems

1. The **rectangular rule** is (Fig. 45)

$$\int_a^b x(t)\,dt \approx h[x(t_1^*)+\cdots+x(t_n^*)], \qquad h=\frac{b-a}{n}$$

where $t_k^* = a+(k-\frac{1}{2})h$. How is this formula obtained? What are the nodes and the coefficients? How can we obtain error bounds for the approximate value given by the formula?

2. The **trapezoidal rule** is (Fig. 46)

$$\int_{t_0}^{t_1} x(t)\,dt \approx \frac{h}{2}(x_0+x_1), \qquad h=\frac{b-a}{n}$$

or

$$\int_a^b x(t)\,dt \approx h(\tfrac{1}{2}x_0+x_1+\cdots+x_{n-1}+\tfrac{1}{2}x_n)$$

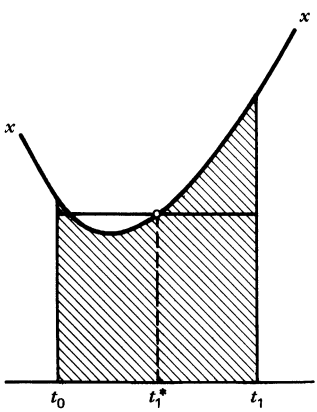


Fig. 45. Rectangular rule

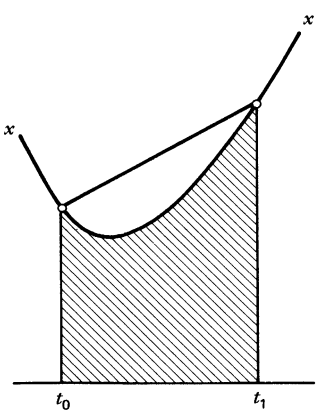


Fig. 46. Trapezoidal rule

where $x_k = x(t_k)$ and $t_k = a + kh$. Explain how the formulas are obtained if we approximate x by a piecewise linear function.

3. **Simpson's rule** is (Fig. 47)

$$\int_{t_0}^{t_2} x(t)\,dt \approx \frac{h}{3}(x_0 + 4x_1 + x_2) \qquad h = \frac{b-a}{n}$$

or

$$\int_a^b x(t)\,dt \approx \frac{h}{3}(x_0 + 4x_1 + 2x_2 + \cdots + 4x_{n-1} + x_n)$$

where n is even, $x_k = x(t_k)$ and $t_k = a + kh$. Show that these formulas are obtained if we approximate x on $[t_0, t_2]$ by a quadratic polynomial with values at t_0, t_1, t_2 equal to those of x; similarly on $[t_2, t_4]$, etc.

4. Let $f(x) = f_n(x) - \varepsilon_n(x)$ where f_n is the approximation obtained by the trapezoidal rule. Show that for any twice continuously differentiable function x we have the error bounds

$$k_n m_2^* \leqq \varepsilon_n(x) \leqq k_n m_2 \qquad \text{where} \qquad k_n = \frac{(b-a)^3}{12n^2}$$

and m_2 and m_2^* are the maximum and minimum of x'' on $[a, b]$.

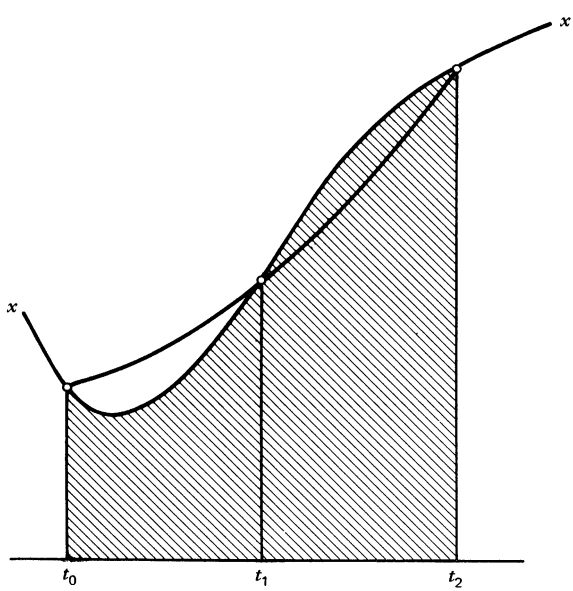


Fig. 47. Simpson's rule

5. Simpson's rule is widely used in practice. To get a feeling for the increase in accuracy, apply both the trapezoidal rule and Simpson's rule with $n = 10$ to the integral

$$I = \int_0^1 e^{-t^2}\, dt$$

and compare the values

0.746211 and 0.746825

with the actual value 0.746824 (exact to 6D).

t	e^{-t^2}
0	1.000 000
0.1	0.990 050
0.2	0.960 789
0.3	0.913 931
0.4	0.852 144
0.5	0.778 801
0.6	0.697 676
0.7	0.612 626
0.8	0.527 292
0.9	0.444 858
1.0	0.367 879

6. Using Prob. 4, show that bounds for the error of 0.746 211 in Prob. 5 are $-0.001\,667$ and $0.000\,614$, so that

$$0.745\,597 \leqq I \leqq 0.747\,878.$$

7. The **three-eights rule** is

$$\int_{t_0}^{t_3} x(t)\, dt \approx \frac{3h}{8}(x_0 + 3x_1 + 3x_2 + x_3)$$

where $x_k = x(t_k)$ and $t_k = a + kh$. Show that this formula is obtained if we approximate x on $[t_0, t_3]$ by a cubical polynomial which equals x at the nodes t_0, t_1, t_2, t_3. (The rules in Probs. 2, 3, 7 are the first members of the sequence of *Newton-Cotes formulas*.)

8. Consider the integration formula

$$\int_{-h}^{h} x(t)\, dt = 2hx(0) + r(x)$$

where r is the error. Assume that $x \in C^1[-h, h]$, that is, x is continuously differentiable on $J = [-h, h]$. Show that then the error can be estimated

$$|r(x)| \leqq h^2 p(x)$$

where

$$p(x) = \max_{t \in J} |x'(t)|.$$

Show that p is a seminorm on the vector space of those functions. (Cf. Prob. 12, Sec. 2.3.)

9. If x is real analytic, show that

$$\int_{-h}^{h} x(t)\,dt = 2h\left(x(0) + x''(0)\frac{h^2}{3!} + x^{IV}(0)\frac{h^4}{5!} + \cdots\right). \tag{16}$$

Assume for the integral an approximate expression of the form $2h(\alpha_{-1}x(-h) + \alpha_0 x(0) + \alpha_1 x(h))$ and determine α_{-1}, α_0, α_1 so that as many powers h, h^2, $\cdots$ as possible agree with (16). Show that this gives Simpson's rule

$$\int_{-h}^{h} x(t)\,dt \approx \frac{h}{3}(x(-h) + 4x(0) + x(h)).$$

Why does this derivation show that the rule is exact for cubical polynomials?

10. In the proof of the Weierstrass approximation theorem we used bounds for the Fourier coefficients of a continuous and piecewise linear function. How can those bounds be obtained?

4.12 Open Mapping Theorem

We have discussed the Hahn-Banach theorem and the uniform boundedness theorem and shall now approach the third "big" theorem in this chapter, the open mapping theorem. It will be concerned with open mappings. These are mappings such that the image of every open set is an open set (definition below). Remembering our discussion of the importance of open sets (cf. Sec. 1.3), we understand that open mappings are of general interest. More specifically, the open mapping theorem states conditions under which a bounded linear operator is an open mapping. As in the uniform boundedness theorem we again need completeness, and the present theorem exhibits another reason why Banach spaces are more satisfactory than incomplete normed spaces. The theorem also gives conditions under which the inverse of a bounded linear operator is bounded. The proof of the open mapping

theorem will be based on Baire's category theorem stated and explained in Sec. 4.7.

Let us begin by introducing the concept of an open mapping.

4.12-1 Definition (Open mapping). Let X and Y be metric spaces. Then T: $\mathcal{D}(T) \longrightarrow Y$ with domain $\mathcal{D}(T) \subset X$ is called an *open mapping* if for every open set in $\mathcal{D}(T)$ the image is an open set in Y. ∎

Note that if a mapping is not surjective, one must take care to distinguish between the assertions that the mapping is open as a mapping from its domain

(a) into Y,

(b) onto its range.

(b) is weaker than (a). For instance, if $X \subset Y$, the mapping $x \longmapsto x$ of X into Y is open if and only if X is an open subset of Y, whereas the mapping $x \longmapsto x$ of X onto its range (which is X) is open in any case.

Furthermore, to avoid confusion, we should remember that, by Theorem 1.3-4, a continuous mapping T: $X \longrightarrow Y$ has the property that for every open set in Y the inverse image is an open set in X. This does *not* imply that T maps open sets in X onto open sets in Y. For example, the mapping $\mathbf{R} \longrightarrow \mathbf{R}$ given by $t \longmapsto \sin t$ is continuous but maps $(0, 2\pi)$ onto $[-1, 1]$.

4.12-2 Open Mapping Theorem, Bounded Inverse Theorem. *A bounded linear operator T from a Banach space X onto a Banach space Y is an open mapping. Hence if T is bijective, T^{-1} is continuous and thus bounded.*

The proof will readily follow from

4.12-3 Lemma (Open unit ball). *A bounded linear operator T from a Banach space X onto a Banach space Y has the property that the image $T(B_0)$ of the open unit ball $B_0 = B(0; 1) \subset X$ contains an open ball about $0 \in Y$.*

Proof. Proceeding stepwise, we prove:

(a) The closure of the image of the open ball $B_1 = B(0; \frac{1}{2})$ contains an open ball B^*.

(b) $\overline{T(B_n)}$ contains an open ball V_n about $0 \in Y$, where $B_n = B(0; 2^{-n}) \subset X$.

(c) $T(B_0)$ contains an open ball about $0 \in Y$.

The details are as follows.

(a) In connection with subsets $A \subset X$ we shall write αA (α a scalar) and $A + w$ ($w \in X$) to mean

(1) $$\alpha A = \{x \in X \mid x = \alpha a,\ a \in A\}$$ (Fig. 48)

(2) $$A + w = \{x \in X \mid x = a + w,\ a \in A\}$$ (Fig. 49)

and similarly for subsets of Y.

