1

Preliminaries

To facilitate readability of the text we shall frequently recall basic definitions and facts as the need arises. Nevertheless, certain topics, such as those related to reflexivity of Banach spaces and the weak and weak* topologies, play such a primary role in our development that they deserve special attention. It is our purpose in this chapter to collect facts which will be used repeatedly throughout the text.

For the most part we use notation and symbols commonly used in textbooks. We only mention that the symbols $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{C}$ are reserved, respectively, for the natural numbers, integers, real numbers, and complex numbers.

If A is a subset of a metric space (M, ρ) and if $x \in M$, then diam A and dist(x, A) denote, respectively, the diameter of A and the distance from x to A. Precisely,

$$\operatorname{diam} A = \sup\{\rho(x, y) : x, y \in A\};$$
$$\operatorname{dist}(x, A) = \inf\{\rho(x, y) : y \in A\}.$$

Also $B(x; r)$ always denotes the *closed* ball centered at x with radius $r > 0$:

$$B(x; r) = \{y \in M : \rho(x, y) \leqslant r\}.$$

We use the symbol $\bar{A}$ to denote the closure of A in M, and if M has a topology τ other than the one induced by the metric, we use ${}^{\tau}\bar{A}$ to denote the closure of A in (M, τ).

For the remainder of the chapter we confine our attention to Banach spaces. All the above definitions carry over to a Banach space setting $(X, \|\ \|)$ by taking $\rho(x, y) = \|x - y\|$.

We remark at the outset that while most of the results presented in this book remain valid for complex Banach spaces, we shall, for convenience, confine our attention to real Banach spaces. Throughout the remainder of the chapter $X = (X, \|\ \|)$ will denote an arbitrary real Banach space.

Convexity We use conv A to denote the *convex hull* of $A \subset X$, i.e., the smallest convex subset of X which contains A. Obviously

$$\operatorname{conv} A = \cap\{K \subset X : K \supset A, K \text{ is convex}\}.$$

Moreover, $x \in \operatorname{conv} A$ if and only if x is of the form $x = \Sigma_{i=1}^{n} \lambda_i x_i$ where $x_i \in A$, $\lambda_i \geqslant 0$ and $\Sigma_{i=1}^{n} \lambda_i = 1$. The closure of conv A is denoted $\overline{\operatorname{conv}}\, A$ and called the *convex closure* of A. As above,

$$\overline{\operatorname{conv}}\, A = \cap \{K \subset X : K \supset A, K \text{ is closed and convex}\}.$$

A fundamental property of the convex closure operation is given by the following:

Theorem 1.1 (Mazur's) *If $\bar{A}$ is compact, then so is $\overline{\operatorname{conv}}\, A$.*

Dual spaces and reflexivity For two Banach spaces X and Y, let $\mathscr{L}(X, Y)$ denote the space of all bounded (continuous) linear operators (mappings) from X to Y. The norm $\|T\|$ of an operator $T \in \mathscr{L}(X, Y)$ is given by

$$\begin{aligned} \|T\| &= \sup\{\|Tx\|/\|x\| : x \in X, x \neq 0\} \\ &= \sup\{\|Tx\| : x \in X, \|x\| = 1\}. \end{aligned}$$

It is easily verified that if X and Y are Banach spaces then so is $(\mathscr{L}(X, Y), \|\ \|)$. The *dual* or *conjugate space* X^* of X is the space $X^* = \mathscr{L}(X, \mathbb{R})$. The elements of X^* are called *continuous linear functionals*. For $x^* \in X^*$, we shall often use the notation which pairs elements of X with elements of X^*:

$$x^*(x) = \langle x, x^* \rangle, x \in X, x^* \in X^*.$$

(This of course is consistent with the pairing in a Hilbert space H where, by the Riesz Representation Theorem, $H = H^*$ and the value of $y(x)$ for $x \in H$, $y \in H^*$ is given by the usual inner product $\langle x, y \rangle$.)