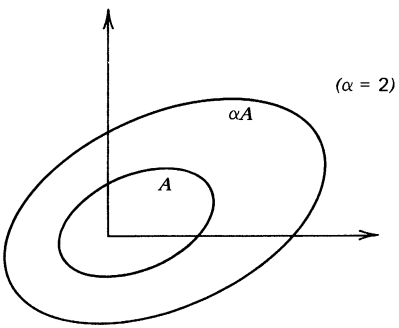


Fig. 48. Illustration of formula (1)

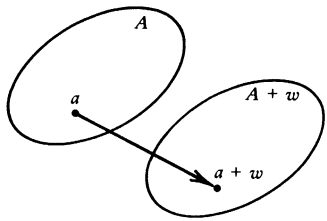


Fig. 49. Illustration of formula (2)

We consider the open ball $B_1 = B(0; \frac{1}{2}) \subset X$. Any fixed $x \in X$ is in kB_1 with real k sufficiently large ($k > 2\|x\|$). Hence

$$X = \bigcup_{k=1}^{\infty} kB_1.$$

Since T is surjective and linear,

(3) $$Y = T(X) = T\left(\bigcup_{k=1}^{\infty} kB_1\right) = \bigcup_{k=1}^{\infty} kT(B_1) = \bigcup_{k=1}^{\infty} \overline{kT(B_1)}.$$

Note that by taking closures we did not add further points to the union since that union was already the whole space Y. Since Y is complete, it is nonmeager in itself, by Baire's category theorem 4.7-2. Hence, noting that (3) is similar to (1) in 4.7-2, we conclude that a $\overline{kT(B_1)}$ must contain some open ball. This implies that $\overline{T(B_1)}$ also contains an open ball, say, $B^* = B(y_0; \varepsilon) \subset \overline{T(B_1)}$. It follows that

(4) $$B^* - y_0 = B(0; \varepsilon) \subset \overline{T(B_1)} - y_0.$$

(b) We prove that $B^* - y_0 \subset \overline{T(B_0)}$, where B_0 is given in the theorem. This we do by showing that [cf. (4)]

$$\overline{T(B_1)} - y_0 \subset \overline{T(B_0)}. \tag{5}$$

Let $y \in \overline{T(B_1)} - y_0$. Then $y + y_0 \in \overline{T(B_1)}$, and we remember that $y_0 \in \overline{T(B_1)}$, too. By 1.4-6($a$) there are

$$u_n = Tw_n \in T(B_1) \qquad \text{such that} \qquad u_n \longrightarrow y + y_0,$$

$$v_n = Tz_n \in T(B_1) \qquad \text{such that} \qquad v_n \longrightarrow y_0.$$

Since w_n, $z_n \in B_1$ and B_1 has radius $\frac{1}{2}$, it follows that

$$\|w_n - z_n\| \leqq \|w_n\| + \|z_n\| < \tfrac{1}{2} + \tfrac{1}{2} = 1,$$

so that $w_n - z_n \in B_0$. From

$$T(w_n - z_n) = Tw_n - Tz_n = u_n - v_n \longrightarrow y$$

we see that $y \in \overline{T(B_0)}$. Since $y \in \overline{T(B_1)} - y_0$ was arbitrary, this proves (5). From (4) we thus have

$$B^* - y_0 = B(0; \varepsilon) \subset \overline{T(B_0)}. \tag{6}$$

Let $B_n = B(0; 2^{-n}) \subset X$. Since T is linear, $\overline{T(B_n)} = 2^{-n}\overline{T(B_0)}$. From (6) we thus obtain

$$V_n = B(0; \varepsilon/2^n) \subset \overline{T(B_n)}. \tag{7}$$

(c) We finally prove that

$$V_1 = B(0; \tfrac{1}{2}\varepsilon) \subset T(B_0)$$

by showing that every $y \in V_1$ is in $T(B_0)$. So let $y \in V_1$. From (7) with $n = 1$ we have $V_1 \subset \overline{T(B_1)}$. Hence $y \in \overline{T(B_1)}$. By 1.4-6(a) there must be a $v \in T(B_1)$ close to y, say, $\|y - v\| < \varepsilon/4$. Now $v \in T(B_1)$ implies $v = Tx_1$ for some $x_1 \in B_1$. Hence

$$\|y - Tx_1\| < \frac{\varepsilon}{4}.$$

From this and (7) with $n=2$ we see that $y-Tx_1 \in V_2 \subset \overline{T(B_2)}$. As before we conclude that there is an $x_2 \in B_2$ such that

$$\|(y-Tx_1)-Tx_2\| < \frac{\varepsilon}{8}.$$

Hence $y-Tx_1-Tx_2 \in V_3 \subset \overline{T(B_3)}$, and so on. In the nth step we can choose an $x_n \in B_n$ such that

$$\left\| y - \sum_{k=1}^{n} Tx_k \right\| < \frac{\varepsilon}{2^{n+1}} \qquad (n=1,2,\cdots). \tag{8}$$

Let $z_n = x_1 + \cdots + x_n$. Since $x_k \in B_k$, we have $\|x_k\| < 1/2^k$. This yields for $n > m$

$$\|z_n - z_m\| \leqq \sum_{k=m+1}^{n} \|x_k\| < \sum_{k=m+1}^{\infty} \frac{1}{2^k} \longrightarrow 0$$

as $m \longrightarrow \infty$. Hence (z_n) is Cauchy. (z_n) converges, say, $z_n \longrightarrow x$ because X is complete. Also $x \in B_0$ since B_0 has radius 1 and

$$\sum_{k=1}^{\infty} \|x_k\| < \sum_{k=1}^{\infty} \frac{1}{2^k} = 1. \tag{9}$$

Since T is continuous, $Tz_n \longrightarrow Tx$, and (8) shows that $Tx = y$. Hence $y \in T(B_0)$. ∎

***Proof of Theorem* 4.12-2.** We prove that for every open set $A \subset X$ the image $T(A)$ is open in Y. This we do by showing that for every $y = Tx \in T(A)$ the set $T(A)$ contains an open ball about $y = Tx$.

Let $y = Tx \in T(A)$. Since A is open, it contains an open ball with center x. Hence $A - x$ contains an open ball with center 0; let the radius of the ball be r and set $k = 1/r$, so that $r = 1/k$. Then $k(A-x)$ contains the open unit ball $B(0;1)$. Lemma 4.12-3 now implies that $T(k(A-x)) = k[T(A) - Tx]$ contains an open ball about 0, and so does $T(A) - Tx$. Hence $T(A)$ contains an open ball about $Tx = y$. Since $y \in T(A)$ was arbitrary, $T(A)$ is open.

Finally, if $T^{-1}: Y \longrightarrow X$ exists, it is continuous by Theorem 1.3-4 because T is open. Since T^{-1} is linear by Theorem 2.6-10, it is bounded by Theorem 2.7-9. ∎

Problems

1. Show that T: $\mathbf{R}^2 \longrightarrow \mathbf{R}$ defined by $(\xi_1, \xi_2) \longmapsto (\xi_1)$ is open. Is the mapping $\mathbf{R}^2 \longmapsto \mathbf{R}^2$ given by $(\xi_1, \xi_2) \longmapsto (\xi_1, 0)$ an open mapping?

2. Show that an open mapping need not map closed sets onto closed sets.

3. Extending (1) and (2), we can define

$$A + B = \{x \in X \mid x = a + b,\ a \in A,\ b \in B\},$$

where A, $B \subset X$. To become familiar with this notation find αA, $A + w$, $A + A$, where $A = \{1, 2, 3, 4\}$. Explain Fig. 50.

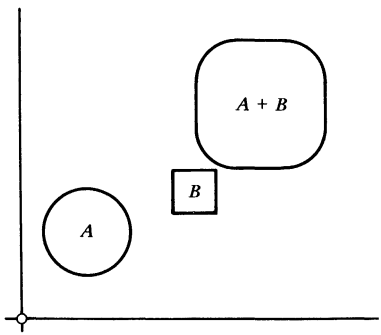


Fig. 50. Sets A, B and $A + B$ in the plane

4. Show that in (9) the inequality is strict.

5. Let X be the normed space whose points are sequences of complex numbers $x = (\xi_j)$ with only finitely many nonzero terms and norm defined by $\|x\| = \sup_j |\xi_j|$. Let T: $X \longrightarrow X$ be defined by

$$y = Tx = \left(\xi_1, \frac{1}{2}\xi_2, \frac{1}{3}\xi_3, \cdots\right).$$

Show that T is linear and bounded but T^{-1} is unbounded. Does this contradict 4.12-2?

6. Let X and Y be Banach spaces and T: $X \longrightarrow Y$ an injective bounded linear operator. Show that T^{-1}: $\mathcal{R}(T) \longrightarrow X$ is bounded if and only if $\mathcal{R}(T)$ is closed in Y.

7. Let T: $X \longrightarrow Y$ be a bounded linear operator, where X and Y are Banach spaces. If T is bijective, show that there are positive real numbers a and b such that $a\|x\| \leqq \|Tx\| \leqq b\|x\|$ for all $x \in X$.

8. **(Equivalent norms)** Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be norms on a vector space X such that $X_1 = (X, \|\cdot\|_1)$ and $X_2 = (X, \|\cdot\|_2)$ are complete. If $\|x_n\|_1 \longrightarrow 0$ always implies $\|x_n\|_2 \longrightarrow 0$, show that convergence in X_1 implies convergence in X_2 and conversely, and there are positive numbers a and b such that for all $x \in X$,

$$a\|x\|_1 \leqq \|x\|_2 \leqq b\|x\|_1.$$

(Note that then these norms are equivalent; cf. Def. 2.4-4.)

9. Let $X_1 = (X, \|\cdot\|_1)$ and $X_2 = (X, \|\cdot\|_2)$ be Banach spaces. If there is a constant c such that $\|x\|_1 \leqq c\|x\|_2$ for all $x \in X$, show that there is a constant k such that $\|x\|_2 \leqq k\|x\|_1$ for all $x \in X$ (so that the two norms are equivalent; cf. Def. 2.4-4).

10. From Sec. 1.3 we know that the set $\mathcal{T}$ of all open subsets of a metric space X is called a *topology* for X. Consequently, each norm on a vector space X defines a topology for X. If we have two norms on X such that $X_1 = (X, \|\cdot\|_1)$ and $X_2 = (X, \|\cdot\|_2)$ are Banach spaces and the topologies $\mathcal{T}_1$ and $\mathcal{T}_2$ defined by $\|\cdot\|_1$ and $\|\cdot\|_2$ satisfy $\mathcal{T}_1 \supset \mathcal{T}_2$, show that $\mathcal{T}_1 = \mathcal{T}_2$.

4.13 Closed Linear Operators. Closed Graph Theorem

Not all linear operators of practical importance are bounded. For instance, the differential operator in 2.7-5 is unbounded, and in quantum mechanics and other applications one needs unbounded operators quite frequently. However, practically all of the linear operators which the analyst is likely to use are so-called closed linear operators. This makes it worthwhile to give an introduction to these operators. In this section we define closed linear operators on normed spaces and consider some of their properties, in particular in connection with the important closed graph theorem which states sufficient conditions under which a closed linear operator on a Banach space is bounded.

A more detailed study of closed and other unbounded operators in Hilbert spaces will be presented in Chap. 10 and applications to quantum mechanics in Chap. 11.

Let us begin with the definition.