The space $X^{**} = \mathscr{L}(X^*, \mathbb{R})$ is called the second dual (or conjugate) space of X. If $x \in X$ is fixed, then the relation $\langle x, x^* \rangle$ defines a continuous linear functional on X^*; thus x is associated in a natural way with an element x^{**} of X^{**}. The mapping $x \mapsto x^{**}$ is called the *canonical* (or *natural*) embedding of X in X^{**}. This embedding is always a linear isometry. If it is also surjective then X is said to be *reflexive* and we write $X = X^{**}$.

The weak and weak* topologies The *weak topology on* X is the topology generated by the family of seminorms $\{p_{x^*}\}$, $x^* \in X^*$, where

$$p_{x^*}(x) = |\langle x, x^* \rangle|, \qquad x \in X.$$

Similarly, the *weak* topology on* X^* is generated by the seminorms $\{p_x\}$, $x \in X$, where

$$p_x(x) = |\langle x, x^* \rangle|, \qquad x^* \in X^*.$$

Both X and X^* are locally convex, linear topological spaces relative to their respective weak and weak* topologies. Note also that X^* also has a weak topology which in general is distinct from its weak* topology. The two, of course, coincide when X is reflexive.

The weak topology on X is the weakest (coarsest) topology for which all the functionals $x^* \in X^*$ are continuous. In particular, a net $\{x_\alpha, \alpha \in A\}$ converges to an element $x \in X$ in the weak topology if and only if $\lim_\alpha \langle x_\alpha, x^* \rangle = \langle x, x^* \rangle$ for each $x^* \in X^*$. When this occurs we say that $\{x_\alpha\}$ is *weakly convergent* or *converges weakly* to x, and we write

$$w\text{-}\lim_{\alpha} x_\alpha = x.$$

Similarly, a net $\{x^*_\alpha, \alpha \in A\}$ in X^* converges to $x^* \in X^*$ in the weak* topology if and only if for each $x \in X$, $\lim_\alpha \langle x, x^*_\alpha \rangle = \langle x, x^* \rangle$, in which case

$$w^*\text{-}\lim_{\alpha} x^*_\alpha = x^*.$$

(We elaborate on the notion of net convergence in Chapter 14.)

We now collect some basic and well-known properties about the weak and weak* topologies.

Property 1.1 *A convex subset K of X is closed if and only if it is weakly closed.*

Property 1.2 *If K is a weakly compact subset of X then* $\overline{\text{conv}}\, K$ *is also weakly compact.*

The above facts do not carry over to the weak* topology. However the following fact about the weak* topology is very important.

Property 1.3 (Alaoglu's Theorem) *The unit ball $B(0; 1)$ in a dual space X^* is always compact in the weak* topology.*

Note that the above implies that any ball or any intersection of balls in a dual space is weak* compact.

If X is reflexive, then $X = X^{**}$. Thus, in view of Alaoglu's Theorem, we have:

Property 1.4 *If X is reflexive, then each ball in X is compact in the weak topology.*

The above, in fact, characterizes reflexive spaces: A Banach space X is reflexive if and only if its unit ball is compact in the weak topology. When combined with Properties 1.1 and 1.2 this fact implies that each bounded, closed and convex subset of a reflexive space is compact in the weak topology (i.e., weakly compact).

The following rather deep fact is of fundamental importance. We invoke this fact repeatedly and without comment.

Property 1.5 (*The Eberlein–Smulian Theorem*) *For any subset A of X the following statements are equivalent.*

(a) *Each sequence $\{x_n\}$ in A has a subsequence that is weakly convergent.*
(b) *Each sequence $\{x_n\}$ in A has a weak cluster point in X.*
(c) *The weak closure ${}^w\bar{A}$ of A is weakly compact.*

Thus weak compactness is always equivalent to sequential weak compactness. This fact does not hold for the weak* topology. However the following is true.

Property 1.6 (*The Krein–Smulian Theorem*) *A subset K of a dual space X^* is weak* closed if and only if for each $r>0$ the sets $\{x^* \in K: \|x^*\| \leqslant r\}$ are weak* closed.*

The fact that the weak* topology on a separable space is metrizable gives rise to the following useful fact.

Property 1.7 *If X is separable and if K is a convex subset of X^*, then K is weak* closed if and only if K is weak* sequentially closed.*

Proofs of the above facts may be found, for example, in Dunford and Schwartz (1957) or in almost any standard text in functional analysis. For our purposes, however, the mere knowledge of these facts will enable the reader to follow all the steps in the proofs which follow.