4.13-1 Definition (Closed linear operator). Let X and Y be normed spaces and T: $\mathcal{D}(T) \longrightarrow Y$ a linear operator with domain $\mathcal{D}(T) \subset X$. Then T is called a *closed linear operator* if its *graph*

$$\mathcal{G}(T) = \{(x, y) \mid x \in \mathcal{D}(T),\ y = Tx\}$$

is closed in the normed space $X \times Y$, where the two algebraic operations of a vector space in $X \times Y$ are defined as usual, that is

$$(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2)$$

$$\alpha(x, y) = (\alpha x, \alpha y)$$

(α a scalar) and the norm on $X \times Y$ is defined by[7]

$$\|(x, y)\| = \|x\| + \|y\|. \tag{1}$$

∎

Under what conditions will a closed linear operator be bounded? An answer is given by the important

4.13-2 Closed Graph Theorem. *Let X and Y be Banach spaces and T: $\mathcal{D}(T) \longrightarrow Y$ a closed linear operator, where $\mathcal{D}(T) \subset X$. Then if $\mathcal{D}(T)$ is closed in X, the operator T is bounded.*

Proof. We first show that $X \times Y$ with norm defined by (1) is complete. Let (z_n) be Cauchy in $X \times Y$, where $z_n = (x_n, y_n)$. Then for every $\varepsilon > 0$ there is an N such that

$$\|z_n - z_m\| = \|x_n - x_m\| + \|y_n - y_m\| < \varepsilon \qquad (m, n > N). \tag{2}$$

Hence (x_n) and (y_n) are Cauchy in X and Y, respectively, and converge, say, $x_n \longrightarrow x$ and $y_n \longrightarrow y$, because X and Y are complete. This implies that $z_n \longrightarrow z = (x, y)$ since from (2) with $m \longrightarrow \infty$ we have $\|z_n - z\| \leqq \varepsilon$ for $n > N$. Since the Cauchy sequence (z_n) was arbitrary, $X \times Y$ is complete.

[7] For other norms, see Prob. 2.

By assumption, $\mathcal{G}(T)$ is closed in $X \times Y$ and $\mathcal{D}(T)$ is closed in X. Hence $\mathcal{G}(T)$ and $\mathcal{D}(T)$ are complete by 1.4-7. We now consider the mapping

$$P: \mathcal{G}(T) \longrightarrow \mathcal{D}(T)$$

$$(x, Tx) \longmapsto x.$$

P is linear. P is bounded because

$$\|P(x, Tx)\| = \|x\| \leqq \|x\| + \|Tx\| = \|(x, Tx)\|.$$

P is bijective; in fact the inverse mapping is

$$P^{-1}: \mathcal{D}(T) \longrightarrow \mathcal{G}(T)$$

$$x \longmapsto (x, Tx).$$

Since $\mathcal{G}(T)$ and $\mathcal{D}(T)$ are complete, we can apply the bounded inverse theorem 4.12-2 and see that P^{-1} is bounded, say, $\|(x, Tx)\| \leqq b\|x\|$ for some b and all $x \in \mathcal{D}(T)$. Hence T is bounded because

$$\|Tx\| \leqq \|Tx\| + \|x\| = \|(x, Tx)\| \leqq b\|x\|$$

for all $x \in \mathcal{D}(T)$. ∎

By definition, $\mathcal{G}(T)$ is closed if and only if $z = (x, y) \in \overline{\mathcal{G}(T)}$ implies $z \in \mathcal{G}(T)$. From Theorem 1.4-6(a) we see that $z \in \overline{\mathcal{G}(T)}$ if and only if there are $z_n = (x_n, Tx_n) \in \mathcal{G}(T)$ such that $z_n \longrightarrow z$, hence

$$x_n \longrightarrow x, \qquad Tx_n \longrightarrow y; \tag{3}$$

and $z = (x, y) \in \mathcal{G}(T)$ if and only if $x \in \mathcal{D}(T)$ and $y = Tx$. This proves the following useful criterion which expresses a property that is often taken as a definition of closedness of a linear operator.

4.13-3 Theorem (Closed linear operator). *Let $T: \mathcal{D}(T) \longrightarrow Y$ be a linear operator, where $\mathcal{D}(T) \subset X$ and X and Y are normed spaces. Then T is closed if and only if it has the following property. If $x_n \longrightarrow x$, where $x_n \in \mathcal{D}(T)$, and $Tx_n \longrightarrow y$, then $x \in \mathcal{D}(T)$ and $Tx = y$.*

Note well that this property is different from the following property of a bounded linear operator. If a linear operator T is bounded

and thus continuous, and if (x_n) is a sequence in $\mathfrak{D}(T)$ which converges in $\mathfrak{D}(T)$, then (Tx_n) also converges; cf. 1.4-8. This need not hold for a closed linear operator. However, if T is closed and two sequences (x_n) and $(\tilde{x}_n)$ in the domain of T converge with the same limit and if the corresponding sequences (Tx_n) and $(T\tilde{x}_n)$ both converge, then the latter have the same limit (cf. Prob. 6).

4.13-4 Example (Differential operator). Let $X = C[0, 1]$ and

$$T\colon \mathfrak{D}(T) \longrightarrow X$$
$$x \longmapsto x'$$

where the prime denotes differentiation and $\mathfrak{D}(T)$ is the subspace of functions $x \in X$ which have a continuous derivative. Then T is not bounded, but is closed.

Proof. We see from 2.7-5 that T is not bounded. We prove that T is closed by applying Theorem 4.13-3. Let (x_n) in $\mathfrak{D}(T)$ be such that both (x_n) and (Tx_n) converge, say,

$$x_n \longrightarrow x \qquad \text{and} \qquad Tx_n = x_n' \longrightarrow y.$$

Since convergence in the norm of $C[0, 1]$ is uniform convergence on $[0, 1]$, from $x_n' \longrightarrow y$ we have

$$\int_0^t y(\tau)\, d\tau = \int_0^t \lim_{n\to\infty} x_n'(\tau)\, d\tau = \lim_{n\to\infty} \int_0^t x_n'(\tau)\, d\tau = x(t) - x(0),$$

that is,

$$x(t) = x(0) + \int_0^t y(\tau)\, d\tau.$$

This shows that $x \in \mathfrak{D}(T)$ and $x' = y$. Theorem 4.13-3 now implies that T is closed. ∎

It is worth noting that in this example, $\mathfrak{D}(T)$ is not closed in X since T would then be bounded by the closed graph theorem.

Closedness does not imply boundedness of a linear operator. Conversely, boundedness does not imply closedness.

Proof. The first statement is illustrated by 4.13-4 and the second one by the following example. Let T: $\mathcal{D}(T) \longrightarrow \mathcal{D}(T) \subset X$ be the identity operator on $\mathcal{D}(T)$, where $\mathcal{D}(T)$ is a proper dense subspace of a normed space X. Then it is trivial that T is linear and bounded. However, T is not closed. This follows immediately from Theorem 4.13-3 if we take an $x \in X - \mathcal{D}(T)$ and a sequence (x_n) in $\mathcal{D}(T)$ which converges to x. ∎

Our present discussion seems to indicate that in connection with unbounded operators the determination of domains and extension problems may play a basic role. This is in fact so, as we shall see in more detail in Chap. 10. The statement which we have just proved is rather negative in spirit. On the positive side we have

4.13-5 Lemma (Closed operator). *Let T: $\mathcal{D}(T) \longrightarrow Y$ be a bounded linear operator with domain $\mathcal{D}(T) \subset X$, where X and Y are normed spaces. Then:*

(a) *If $\mathcal{D}(T)$ is a closed subset of X, then T is closed.*

(b) *If T is closed and Y is complete, then $\mathcal{D}(T)$ is a closed subset of X.*

Proof. **(a)** If (x_n) is in $\mathcal{D}(T)$ and converges, say, $x_n \longrightarrow x$, and is such that (Tx_n) also converges, then $x \in \overline{\mathcal{D}(T)} = \mathcal{D}(T)$ since $\mathcal{D}(T)$ is closed, and $Tx_n \longrightarrow Tx$ since T is continuous. Hence T is closed by Theorem 4.13-3.

(b) For $x \in \overline{\mathcal{D}(T)}$ there is a sequence (x_n) in $\mathcal{D}(T)$ such that $x_n \longrightarrow x$; cf. 1.4-6. Since T is bounded,

$$\|Tx_n - Tx_m\| = \|T(x_n - x_m)\| \leqq \|T\| \, \|x_n - x_m\|.$$

This shows that (Tx_n) is Cauchy. (Tx_n) converges, say, $Tx_n \longrightarrow y \in Y$ because Y is complete. Since T is closed, $x \in \mathcal{D}(T)$ by 4.13-3 (and $Tx = y$). Hence $\mathcal{D}(T)$ is closed because $x \in \overline{\mathcal{D}(T)}$ was arbitrary. ∎

Problems

1. Prove that (1) defines a norm on $X \times Y$.

2. Other frequently used norms on the product $X \times Y$ of normed spaces

X and Y are defined by

$$\|(x, y)\| = \max \{\|x\|, \|y\|\}$$

and

$$\|(x, y)\|_0 = (\|x\|^2 + \|y\|^2)^{1/2}.$$

Verify that these are norms.

3. Show that the graph $\mathcal{G}(T)$ of a linear operator $T: X \longrightarrow Y$ is a vector subspace of $X \times Y$.

4. If X and Y in Def. 4.13-1 are Banach spaces, show that $V = X \times Y$ with norm defined by (1) is a Banach space.

5. (Inverse) If the inverse T^{-1} of a closed linear operator exists, show that T^{-1} is a closed linear operator.

6. Let T be a closed linear operator. If two sequences (x_n) and $(\tilde{x}_n)$ in $\mathcal{D}(T)$ converge with the same limit x and if (Tx_n) and $(T\tilde{x}_n)$ both converge, show that (Tx_n) and $(T\tilde{x}_n)$ have the same limit.

7. Obtain the second statement in Theorem 4.12-2 from the closed graph theorem.

8. Let X and Y be normed spaces and let $T: X \longrightarrow Y$ be a closed linear operator. (*a*) Show that the image A of a compact subset $C \subset X$ is closed in Y. (*b*) Show that the inverse image B of a compact subset $K \subset Y$ is closed in X. (Cf. Def. 2.5-1.)

9. If $T: X \longrightarrow Y$ is a closed linear operator, where X and Y are normed spaces and Y is compact, show that T is bounded.

10. Let X and Y be normed spaces and X compact. If $T: X \longrightarrow Y$ is a bijective closed linear operator, show that T^{-1} is bounded.

11. (Null space) Show that the null space $\mathcal{N}(T)$ of a closed linear operator $T: X \longrightarrow Y$ is a closed subspace of X.

12. Let X and Y be normed spaces. If $T_1: X \longrightarrow Y$ is a closed linear operator and $T_2 \in B(X, Y)$, show that $T_1 + T_2$ is a closed linear operator.

13. Let T be a closed linear operator with domain $\mathcal{D}(T)$ in a Banach space X and range $\mathcal{R}(T)$ in a normed space Y. If T^{-1} exists and is bounded, show that $\mathcal{R}(T)$ is closed.

14. Assume that the terms of the series $u_1+u_2+\cdots$ are continuously differentiable functions on the interval $J=[0, 1]$ and that the series is uniformly convergent on J and has the sum x. Furthermore, suppose that $u_1'+u_2'+\cdots$ also converges uniformly on J. Show that then x is continuously differentiable on $(0, 1)$ and $x'=u_1'+u_2'+\cdots$.

15. (Closed extension) Let T: $\mathcal{D}(T)\longrightarrow Y$ be a linear operator with graph $\mathcal{G}(T)$, where $\mathcal{D}(T)\subset X$ and X and Y are Banach spaces. Show that T has an extension $\tilde{T}$ which is a closed linear operator with graph $\overline{\mathcal{G}(T)}$ if and only if $\overline{\mathcal{G}(T)}$ does not contain an element of the form $(0, y)$, where $y\neq 0$.

CHAPTER 5

FURTHER APPLICATIONS: BANACH FIXED POINT THEOREM

This chapter is optional. Its material will not be used in the remaining chapters.

Prerequisite is Chap. 1 (but not Chaps. 2 to 4), so that the present chapter can also be studied immediately after Chap. 1 if so desired.

The Banach fixed point theorem is important as a source of existence and uniqueness theorems in different branches of analysis. In this way the theorem provides an impressive illustration of the unifying power of functional analytic methods and of the usefulness of fixed point theorems in analysis.

Brief orientation about main content

The *Banach fixed point theorem* or *contraction theorem* 5.1-2 concerns certain mappings (*contractions*, cf. 5.1-1) of a complete metric space into itself. It states conditions sufficient for the existence and uniqueness of a *fixed point* (point that is mapped onto itself). The theorem also gives an iterative process by which we can obtain approximations to the fixed point and error bounds (cf. 5.1-3). We consider three important fields of application of the theorem, namely,

linear algebraic equations (Sec. 5.2),
ordinary differential equations (Sec. 5.3),
integral equations (Sec. 5.4).

There are other applications (for instance, partial differential equations) whose discussion would require more prerequisites.

5.1 Banach Fixed Point Theorem

A **fixed point** of a mapping T: $X \longrightarrow X$ of a set X into itself is an $x \in X$ which is mapped onto itself (is "kept fixed" by T), that is,

$$Tx = x,$$

the image Tx coincides with x.

For example, a translation has no fixed points, a rotation of the plane has a single fixed point (the center of rotation), the mapping $x \longmapsto x^2$ of $\mathbf{R}$ into itself has two fixed points (0 and 1) and the projection $(\xi_1, \xi_2) \longmapsto \xi_1$ of $\mathbf{R}^2$ onto the ξ_1-axis has infinitely many fixed points (all points of the ξ_1-axis).

The Banach fixed point theorem to be stated below is an existence and uniqueness theorem for fixed points of certain mappings, and it also gives a constructive procedure for obtaining better and better approximations to the fixed point (the solution of the practical problem). This procedure is called an **iteration**. By definition, this is a method such that we choose an arbitrary x_0 in a given set and calculate recursively a sequence $x_0, x_1, x_2, \cdots$ from a relation of the form

$$x_{n+1} = Tx_n \qquad n = 0, 1, 2, \cdots;$$

that is, we choose an arbitrary x_0 and determine successively $x_1 = Tx_0$, $x_2 = Tx_1, \cdots$.

Iteration procedures are used in nearly every branch of applied mathematics, and convergence proofs and error estimates are very often obtained by an application of Banach's fixed point theorem (or more difficult fixed point theorems). Banach's theorem gives sufficient conditions for the existence (and uniqueness) of a fixed point for a class of mappings, called contractions. The definition is as follows.

5.1-1 Definition (Contraction). Let $X = (X, d)$ be a metric space. A mapping $T: X \longrightarrow X$ is called a *contraction on* X if there is a positive real number $\alpha < 1$ such that for all $x, y \in X$

(1) $$d(Tx, Ty) \leqq \alpha d(x, y) \qquad (\alpha < 1).$$

Geometrically this means that any points x and y have images that are closer together than those points x and y; more precisely, the ratio $d(Tx, Ty)/d(x, y)$ does not exceed a constant α which is strictly less than 1.

5.1-2 Banach Fixed Point Theorem (Contraction Theorem). *Consider a metric space $X = (X, d)$, where $X \neq \emptyset$. Suppose that X is complete and let $T: X \longrightarrow X$ be a contraction on X. Then T has precisely one fixed point.*

Proof. We construct a sequence (x_n) and show that it is Cauchy, so that it converges in the complete space X, and then we prove that its

limit x is a fixed point of T and T has no further fixed points. This is the idea of the proof.

We choose any $x_0 \in X$ and define the "iterative sequence" (x_n) by

(2) $$x_0, \qquad x_1 = Tx_0, \qquad x_2 = Tx_1 = T^2x_0, \quad \cdots, \quad x_n = T^nx_0, \quad \cdots.$$

Clearly, this is the sequence of the images of x_0 under repeated application of T. We show that (x_n) is Cauchy. By (1) and (2),

$$\begin{aligned} d(x_{m+1}, x_m) &= d(Tx_m, Tx_{m-1}) \\ &\leqq \alpha d(x_m, x_{m-1}) \\ &= \alpha d(Tx_{m-1}, Tx_{m-2}) \\ &\leqq \alpha^2 d(x_{m-1}, x_{m-2}) \\ &\cdots \leqq \alpha^m d(x_1, x_0). \end{aligned} \tag{3}$$

Hence by the triangle inequality and the formula for the sum of a geometric progression we obtain for $n > m$

$$\begin{aligned} d(x_m, x_n) &\leqq d(x_m, x_{m+1}) + d(x_{m+1}, x_{m+2}) + \cdots + d(x_{n-1}, x_n) \\ &\leqq (\alpha^m + \alpha^{m+1} + \cdots + \alpha^{n-1}) d(x_0, x_1) \\ &= \alpha^m \frac{1-\alpha^{n-m}}{1-\alpha} d(x_0, x_1). \end{aligned}$$

Since $0 < \alpha < 1$, in the numerator we have $1 - \alpha^{n-m} < 1$. Consequently,

$$d(x_m, x_n) \leqq \frac{\alpha^m}{1-\alpha} d(x_0, x_1) \qquad (n > m). \tag{4}$$

On the right, $0 < \alpha < 1$ and $d(x_0, x_1)$ is fixed, so that we can make the right-hand side as small as we please by taking m sufficiently large (and $n > m$). This proves that (x_m) is Cauchy. Since X is complete, (x_m) converges, say, $x_m \longrightarrow x$. We show that this limit x is a fixed point of the mapping T.

From the triangle inequality and (1) we have

$$\begin{aligned} d(x, Tx) &\leqq d(x, x_m) + d(x_m, Tx) \\ &\leqq d(x, x_m) + \alpha d(x_{m-1}, x) \end{aligned}$$

and can make the sum in the second line smaller than any preassigned $\varepsilon > 0$ because $x_m \longrightarrow x$. We conclude that $d(x, Tx) = 0$, so that $x = Tx$ by (M2), Sec. 1.1. This shows that x is a fixed point of T.

x is the only fixed point of T because from $Tx = x$ and $T\tilde{x} = \tilde{x}$ we obtain by (1)

$$d(x, \tilde{x}) = d(Tx, T\tilde{x}) \leqq \alpha d(x, \tilde{x})$$

which implies $d(x, \tilde{x}) = 0$ since $\alpha < 1$. Hence $x = \tilde{x}$ by (M2) and the theorem is proved. ∎

5.1-3 Corollary (Iteration, error bounds). *Under the conditions of Theorem* 5.1-2 *the iterative sequence* (2) *with arbitrary* $x_0 \in X$ *converges to the unique fixed point* x *of* T. *Error estimates are the* **prior estimate**

$$\textbf{(5)} \qquad d(x_m, x) \leqq \frac{\alpha^m}{1-\alpha} d(x_0, x_1)$$

and the **posterior estimate**

$$\textbf{(6)} \qquad d(x_m, x) \leqq \frac{\alpha}{1-\alpha} d(x_{m-1}, x_m).$$

Proof. The first statement is obvious from the previous proof. Inequality (5) follows from (4) by letting $n \longrightarrow \infty$. We derive (6). Taking $m = 1$ and writing y_0 for x_0 and y_1 for x_1, we have from (5)

$$d(y_1, x) \leqq \frac{\alpha}{1-\alpha} d(y_0, y_1).$$

Setting $y_0 = x_{m-1}$, we have $y_1 = Ty_0 = x_m$ and obtain (6). ∎

The prior error bound (5) can be used at the beginning of a calculation for estimating the number of steps necessary to obtain a given accuracy. (6) can be used at intermediate stages or at the end of a calculation. It is at least as accurate as (5) and may be better; cf. Prob. 8.

From the viewpoint of applied mathematics the situation is not yet completely satisfactory because it frequently happens that a mapping T is a contraction not on the entire space X but merely on a subset Y of X. However, if Y is closed, it is complete by Theorem 1.4-7, so that T

has a fixed point x in Y, and $x_m \longrightarrow x$ as before, provided we impose a suitable restriction on the choice of x_0, so that the x_m's remain in Y. A typical and practically useful result of this kind is as follows.

5.1-4 Theorem (Contraction on a ball). *Let T be a mapping of a complete metric space $X=(X, d)$ into itself. Suppose T is a contraction on a closed ball $Y=\{x \mid d(x, x_0) \leqq r\}$, that is, T satisfies (1) for all $x, y \in Y$. Moreover, assume that*

$$d(x_0, Tx_0) < (1-\alpha)r. \tag{7}$$

Then the iterative sequence (2) converges to an $x \in Y$. This x is a fixed point of T and is the only fixed point of T in Y.

Proof. We merely have to show that all x_m's as well as x lie in Y. We put $m=0$ in (4), change n to m and use (7) to get

$$d(x_0, x_m) \leqq \frac{1}{1-\alpha}\, d(x_0, x_1) < r.$$

Hence all x_m's are in Y. Also $x \in Y$ since $x_m \longrightarrow x$ and Y is closed. The assertion of the theorem now follows from the proof of Banach's theorem 5.1-2. ∎

For later use the reader may give the simple proof of

5.1-5 Lemma (Continuity). *A contraction T on a metric space X is a continuous mapping.*

Problems

1. Give further examples of mappings in elementary geometry which have (*a*) a single fixed point, (*b*) infinitely many fixed points.