We will now list several properties which characterize reflexivity. Other such characterizations will be introduced in the text as needed.

Property 1.8 *A Banach space X is reflexive if and only if one of the following (equivalent) conditions holds.*

(a) *X^* is reflexive.*
(b) *$B(0; 1)$ is weakly compact in X^*.*
(c) *Any bounded sequence in X has a weakly convergent subsequence.*

(d) (James (1957)). *For any* $x^* \in X^*$ *there exists* $x \in B(0; 1)$ *such that* $x^*(x) = \|x^*\|$.
(e) (James (1964)). *For any bounded, closed and convex subset* K *of* X *and any* $x^* \in X^*$ *there exists* $x \in K$ *such that* $x^*(x) = \sup\{x^*(y) : y \in K\}$.
(f) (Smulian (1939)). *For any decreasing sequence* $\{K_n\}$ *of nonempty, bounded, closed and convex subsets of* X, $\bigcap_{n=1}^{\infty} K_n \neq \emptyset$.

There is one important fact that one should keep in mind when dealing with the weak* topology on a dual space X^*. Namely, two different Banach spaces, X_1 and X_2 may have the same dual space: $X^* = X_1^* = X_2^*$ (where the dual norms on X_1^* and X_2^* coincide). However the weak* topologies induced on X^* by X_1 and X_2 may differ. This occurs, for example, with the classical space l^1 which is dual to both c (the space of convergent real sequences) and c_0 (the space of such sequences which converge to 0).

We also assume the reader is familiar with the classical Banach spaces which arise frequently in the literature (e.g., the l^p- and L^p- spaces, $1 \leqslant p \leqslant \infty$, $\mathscr{C}[0, 1]$, c, c_0, etc.) along with the facts: $(L^p)^*$ and $(l^p)^* (1 < p < \infty)$ are represented, respectively, by L^q and l^q where $p^{-1} + q^{-1} = 1$, $(l^1)^*$ is represented by l^∞, etc.

Finally, we shall require some basic facts about Schauder bases. A sequence $\{e^n\}$ in a Banach space X is called a *Schauder basis* for X if for any $x \in X$ there is a unique sequence $\{\xi_n\} = \{\xi_n(x)\}$ of real numbers such that

$$x = \sum_{n=1}^{\infty} \xi_n e^n = \lim_{k \to \infty} \sum_{n=1}^{k} \xi_n e^n.$$

Obviously if X has such a basis then all the e^n are linearly independent. Also there exist two sequences $\{P_k\}$ and $\{R_k\}$ of natural operations; the projections:

$$P_k : x \mapsto \sum_{n=1}^{k} \xi_n e^n;$$

and the remainders:

$$R_k : x \mapsto \sum_{n=k+1}^{k} \xi_n e^n.$$

The family $\{P_k\}$ (as well as $\{R_k\}$) is equicontinuous (Banach-Steinhaus Theorem) and the constant $K = \sup_n \|P_n\|$ is called the basis constant for X. Moreover, for any $x \in X$, $\lim_{n \to \infty} P_n x = x$; $\lim_{n \to \infty} R_n x = 0$. Consequently, any linear functional $f \in X^*$ has the following representation:

$$fx = f\left(\sum_{n=1}^{\infty} \xi_n e^n\right) = \sum_{n=1}^{\infty} \xi_n f(e^n).$$

Thus f is completely determined by its values at the basis elements. In particular, the coordinate functionals $\{f_n\}$ defined by the relation $f_n(e^i)=\delta_{in}$ are *biorthogonal* to the basis $\{e^n\}$. The family $\{f_n\}$ is equibounded, and the value $\xi_n = f_n(x)$ is called the nth coordinate of x with respect to the basis $\{e^n\}$. We shall call a sequence $\{x_n\}$ in X *coordinate-wise convergent* if for each $n \in \mathbb{N}$, the sequence $\{f_n(x_k)\}_{k=1}^{\infty}$ converges. (In general this type of convergence is not equivalent to either weak or weak* convergence.)