2. Let $X=\{x \in \mathbf{R} \mid x \geqq 1\} \subset \mathbf{R}$ and let the mapping T: $X \longrightarrow X$ be defined by $Tx = x/2 + x^{-1}$. Show that T is a contraction and find the smallest α.

3. Illustrate with an example that in Theorem 5.1-2, completeness is essential and cannot be omitted.

4. It is important that in Banach's theorem 5.1-2 the condition (1) cannot be replaced by $d(Tx, Ty) < d(x, y)$ when $x \neq y$. To see this, consider

$X = \{x \mid 1 \leqq x < +\infty\}$, taken with the usual metric of the real line, and T: $X \longrightarrow X$ defined by $x \longmapsto x + x^{-1}$. Show that $|Tx - Ty| < |x - y|$ when $x \neq y$, but the mapping has no fixed points.

5. If T: $X \longrightarrow X$ satisfies $d(Tx, Ty) < d(x, y)$ when $x \neq y$ and T has a fixed point, show that the fixed point is unique; here, (X, d) is a metric space.

6. If T is a contraction, show that T^n $(n \in \mathbf{N})$ is a contraction. If T^n is a contraction for an $n > 1$, show that T need not be a contraction.

7. Prove Lemma 5.1-5.

8. Show that the error bounds given by (5) form a proper monotone decreasing sequence. Show that (6) is at least as good as (5).

9. Show that in the case of Theorem 5.1-4 we have the prior error estimate $d(x_m, x) < \alpha^m r$ and the posterior estimate (6).

10. In analysis, a usual sufficient condition for the convergence of an iteration $x_n = g(x_{n-1})$ is that g be continuously differentiable and

$$|g'(x)| \leqq \alpha < 1.$$

Verify this by the use of Banach's fixed point theorem.

11. To find approximate numerical solutions of a given equation $f(x) = 0$, we may convert the equation to the form $x = g(x)$, choose an initial value x_0 and compute

$$x_n = g(x_{n-1}) \qquad n = 1, 2, \cdots.$$

Suppose that g is continuously differentiable on some interval $J = [x_0 - r, x_0 + r]$ and satisfies $|g'(x)| \leqq \alpha < 1$ on J as well as

$$|g(x_0) - x_0| < (1 - \alpha) r.$$

Show that then $x = g(x)$ has a unique solution x on J, the iterative sequence (x_m) converges to that solution, and one has the error estimates

$$|x - x_m| < \alpha^m r, \qquad |x - x_m| \leqq \frac{\alpha}{1 - \alpha} |x_m - x_{m-1}|.$$

12. Using Banach's theorem 5.1-2, set up an iteration process for solving $f(x)=0$ if f is continuously differentiable on an interval $J=[a, b]$, $f(a)<0, f(b)>0$ and $0<k_1 \leqq f'(x) \leqq k_2$ $(x \in J)$; use $g(x)=x-\lambda f(x)$ with a suitable λ.

13. Consider an iteration process for solving $f(x)=x^3+x-1=0$; proceed as follows. (*a*) Show that one possibility is

$$x_n = g(x_{n-1}) = (1+x_{n-1}^2)^{-1}.$$

Choose $x_0=1$ and perform three steps. Is $|g'(x)|<1$? (Cf. Prob. 10.) Show that the iteration can be illustrated by Fig. 51. (*b*) Estimate the errors by (5). (*c*) We can write $f(x)=0$ in the form $x=1-x^3$. Is this form suitable for iteration? Try $x_0=1$, $x_0=0.5$, $x_0=2$ and see what happens.

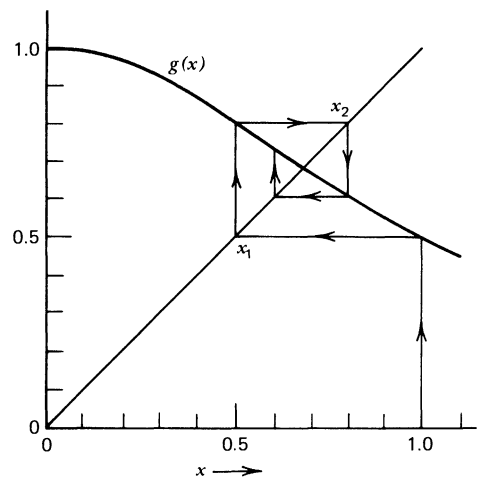


Fig. 51. Iteration in Prob. 13(*a*)

14. Show that another iteration process for the equation in Prob. 13 is

$$x_n = x_{n-1}^{1/2}(1+x_{n-1}^2)^{-1/2}.$$

Choose $x_0=1$. Determine x_1, x_2, x_3. What is the reason for the rapid convergence? (The real root is 0.682 328, 6D.)

15. (Newton's method) Let f be real-valued and twice continuously differentiable on an interval $[a, b]$, and let $\hat{x}$ be a simple zero of f in

(a, b). Show that *Newton's method* defined by

$$x_{n+1} = g(x_n), \qquad g(x_n) = x_n - \frac{f(x_n)}{f'(x_n)}$$

is a contraction in some neighborhood of $\hat{x}$ (so that the iterative sequence converges to $\hat{x}$ for any x_0 sufficiently close to $\hat{x}$).

16. (Square root) Show that an iteration for calculating the square root of a given positive number c is

$$x_{n+1} = g(x_n) = \frac{1}{2}\left(x_n + \frac{c}{x_n}\right)$$

where $n = 0, 1, \cdots$. What condition do we get from Prob. 10? Starting from $x_0 = 1$, calculate approximations $x_1, \cdots, x_4$ for $\sqrt{2}$.

17. Let T: $X \longrightarrow X$ be a contraction on a complete metric space, so that (1) holds. Because of rounding errors and for other reasons, instead of T one often has to take a mapping S: $X \longrightarrow X$ such that for all $x \in X$,

$$d(Tx, Sx) \leqq \eta \qquad (\eta > 0, \text{suitable}).$$

Using induction, show that then for any $x \in X$,

$$d(T^m x, S^m x) \leqq \eta \frac{1-\alpha^m}{1-\alpha} \qquad (m = 1, 2, \cdots).$$

18. The mapping S in Prob. 17 may not have a fixed point; but in practice, S^n often has a fixed point y for some n. Using Prob. 17, show that then for the distance from y to the fixed point x of T we have

$$d(x, y) \leqq \frac{\eta}{1-\alpha}.$$

19. In Prob. 17, let $x = Tx$ and $y_m = S^m y_0$. Using (5) and Prob. 17, show that

$$d(x, y_m) \leqq \frac{1}{1-\alpha}[\eta + \alpha^m d(y_0, Sy_0)].$$

What is the significance of this formula in applications?

20. (Lipschitz condition) A mapping T: $[a, b] \longrightarrow [a, b]$ is said to satisfy a *Lipschitz condition* with a *Lipschitz constant* k on $[a, b]$ if there is a constant k such that for all $x, y \in [a, b]$,

$$|Tx - Ty| \leqq k\,|x - y|.$$

(a) Is T a contraction? (b) If T is continuously differentiable, show that T satisfies a Lipschitz condition. (c) Does the converse of (b) hold?

5.2 Application of Banach's Theorem to Linear Equations

Banach's fixed point theorem has important applications to iteration methods for solving systems of linear algebraic equations and yields sufficient conditions for convergence and error bounds.

To understand the situation, we first remember that for solving such a system there are various *direct methods* (methods that would yield the exact solution after finitely many arithmetical operations if the precision—the word length of our computer—were unlimited); a familiar example is Gauss' elimination method (roughly, a systematic version of the elimination taught in school). However, an iteration, or *indirect method*, may be more efficient if the system is special, for instance, if it is *sparse*, that is, if it consists of many equations but has only a small number of nonzero coefficients. (Vibrational problems, networks and difference approximations of partial differential equations often lead to sparse systems.) Moreover, the usual direct methods require about $n^3/3$ arithmetical operations (n = number of equations = number of unknowns), and for large n, rounding errors may become quite large, whereas in an iteration, errors due to roundoff (or even blunders) may be damped out eventually. In fact, iteration methods are frequently used to improve "solutions" obtained by direct methods.

To apply Banach's theorem, we need a complete metric space and a contraction mapping on it. We take the set X of all ordered n-tuples of real numbers, written

$$x = (\xi_1, \cdots, \xi_n), \qquad y = (\eta_1, \cdots, \eta_n), \qquad z = (\zeta_1, \cdots, \zeta_n),$$

etc. On X we define a metric d by

$$d(x, z) = \max_j |\xi_j - \zeta_j|. \tag{1}$$

$X = (X, d)$ is complete; the simple proof is similar to that in Example 1.5-1.

On X we define T: $X \longrightarrow X$ by

$$y = Tx = Cx + b \tag{2}$$

where $C = (c_{jk})$ is a fixed real $n \times n$ matrix and $b \in X$ a fixed vector. Here and later in this section, all vectors are *column vectors*, because of the usual conventions of matrix multiplication.

Under what condition will T be a contraction? Writing (2) in components, we have

$$\eta_j = \sum_{k=1}^{n} c_{jk}\xi_k + \beta_j \qquad j = 1, \cdots, n,$$

where $b = (\beta_j)$. Setting $w = (\omega_j) = Tz$, we thus obtain from (1) and (2)

$$\begin{aligned} d(y, w) = d(Tx, Tz) &= \max_j |\eta_j - \omega_j| \\ &= \max_j \left| \sum_{k=1}^{n} c_{jk}(\xi_k - \zeta_k) \right| \\ &\leqq \max_i |\xi_i - \zeta_i| \max_j \sum_{k=1}^{n} |c_{jk}| \\ &= d(x, z) \max_j \sum_{k=1}^{n} |c_{jk}|. \end{aligned}$$

We see that this can be written $d(y, w) \leqq \alpha d(x, z)$, where

$$\alpha = \max_j \sum_{k=1}^{n} |c_{jk}|. \tag{3}$$

Banach's theorem 5.1-2 thus yields

5.2-1 Theorem (Linear equations). *If a system*

$$x = Cx + b \qquad (C = (c_{jk}),\ b \text{ given}) \tag{4}$$

of n linear equations in n unknowns $\xi_1, \cdots, \xi_n$ *(the components of x) satisfies*

$$\sum_{k=1}^{n} |c_{jk}| < 1 \qquad (j = 1, \cdots, n), \tag{5}$$

it has precisely one solution x. This solution can be obtained as the limit of the iterative sequence $(x^{(0)}, x^{(1)}, x^{(2)}, \cdots)$, *where* $x^{(0)}$ *is arbitrary and*

$$x^{(m+1)} = Cx^{(m)} + b \qquad m = 0, 1, \cdots. \tag{6}$$

Error bounds are [cf. (3)]

$$d(x^{(m)}, x) \leqq \frac{\alpha}{1-\alpha}\, d(x^{(m-1)}, x^{(m)}) \leqq \frac{\alpha^m}{1-\alpha}\, d(x^{(0)}, x^{(1)}). \tag{7}$$

Condition (5) is sufficient for convergence. It is a **row sum criterion** because it involves *row sums* obtained by summing the absolute values of the elements in a row of C. If we replaced (1) by other metrics, we would obtain other conditions. Two cases of practical importance are included in Probs. 7 and 8.

How is Theorem 5.2-1 related to methods used in practice? A system of n linear equations in n unknowns is usually written

$$Ax = c, \tag{8}$$

where A is an n-rowed square matrix. Many iterative methods for (8) with $\det A \neq 0$ are such that one writes $A = B - G$ with a suitable nonsingular matrix B. Then (8) becomes

$$Bx = Gx + c$$

or

$$x = B^{-1}(Gx + c).$$

This suggests the iteration (6) where

$$C = B^{-1}G, \qquad b = B^{-1}c. \tag{9}$$

Let us illustrate this by two standard methods, the Jacobi iteration, which is largely of theoretical interest, and the Gauss-Seidel iteration, which is widely used in applied mathematics.

5.2-2 Jacobi iteration. This iteration method is defined by

$$\xi_j^{(m+1)} = \frac{1}{a_{jj}}\left(\gamma_j - \sum_{\substack{k=1 \\ k\neq j}}^{n} a_{jk}\xi_k^{(m)}\right) \qquad j = 1, \cdots, n, \tag{10}$$

where $c = (\gamma_j)$ in (8) and we assume $a_{jj} \neq 0$ for $j = 1, \cdots, n$. This iteration is suggested by solving the jth equation in (8) for ξ_j. It is not difficult to verify that (10) can be written in the form (6) with

$$C = -D^{-1}(A - D), \qquad b = D^{-1}c \tag{11}$$

where $D = \operatorname{diag}(a_{jj})$ is the diagonal matrix whose nonzero elements are those of the principal diagonal of A.

Condition (5) applied to C in (11) is sufficient for the convergence of the Jacobi iteration. Since C in (11) is relatively simple, we can express (5) directly in terms of the elements of A. The result is the row sum criterion for the Jacobi iteration

$$\sum_{\substack{k=1 \\ k\neq j}}^{n} \left|\frac{a_{jk}}{a_{jj}}\right| < 1 \qquad j = 1, \cdots, n, \tag{12}$$

or

$$\sum_{\substack{k=1 \\ k\neq j}}^{n} |a_{jk}| < |a_{jj}| \qquad j = 1, \cdots, n. \tag{12*}$$

This shows that, roughly speaking, convergence is guaranteed if the elements in the principal diagonal of A are sufficiently large.

Note that in the Jacobi iteration some components of $x^{(m+1)}$ may already be available at a certain instant but are not used while the computation of the remaining components is still in progress, that is,

all the components of a new approximation are introduced simultaneously at the end of an iterative cycle. We express this fact by saying that the Jacobi iteration is a method of *simultaneous corrections.*

5.2-3 Gauss-Seidel iteration. This is a method of *successive corrections,* in which at every instant all of the latest known components are used. The method is defined by

$$\xi_j^{(m+1)} = \frac{1}{a_{jj}} \left(\gamma_j - \sum_{k=1}^{j-1} a_{jk} \xi_k^{(m+1)} - \sum_{k=j+1}^{n} a_{jk} \xi_k^{(m)} \right), \tag{13}$$

where $j = 1, \cdots, n$ and we again assume $a_{jj} \neq 0$ for all j.

We obtain a matrix form of (13) by writing (Fig. 52)

$$A = -L + D - U$$

where D is as in the Jacobi iteration and L and U are lower and upper triangular, respectively, with principal diagonal elements all zero, the minus signs being a matter of convention and convenience. We now imagine that each equation in (13) is multiplied by a_{jj}. Then we can write the resulting system in the form

$$Dx^{(m+1)} = c + Lx^{(m+1)} + Ux^{(m)}$$

or

$$(D - L)x^{(m+1)} = c + Ux^{(m)}.$$

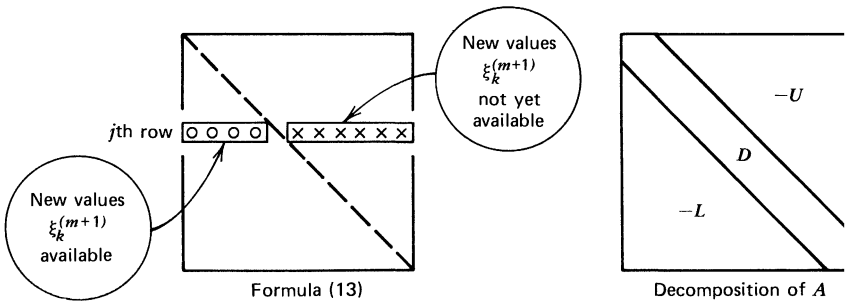


Fig. 52. Explanation of the Gauss-Seidel formulas (13) and (14)

Multiplication by $(D-L)^{-1}$ gives (6) with

$$C=(D-L)^{-1}U, \qquad b=(D-L)^{-1}c. \tag{14}$$

Condition (5) applied to C in (14) is sufficient for the convergence of the Gauss-Seidel iteration. Since C is complicated, the remaining practical problem is to get simpler conditions sufficient for the validity of (5). We mention without proof that (12) is sufficient, but there are better conditions, which the interested reader can find in J. Todd (1962), pp. 494, 495, 500.

Problems

1. Verify (11) and (14).

2. Consider the system

$$\begin{aligned} 5\xi_1 - \xi_2 &= 7 \\ -3\xi_1 + 10\xi_2 &= 24. \end{aligned}$$

(*a*) Determine the exact solution. (*b*) Apply the Jacobi iteration. Does C satisfy (5)? Starting from $x^{(0)}$ with components 1, 1, calculate $x^{(1)}$, $x^{(2)}$ and the error bounds (7) for $x^{(2)}$. Compare these bounds with the actual error of $x^{(2)}$. (*c*) Apply the Gauss-Seidel iteration, performing the same tasks as in (*b*).

3. Consider the system

$$\begin{aligned} \xi_1 - 0.25\xi_2 - 0.25\xi_3 &= 0.50 \\ -0.25\xi_1 + \xi_2 - 0.25\xi_4 &= 0.50 \\ -0.25\xi_1 + \xi_3 - 0.25\xi_4 &= 0.25 \\ -0.25\xi_2 - 0.25\xi_3 + \xi_4 &= 0.25. \end{aligned}$$

(Equations of this form arise in the numerical solution of partial differential equations.) (*a*) Apply the Jacobi iteration, starting from $x^{(0)}$ with components 1, 1, 1, 1 and performing three steps. Compare the approximations with the exact values $\xi_1=\xi_2=0.875$, $\xi_3=\xi_4=0.625$. (*b*) Apply the Gauss-Seidel iteration, performing the same tasks as in (*a*).

4. Gershgorin's theorem states that if λ is an eigenvalue of a square matrix $C=(c_{jk})$, then for some j, where $1 \leqq j \leqq n$,

$$|c_{jj}-\lambda| \leqq \sum_{\substack{k=1 \\ k \neq j}}^{n} |c_{jk}|.$$

(An *eigenvalue* of C is a number λ such that $Cx=\lambda x$ for some $x \neq 0$.) (*a*) Show that (4) can be written $Kx=b$, where $K=I-C$, and Gershgorin's theorem and (5) together imply that K cannot have an eigenvalue 0 (so that K is nonsingular, that is, $\det K \neq 0$, and $Kx=b$ has a unique solution). (*b*) Show that (5) and Gershgorin's theorem imply that C in (6) has spectral radius less than 1. (It can be shown that this is necessary and sufficient for the convergence of the iteration. The *spectral radius* of C is $\max_j |\lambda_j|$, where $\lambda_1, \cdots, \lambda_n$ are the eigenvalues of C.)

5. An example of a system for which the Jacobi iteration diverges whereas the Gauss-Seidel iteration converges is

$$\begin{aligned} 2\xi_1 + \xi_2 + \xi_3 &= 4 \\ \xi_1 + 2\xi_2 + \xi_3 &= 4 \\ \xi_1 + \xi_2 + 2\xi_3 &= 4. \end{aligned}$$

Starting from $x^{(0)}=0$, verify divergence of the Jacobi iteration and perform the first few steps of the Gauss-Seidel iteration to obtain the impression that the iteration seems to converge to the exact solution $\xi_1=\xi_2=\xi_3=1$.

6. It is plausible to think that the Gauss-Seidel iteration is better than the Jacobi iteration in all cases. Actually, the two methods are not comparable. This is surprising. For example, in the case of the system

$$\begin{aligned} \xi_1 \quad\quad + \xi_3 &= 2 \\ -\xi_1 + \xi_2 \quad\quad &= 0 \\ \xi_1 + 2\xi_2 - 3\xi_3 &= 0 \end{aligned}$$

the Jacobi iteration converges whereas the Gauss-Seidel iteration diverges. Derive these two facts from the necessary and sufficient conditions stated in Prob. 4(*b*).

7. (Column sum criterion) To the metric in (1) there corresponds the condition (5). If we use on X the metric d_1 defined by

$$d_1(x, z) = \sum_{j=1}^{n} |\xi_j - \zeta_j|,$$

show that instead of (5) we obtain the condition

$$\sum_{j=1}^{n} |c_{jk}| < 1 \qquad (k = 1, \cdots, n). \tag{15}$$

8. (Square sum criterion) To the metric in (1) there corresponds the condition (5). If we use on X the Euclidean metric d_2 defined by

$$d_2(x, z) = \left[\sum_{j=1}^{n} (\xi_j - \zeta_j)^2\right]^{1/2},$$

show that instead of (5) we obtain the condition

$$\sum_{j=1}^{n} \sum_{k=1}^{n} c_{jk}^2 < 1. \tag{16}$$

9. (Jacobi iteration) Show that for the Jacobi iteration the sufficient convergence conditions (5), (15) and (16) take the form

$$\sum_{\substack{k=1 \\ k \neq j}}^{n} \frac{|a_{jk}|}{|a_{jj}|} < 1, \qquad \sum_{\substack{j=1 \\ j \neq k}}^{n} \frac{|a_{jk}|}{|a_{jj}|} < 1, \qquad \sum_{\substack{j=1 \\ j \neq k}}^{n} \sum_{k=1}^{n} \frac{a_{jk}^2}{a_{jj}^2} < 1.$$

10. Find a matrix C which satisfies (5) but neither (15) nor (16).

5.3 Application of Banach's Theorem to Differential Equations

The most interesting applications of Banach's fixed point theorem arise in connection with function spaces. The theorem then yields existence and uniqueness theorems for differential and integral equations, as we shall see.

In fact, in this section let us consider an *explicit ordinary differential equation of the first order*

(1a) $$x' = f(t, x) \qquad (' = d/dt).$$

An **initial value problem** for such an equation consists of the equation and an *initial condition*

(1b) $$x(t_0) = x_0$$

where t_0 and x_0 are given real numbers.