The sequence $\{f_n\}$ of coordinate functionals is not necessarily a basis for the space X^*. However this is the case if the basis $\{e^n\}$ has the following property: For any $f \in X^*$ the norm of f restricted to span $\{e^k, e^{k+1}, \ldots\}$ converges to 0 as $k \mapsto \infty$. Bases which have this property are said to be *shrinking*. Also, a basis $\{e^n\}$ of X is said to be *boundedly complete* if for every sequence $\{\xi_n\}$ of scalars for which $\sup_k \|\Sigma_{n=1}^k \xi_n e^n\| < \infty$, the series $\Sigma_{n=1}^{\infty} \xi_n e^n$ converges, and thus represents an element of X. These concepts provide a connection between basis theory and reflexivity via the following result.

Property 1.9 (James (1950)). *A Banach space X with a basis $\{e^n\}$ is reflexive if and only if $\{e^n\}$ is shrinking and boundedly complete.*

In view of Property 1.9 a sequence $\{x_k\}$ of elements of a reflexive Banach space which has a basis converges weakly if and only if it is bounded and coordinate-wise convergent.

We shall use the above facts mostly in special settings, for example, in such spaces as c_0, l^1, l^p $(1<p<\infty)$ with the natural basis $\{e^n\}=\{\delta_{in}\}$. We note, in particular, that $\{e^n\}$ is shrinking but *not* boundedly complete in c_0 and l^1.

For those interested in proof of most of the above facts, we suggest the book by van Dulst (1978) which is devoted exclusively to reflexive spaces. Also, persons interested in knowing more about the geometry of Banach spaces might wish to consult such books as Day (1973), Lindenstrauss and Tzafriri (1977, 1979) or Diestel (1975).

2

Banach's Contraction Principle

The fixed point theorem, generally known as the Banach Contraction Principle, appeared in explicit form in Banach's thesis in 1922 where it was used to establish the existence of a solution for an integral equation. Since then, because of its simplicity and usefulness, it has become a very popular tool in solving existence problems in many branches of mathematical analysis. In this chapter we prove Banach's Contraction Principle, discuss some of its more useful variants, and present a few diverse examples of its applications.

Let M be a metric space with distance function (metric) ρ. A mapping $T: M \to M$ is said to be *lipschitzian* if there exists $k \geqslant 0$ such that for all $x, y \in M$,

$$\rho(Tx, Ty) \leqslant k\rho(x, y). \tag{2.1}$$

The smallest k for which (2.1) holds is said to be the *Lipschitz constant* for T. We shall often denote the respective Lipschitz constants of different mappings T and S by $k(T)$ and $k(S)$ and when relevant $k_\rho(T)$ will be used to denote the Lipschitz constant of T with respect to the metric ρ.

For two mappings $S, T: M \to M$,

$$k(T \circ S) \leqslant k(T)k(S)$$

and, in particular,

$$k(T^n) \leqslant k^n(T), \qquad n = 1, 2, \ldots.$$

If M is a linear space whose metric is generated by a norm, $k(T+S) \leqslant k(T) + k(S)$ and, for $\alpha \geqslant 0$, $k(\alpha T) = \alpha k(T)$.

A mapping $T: M \to M$ is said to be a *contraction* if $k(T) < 1$; more precisely, T is a k-contraction with respect to ρ if $k_\rho(T) \leqslant k < 1$.

Theorem 2.1 (Banach's Contraction Principle) *Let* (M, ρ) *be a complete metric space and let* $T: M \to M$ *be a contraction. Then* T *has a unique fixed point in* M*, and for each* $x_0 \in M$ *the sequence of iterates* $\{T^n x_0\}$ *converges to this fixed point.*

We give three proofs of Theorem 2.1. The first is nonconstructive and establishes only the existence part of the theorem while the second, which is a variant of the original proof, not only provides the existence of a fixed point but, as in the original proof, also provides a method for its approximation. We then give the original (and commonly known) proof.

Proof 1 Let $a=\inf\{\rho(x,Tx): x\in M\}$ and $k=k_\rho(T)$. To see that $a=0$, let $\epsilon>0$ and select $x\in M$ so that $\rho(x,Tx)\leqslant a+\epsilon$. Then

$$a\leqslant \rho(Tx,T^2x)\leqslant k\rho(x,Tx)\leqslant k(a+\epsilon).$$

Since $k<1$ and ϵ can be taken arbitrarily small, $a=0$.