We shall use Banach's theorem to prove the famous Picard's theorem which, while not the strongest of its type that is known, plays a vital role in the theory of ordinary differential equations. The idea of approach is quite simple: (1) will be converted to an integral equation, which defines a mapping T, and the conditions of the theorem will imply that T is a contraction such that its fixed point becomes the solution of our problem.

5.3-1 Picard's Existence and Uniqueness Theorem (Ordinary differential equations). *Let f be continuous on a rectangle* (Fig. 53)

$$R = \{(t, x) \mid |t - t_0| \leqq a, |x - x_0| \leqq b\}$$

and thus bounded on R, say (see Fig. 54)

(2) $$|f(t, x)| \leqq c \qquad \text{for all } (t, x) \in R$$

Suppose that f satisfies a **Lipschitz condition** *on R with respect to its second argument, that is, there is a constant k* (Lipschitz constant) *such*

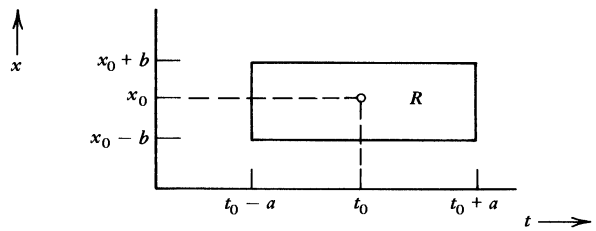
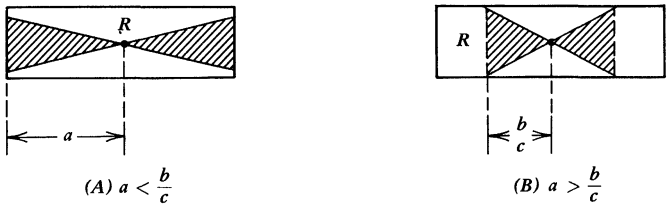


Fig. 53. The rectangle R



(A) $a < \frac{b}{c}$ (B) $a > \frac{b}{c}$

Fig. 54. Geometric illustration of inequality (2) for (*A*) relatively small *c*, (*B*) relatively large *c*. The solution curve must remain in the shaded region bounded by straight lines with slopes $\pm c$.

that for (t, x), $(t, v) \in R$

$$|f(t, x) - f(t, v)| \leqq k\,|x - v|. \tag{3}$$

Then the initial value problem (1) *has a unique solution. This solution exists on an interval* $[t_0 - \beta, t_0 + \beta]$, *where*[1]

$$\beta < \min\left\{a, \frac{b}{c}, \frac{1}{k}\right\}. \tag{4}$$

Proof. Let $C(J)$ be the metric space of all real-valued continuous functions on the interval $J = [t_0 - \beta, t_0 + \beta]$ with metric d defined by

$$d(x, y) = \max_{t \in J} |x(t) - y(t)|.$$

$C(J)$ is complete, as we know from 1.5-5. Let $\tilde{C}$ be the subspace of $C(J)$ consisting of all those functions $x \in C(J)$ that satisfy

$$|x(t) - x_0| \leqq c\beta. \tag{5}$$

It is not difficult to see that $\tilde{C}$ is closed in $C(J)$ (cf. Prob. 6), so that $\tilde{C}$ is complete by 1.4-7.

By integration we see that (1) can be written $x = Tx$, where $T\colon \tilde{C} \longrightarrow \tilde{C}$ is defined by

$$Tx(t) = x_0 + \int_{t_0}^{t} f(\tau, x(\tau))\, d\tau. \tag{6}$$

[1] In the classical proof, $\beta < \min\{a, b/c\}$, which is better. This could also be obtained by a modification of the present proof (by the use of a more complicated metric); cf. A. Bielecki (1956) in the references in Appendix 3.

Indeed, T is defined for all $x \in \tilde{C}$, because $c\beta < b$ by (4), so that if $x \in \tilde{C}$, then $\tau \in J$ and $(\tau, x(\tau)) \in R$, and the integral in (6) exists since f is continuous on R. To see that T maps $\tilde{C}$ into itself, we can use (6) and (2), obtaining

$$|Tx(t) - x_0| = \left| \int_{t_0}^{t} f(\tau, x(\tau))\, d\tau \right| \leqq c\,|t - t_0| \leqq c\beta.$$

We show that T is a contraction on $\tilde{C}$. By the Lipschitz condition (3),

$$\begin{aligned} |Tx(t) - Tv(t)| &= \left| \int_{t_0}^{t} [f(\tau, x(\tau)) - f(\tau, v(\tau))]\, d\tau \right| \\ &\leqq |t - t_0| \max_{\tau \in J} k\,|x(\tau) - v(\tau)| \\ &\leqq k\beta\, d(x, v). \end{aligned}$$

Since the last expression does not depend on t, we can take the maximum on the left and have

$$d(Tx, Tv) \leqq \alpha d(x, v) \qquad \text{where} \qquad \alpha = k\beta.$$

From (4) we see that $\alpha = k\beta < 1$, so that T is indeed a contraction on $\tilde{C}$. Theorem 5.1-2 thus implies that T has a unique fixed point $x \in \tilde{C}$, that is, a continuous function x on J satisfying $x = Tx$. Writing $x = Tx$ out, we have by (6)

$$x(t) = x_0 + \int_{t_0}^{t} f(\tau, x(\tau))\, d\tau. \tag{7}$$

Since $(\tau, x(\tau)) \in R$ where f is continuous, (7) may be differentiated. Hence x is even differentiable and satisfies (1). Conversely, every solution of (1) must satisfy (7). This completes the proof. ∎

Banach's theorem also implies that the solution x of (1) is the limit of the sequence $(x_0, x_1, \cdots)$ obtained by the *Picard iteration*

$$x_{n+1}(t) = x_0 + \int_{t_0}^{t} f(\tau, x_n(\tau))\, d\tau \tag{8}$$

where $n = 0, 1, \cdots$. However, the practical usefulness of this way of obtaining approximations to the solution of (1) and corresponding error bounds is rather limited because of the integrations involved.

We finally mention the following. It can be shown that continuity of f is sufficient (but not necessary) for the existence of a solution of the problem (1), but is not sufficient for uniqueness. A Lipschitz condition is sufficient (as Picard's theorem shows), but not necessary. For details, see the book by E. L. Ince (1956), which also contains historical remarks about Picard's theorem (on p. 63) as well as a classical proof, so that the reader may compare our present approach with the classical one.

Problems

1. If the partial derivative $\partial f/\partial x$ of f exists and is continuous on the rectangle R (cf. Picard's theorem), show that f satisfies a Lipschitz condition on R with respect to its second argument.

2. Show that f defined by $f(t, x) = |\sin x| + t$ satisfies a Lipschitz condition on the whole tx-plane with respect to its second argument, but $\partial f/\partial x$ does not exist when $x = 0$. What fact does this illustrate?

3. Does f defined by $f(t, x) = |x|^{1/2}$ satisfy a Lipschitz condition?

4. Find all initial conditions such that the initial value problem $tx' = 2x$, $x(t_0) = x_0$, has (*a*) no solutions, (*b*) more than one solution, (*c*) precisely one solution.

5. Explain the reasons for the restrictions $\beta < b/c$ and $\beta < 1/k$ in (4).

6. Show that in the proof of Picard's theorem, $\tilde{C}$ is closed in $C(J)$.

7. Show that in Picard's theorem, instead of the constant x_0 we can take any other function $y_0 \in \tilde{C}$, $y_0(t_0) = x_0$, as the initial function of the iteration.

8. Apply the Picard iteration (8) to $x' = 1 + x^2$, $x(0) = 0$. Verify that for x_3, the terms involving t, t^2, $\cdots$, t^5 are the same as those of the exact solution.

9. Show that $x' = 3x^{2/3}$, $x(0) = 0$, has infinitely many solutions x, given by

$$x(t) = 0 \quad \text{if } t < c, \qquad x(t) = (t - c)^3 \quad \text{if } t \geqq c,$$

where $c > 0$ is any constant. Does $3x^{2/3}$ on the right satisfy a Lipschitz condition?

10. Show that solutions of the initial value problem

$$x' = |x|^{1/2}, \qquad x(0) = 0$$

are $x_1 = 0$ and x_2, where $x_2(t) = t\,|t|/4$. Does this contradict Picard's theorem? Find further solutions.

5.4 Application of Banach's Theorem to Integral Equations

We finally consider the Banach fixed point theorem as a source of existence and uniqueness theorems for integral equations. An integral equation of the form

$$\textbf{(1)} \qquad x(t) - \mu \int_a^b k(t, \tau) x(\tau)\, d\tau = v(t)$$

is called a **Fredholm equation** *of the second kind.*[2] Here, $[a, b]$ is a given interval. x is a function on $[a, b]$ which is unknown. μ is a parameter. The **kernel** k of the equation is a given function on the square $G = [a, b] \times [a, b]$ shown in Fig. 55, and v is a given function on $[a, b]$.

Integral equations can be considered on various function spaces. In this section we consider (1) on $C[a, b]$, the space of all continuous functions defined on the interval $J = [a, b]$ with metric d given by

$$(2) \qquad d(x, y) = \max_{t \in J} |x(t) - y(t)|;$$

[2] The presence of the term $x(t)$ enables us to apply iteration, as Theorem 5.4-1 shows. An equation without that term is of the form

$$\int_a^b k(t, \tau) x(\tau)\, d\tau = v(t)$$

and is said to be of the *first kind.*

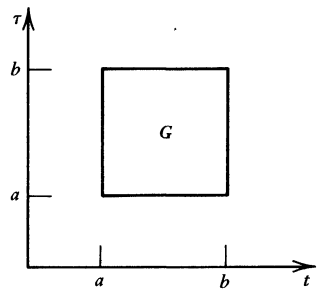


Fig. 55. Domain of definition G of the kernel k in the integral equation (1) in the case of positive a and b

cf. 1.5-5. For the proposed application of Banach's theorem it is important to note that $C[a, b]$ is complete. We assume that $v \in C[a, b]$ and k is continuous on G. Then k is a bounded function on G, say,

$$|k(t, \tau)| \leqq c \qquad \text{for all } (t, \tau) \in G. \tag{3}$$

Obviously, (1) can be written $x = Tx$ where

$$Tx(t) = v(t) + \mu \int_a^b k(t, \tau)x(\tau)\, d\tau. \tag{4}$$

Since v and k are continuous, formula (4) defines an operator T: $C[a, b] \longrightarrow C[a, b]$. We now impose a restriction on μ such that T becomes a contraction. From (2) to (4) we have

$$\begin{aligned} d(Tx, Ty) &= \max_{t \in J} |Tx(t) - Ty(t)| \\ &= |\mu| \max_{t \in J} \left| \int_a^b k(t, \tau)[x(\tau) - y(\tau)]\, d\tau \right| \\ &\leqq |\mu| \max_{t \in J} \int_a^b |k(t, \tau)|\, |x(\tau) - y(\tau)|\, d\tau \\ &\leqq |\mu|\, c \max_{\sigma \in J} |x(\sigma) - y(\sigma)| \int_a^b d\tau \\ &= |\mu|\, c\, d(x, y)(b - a). \end{aligned}$$

This can be written $d(Tx, Ty) \leqq \alpha\, d(x, y)$, where

$$\alpha = |\mu|\, c(b-a).$$

We see that T becomes a contraction ($\alpha < 1$) if

$$|\mu| < \frac{1}{c(b-a)}. \tag{5}$$

Banach's fixed point theorem 5.1-2 now gives

5.4-1 Theorem (Fredholm integral equation). *Suppose k and v in* (1) *to be continuous on $J \times J$ and $J = [a, b]$, respectively, and assume that μ satisfies* (5) *with c defined in* (3). *Then* (1) *has a unique solution x on J. This function x is the limit of the iterative sequence $(x_0, x_1, \cdots)$, where x_0 is any continuous function on J and for $n = 0, 1, \cdots$,*

$$x_{n+1}(t) = v(t) + \mu \int_a^b k(t, \tau) x_n(\tau)\, d\tau. \tag{6}$$

Fredholm's famous theory of integral equations will be discussed in Chap. 8.