Now for any $\epsilon>0$ the set

$$M_\epsilon=\{x\in M: \rho(x,Tx)\leqslant\epsilon\}$$

is nonempty and closed. Moreover, for any $x,y\in M_\epsilon$,

$$\rho(x,y)\leqslant\rho(x,Tx)+\rho(Tx,Ty)+\rho(Ty,y)\leqslant 2\epsilon+k\rho(x,y),$$

yielding

$$\rho(x,y)\leqslant\frac{2\epsilon}{1-k}$$

from which $\lim_{\epsilon\to 0}\operatorname{diam} M_\epsilon=0$.

Since the family $\{M_\epsilon\}$ descends as $\epsilon\downarrow 0$, the Cantor Intersection Theorem implies $\bigcap_{\epsilon>0}M_\epsilon$ consists of exactly one point, say x, which must be fixed under T $(x=Tx)$.

Proof 2 Set $\varphi(x)=(1-k)^{-1}\rho(x,Tx)$ for $x\in M$ (where $k=k_\rho(T)$). Then

$$\rho(x,Tx)-k\rho(x,Tx)\leqslant\rho(x,Tx)-\rho(Tx,T^2x);$$

hence

$$\rho(x,Tx)\leqslant\varphi(x)-\varphi(Tx),\qquad x\in M. \tag{2.2}$$

Thus for $x_0\in M$ and $n,m\in\mathbb{N}$ with $n<m$,

$$\rho(T^nx_0,T^{m+1}x_0)\leqslant\sum_{i=n}^{m}\rho(T^ix_0,T^{i+1}x_0)\leqslant\varphi(T^nx_0)-\varphi(T^{m+1}x_0). \tag{2.3}$$

In particular, $\Sigma_{i=0}^{\infty}\rho(T^ix_0,T^{i+1}x_0)<+\infty$. Therefore $\{T^nx_0\}$ is a Cauchy sequence and, since T is continuous, it converges to a fixed point x of T. The rate of this convergence may be obtained from (2.3) by letting $m\to\infty$:

$$\rho(T^nx_0,x)\leqslant\varphi(T^nx_0)=(1-k)^{-1}\rho(T^nx_0,T^{n+1}x_0)\leqslant\frac{k^n}{1-k}\rho(x_0,Tx_0).$$

Remark 2.1 The above proof shows that *any* continuous mapping which satisfies (2.2) for arbitrary $\varphi: M \to \mathbb{R}^+$ must have a fixed point. In fact, it can be shown by other means that if φ is lower semicontinuous, then an *arbitrary* mapping $T: M \to M$ satisfying (2.2) must have a fixed point. This fact, which is generally known as the Caristi Theorem, is presented in detail later. It is equivalent to the Ekeland Minimization Principle (Ekeland, 1974) (assuming the Axiom of choice) and has many applications in analysis (see, e.g., Brezis and Browder, 1976, for a thorough discussion). The fixed point in both the above cases need not be unique and in the second instance the sequence $\{T^n x_0\}$ need not even converge to a fixed point of T.

Proof 3 Select $x_0 \in M$ and define the iterative sequence $\{x_n\}$ by $x_{n+1} = Tx_n$ (equivalently, $x_n = T^n x_0$), $n = 0, 1, 2, \ldots$. Observe that for any indices $n, p \in \mathbb{N}$,

$$\begin{aligned}\rho(x_n, x_{n+p}) &= \rho(T^n x_0, T^{n+p} x_0) = \rho(T^n x_0, T^n \circ T^p x_0) \leqslant k(T^n)\rho(x_0, T^p x_0)\\ &\leqslant k^n [\rho(x_0, Tx_0) + \rho(Tx_0, T^2 x_0) + \cdots + \rho(T^{p-1} x_0, T^p x_0)]\\ &\leqslant k^n (1 + k + \cdots + k^{p-1}) \rho(x_0, Tx_0)\\ &\leqslant k^n \left(\frac{1-k^p}{1-k}\right) \rho(x_0, Tx_0). \end{aligned} \tag{2.4}$$

This shows that $\{x_n\}$ is a Cauchy sequence, and since M is complete there exists $x \in M$ such that $\lim_{n\to\infty} x_n = x$. To see that x is the unique fixed point of T, observe that

$$x = \lim_{n\to\infty} x_n = \lim_{n\to\infty} x_{n+1} = \lim_{n\to\infty} Tx_n = Tx$$

and, moreover, $x = Tx$ and $y = Ty$ imply

$$\rho(x, y) = \rho(Tx, Ty) \leqslant k\rho(x, y),$$

yielding $\rho(x, y) = 0$.

As in the second proof, letting $p \to \infty$ in (2.4) yields

$$\rho(x_n, x) = \rho(T^n x_0, x) \leqslant \frac{k^n}{1-k} \rho(x_0, Tx_0). \tag{2.5}$$

Remark 2.2 An analysis of the third proof reveals that the assumption $k(T) < 1$ is stronger than necessary. It suffices to assume $k(T^n) < 1$ for at least one fixed $n \in \mathbb{N}$. This implies T^n is a contraction and (by Theorem 2.1) has a unique fixed point x. But $Tx = T^{n+1} x = T^n \circ Tx$, so Tx is also a fixed point of T^n. Hence $x = Tx$, proving x is also a fixed point of T (and the unique one). It is not difficult to find examples of mappings T (even on the interval $[0, 1]$) which are continuous (or discontinuous) and for which $k(T^n) < 1$ while $k(T) \geqslant 1$. However, these examples seem exceptional, so we shall confine our attention to more typical situations.

We now expand on the idea of Remark 2.2. Let $T: M \to M$ be lipschitzian, fix $x_0 \in M$, and let $x_n = T^n x_0$. The counterpart of the estimate (1.2) for this more general class of mappings is

$$\rho(x_n, x_{n+p}) \leqslant \sum_{i=n}^{n+p-1} \rho(T^i x_0, T^{i+1} x_0)$$
$$\leqslant \left[\sum_{i=0}^{p} k(T^{n+i}) \right] \rho(x_0, Tx_0).$$

Thus $\{x_n\}$ is a Cauchy sequence if it is the case that

$$\sum_{i=1}^{\infty} k(T^i) < +\infty. \tag{2.6}$$

Using the fact that $k(T^n)$ is multiplicative, i.e., since $k(T^{n+m}) \leqslant k(T^n)k(T^m)$, it is easy to see that there exists a number $k_\infty(T)$ which satisfies:

$$k_\infty(T) = \lim_{n \to \infty} [k(T^n)]^{1/n} = \inf\{[k(T^n)]^{1/n}: n = 1, 2, \ldots\}. \tag{2.7}$$

Thus (2.6) holds if and only if $k_\infty(T) < 1$, so the assumption $k(T) < 1$ in Theorem 2.1 can be replaced with $k_\infty(T) < 1$.

Of course the question remains of whether the weaker assumption $k_\infty(T) < 1$ actually provides a stronger version of Theorem 2.1. To respond to this, we introduce the notion of equivalence between metrics: Two metrics ρ and r defined on a given set M are said to be equivalent if there exist two positive constants a and b such that for all $x, y \in M$,

$$ar(x, y) \leqslant \rho(x, y) \leqslant br(x, y). \tag{2.8}$$

For two such metrics, any sequence which is Cauchy with respect to r is also Cauchy with respect to ρ (and conversely). Consequently, (M, ρ) is complete if and only if (M, r) is also.

For a ρ-lipschitzian mapping $T: M \to M$, (2.8) implies

$$r(Tx, Ty) \leqslant (1/a)\rho(Tx, Ty) \leqslant (1/a)k_\rho(T)\rho(x, y) \leqslant (b/a)k_\rho(T)r(x, y),$$

and thus $k_r(T) \leqslant (b/a)k_\rho(T)$. Similarly, $k_\rho(T) \leqslant (b/a)k_r(T)$, and so for any $n \in \mathbb{N}$,

$$\frac{a}{b} k_\rho(T^n) \leqslant k_r(T^n) \leqslant \frac{b}{a} k_\rho(T^n).$$