We now consider the **Volterra integral equation**

$$x(t) - \mu \int_a^t k(t, \tau) x(\tau)\, d\tau = v(t). \tag{7}$$

The difference between (1) and (7) is that in (1) the upper limit of integration b is constant, whereas here in (7) it is variable. This is essential. In fact, without any restriction on μ we now get the following existence and uniqueness theorem.

5.4-2 Theorem (Volterra integral equation). *Suppose that v in* (7) *is continuous on $[a, b]$ and the kernel k is continuous on the triangular region R in the $t\tau$-plane given by $a \leqq \tau \leqq t$, $a \leqq t \leqq b$; see Fig.* 56. *Then* (7) *has a unique solution x on $[a, b]$ for every μ.*

Proof. We see that equation (7) can be written $x = Tx$ with $T\colon C[a, b] \longrightarrow C[a, b]$ defined by

$$Tx(t) = v(t) + \mu \int_a^t k(t, \tau) x(\tau)\, d\tau. \tag{8}$$

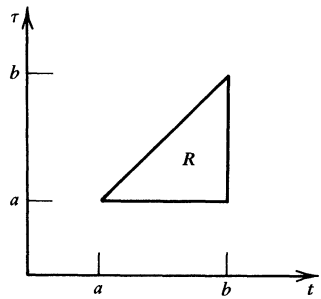


Fig. 56. Triangular region R in Theorem 5.4-2 in the case of positive a and b

Since k is continuous on R and R is closed and bounded, k is a bounded function on R, say,

$$|k(t, \tau)| \leqq c \qquad \text{for all } (t, \tau) \in R.$$

Using (2), we thus obtain for all $x, y \in C[a, b]$

$$|Tx(t) - Ty(t)| = |\mu| \left| \int_a^t k(t, \tau)[x(\tau) - y(\tau)]\, d\tau \right|$$

(9)
$$\leqq |\mu|\, c\, d(x, y) \int_a^t d\tau$$

$$= |\mu|\, c(t-a)\, d(x, y).$$

We show by induction that

(10) $$|T^m x(t) - T^m y(t)| \leqq |\mu|^m c^m \frac{(t-a)^m}{m!}\, d(x, y).$$

For $m = 1$ this is (9). Assuming that (10) holds for any m, we obtain from (8)

$$|T^{m+1}x(t) - T^{m+1}y(t)| = |\mu| \left| \int_a^t k(t, \tau)[T^m x(\tau) - T^m y(\tau)]\, d\tau \right|$$

$$\leqq |\mu|\, c \int_a^t |\mu|^m\, c^m \frac{(\tau-a)^m}{m!}\, d\tau\, d(x, y)$$

$$= |\mu|^{m+1}\, c^{m+1} \frac{(t-a)^{m+1}}{(m+1)!}\, d(x, y),$$

which completes the inductive proof of (10).

Using $t-a \leqq b-a$ on the right-hand side of (10) and then taking the maximum over $t \in J$ on the left, we obtain from (10)

$$d(T^m x, T^m y) \leqq \alpha_m d(x, y)$$

where

$$\alpha_m = |\mu|^m c^m \frac{(b-a)^m}{m!}.$$

For any fixed μ and sufficiently large m we have $\alpha_m < 1$. Hence the corresponding T^m is a contraction on $C[a, b]$, and the assertion of our theorem follows from

5.4-3 Lemma (Fixed point). *Let T: $X \longrightarrow X$ be a continuous mapping* (cf. 1.3-3) *on a complete metric space $X = (X, d)$, and suppose that T^m is a contraction on X for some positive integer m. Then T has a unique fixed point.*

Proof. By assumption, $B = T^m$ is a contraction on X, that is, $d(Bx, By) \leqq \alpha d(x, y)$ for all $x, y \in X$; here $\alpha < 1$. Hence for every $x_0 \in X$,

$$\begin{aligned} d(B^n Tx_0, B^n x_0) &\leqq \alpha d(B^{n-1} Tx_0, B^{n-1} x_0) \\ &\cdots \leqq \alpha^n d(Tx_0, x_0) \longrightarrow 0 \qquad (n \longrightarrow \infty). \end{aligned} \tag{11}$$

Banach's theorem 5.1-2 implies that B has a unique fixed point, call it x, and $B^n x_0 \longrightarrow x$. Since the mapping T is continuous, this implies $B^n Tx_0 = TB^n x_0 \longrightarrow Tx$. Hence by 1.4-2(*b*),

$$d(B^n Tx_0, B^n x_0) \longrightarrow d(Tx, x),$$

so that $d(Tx, x) = 0$ by (11). This shows that x is a fixed point of T. Since every fixed point of T is also a fixed point of B, we see that T cannot have more than one fixed point. ∎

We finally note that a Volterra equation can be regarded as a special Fredholm equation whose kernel k is zero in the part of the square $[a, b] \times [a, b]$ where $\tau > t$ (see Figs. 55 and 56) and may not be continuous at points on the diagonal ($\tau = t$).

Problems

1. Solve by iteration, choosing $x_0 = v$:

$$x(t) - \mu \int_0^1 e^{t-\tau} x(\tau)\, d\tau = v(t) \qquad (|\mu| < 1).$$

2. (Nonlinear integral equation) If v and k are continuous on $[a, b]$ and $C = [a, b] \times [a, b] \times \mathbf{R}$, respectively, and k satisfies on G a Lipschitz condition of the form

$$|k(t, \tau, u_1) - k(t, \tau, u_2)| \leqq l\, |u_1 - u_2|,$$

show that the nonlinear integral equation

$$x(t) - \mu \int_a^b k(t, \tau, x(\tau))\, d\tau = v(t)$$

has a unique solution x for any μ such that $|\mu| < 1/l(b-a)$.

3. It is important to understand that integral equations also arise from problems in differential equations. (*a*) For example, write the initial value problem

$$\frac{dx}{dt} = f(t, x), \qquad x(t_0) = x_0$$

as an integral equation and indicate what kind of equation it is. (*b*) Show that an initial value problem

$$\frac{d^2x}{dt^2} = f(t, x), \qquad x(t_0) = x_0, \qquad x'(t_0) = x_1$$

involving a second order differential equation can be transformed into a Volterra integral equation.

4. (Neumann series) Defining an operator S by

$$Sx(t) = \int_a^b k(t, \tau) x(\tau)\, d\tau$$

and setting $z_n = x_n - x_{n-1}$, show that (6) implies

$$z_{n+1} = \mu S z_n.$$

Choosing $x_0 = v$, show that (6) yields the *Neumann series*

$$x = \lim_{n\to\infty} x_n = v + \mu S v + \mu^2 S^2 v + \mu^3 S^3 v + \cdots.$$

5. Solve the following integral equation (*a*) by a Neumann series, (*b*) by a direct approach.

$$x(t) - \mu \int_0^1 x(\tau)\, d\tau = 1.$$

6. Solve

$$x(t) - \mu \int_a^b c x(\tau)\, d\tau = \tilde{v}(t)$$

where c is a constant, and indicate how the corresponding Neumann series can be used to obtain the convergence condition (5) for the Neumann series of (1).

7. (Iterated kernel, resolvent kernel) Show that in the Neumann series in Prob. 4 we can write

$$(S^n v)(t) = \int_a^b k_{(n)}(t, \tau) v(\tau)\, d\tau \qquad n = 2, 3, \cdots,$$

where the *iterated kernel* $k_{(n)}$ is given by

$$k_{(n)}(t, \tau) = \int_a^b \cdots \int_a^b k(t, t_1) k(t_1, t_2) \cdots k(t_{n-1}, \tau)\, dt_1 \cdots dt_{n-1}$$

so that the Neumann series can be written

$$x(t) = v(t) + \mu \int_a^b k(t, \tau) v(\tau)\, d\tau + \mu^2 \int_a^b k_{(2)}(t, \tau) v(\tau)\, d\tau + \cdots$$

or, by the use of the *resolvent kernel* $\tilde{k}$ defined by

$$\tilde{k}(t, \tau, \mu) = \sum_{j=0}^{\infty} \mu^j k_{(j+1)}(t, \tau) \qquad (k_{(1)} = k)$$

it can be written

$$x(t) = v(t) + \mu \int_a^b \tilde{k}(t, \tau, \mu) v(\tau)\, d\tau.$$

8. It is of interest that the Neumann series in Prob. 4 can also be obtained by substituting a power series in μ,

$$x(t) = v_0(t) + \mu v_1(t) + \mu^2 v_2(t) + \cdots$$

into (1), integrating termwise and comparing coefficients. Show that this gives

$$v_0(t) = v(t), \qquad v_n(t) = \int_a^b k(t, \tau) v_{n-1}(\tau)\, d\tau, \qquad n = 1, 2, \cdots.$$

Assuming that $|v(t)| \leqq c_0$ and $|k(t, \tau)| \leqq c$, show that

$$|v_n(t)| \leqq c_0[c(b-a)]^n,$$

so that (5) implies convergence.

9. Using Prob. 7, solve (1), where $a = 0$, $b = 2\pi$ and

$$k(t, \tau) = \sum_{n=1}^{N} a_n \sin nt \cos n\tau.$$

10. In (1), let $a = 0$, $b = \pi$ and

$$k(t, \tau) = a_1 \sin t \sin 2\tau + a_2 \sin 2t \sin 3\tau.$$

Write the solution in terms of the resolvent kernel (cf. Prob. 7).