FUNCTIONAL ANALYSIS

PIOTR HAJŁASZ

1. Banach and Hilbert spaces

In what follows $\mathbb{K}$ will denote $\mathbb{R}$ of $\mathbb{C}$.

Definition. A *normed* space is a pair $(X, \|\cdot\|)$, where X is a linear space over $\mathbb{K}$ and

$$\|\cdot\| : X \to [0,\infty)$$

is a function, called a *norm*, such that

(1) $\|x+y\| \leq \|x\| + \|y\|$ for all $x, y \in X$;
(2) $\|\alpha x\| = |\alpha|\|x\|$ for all $x \in X$ and $\alpha \in \mathbb{K}$;
(3) $\|x\| = 0$ if and only if $x = 0$.

Since $\|x-y\| \leq \|x-z\| + \|z-y\|$ for all $x, y, z \in X$,

$$d(x,y) = \|x-y\|$$

defines a metric in a normed space. In what follows normed paces will always be regarded as metric spaces with respect to the metric d. A normed space is called a *Banach space* if it is complete with respect to the metric d.

Definition. Let X be a linear space over $\mathbb{K}$ (=$\mathbb{R}$ or $\mathbb{C}$). The *inner product* (*scalar product*) is a function

$$\langle\cdot,\cdot\rangle : X \times X \to \mathbb{K}$$

such that

(1) $\langle x, x\rangle \geq 0$;
(2) $\langle x, x\rangle = 0$ if and only if $x = 0$;
(3) $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle$;
(4) $\langle x_1 + x_2, y\rangle = \langle x_1, y\rangle + \langle x_2, y\rangle$;
(5) $\langle x, y\rangle = \overline{\langle y, x\rangle}$,

for all $x, x_1, x_2, y \in X$ and all $\alpha \in \mathbb{K}$.

As an obvious corollary we obtain

$$\langle x, y_1 + y_2\rangle = \langle x, y_1\rangle + \langle x, y_2\rangle, \quad \langle x, \alpha y\rangle = \overline{\alpha}\langle x, y\rangle\,,$$

Date: February 12, 2009.

for all $x, y_1, y_2 \in X$ and $\alpha \in \mathbb{K}$.

For a space with an inner product we define

$$\|x\| = \sqrt{\langle x, x\rangle}\,.$$

Lemma 1.1 (Schwarz inequality). *If X is a space with an inner product $\langle\cdot,\cdot\rangle$, then*

$$|\langle x, y\rangle| \leq \|x\|\,\|y\| \quad \textit{for all } x, y \in X\,.$$

Proof. We can assume that $\langle x, y\rangle \neq 0$. For $x, y \in X$ and $t \in \mathbb{R}$ we have

$$\begin{aligned} 0 &\leq \langle x + ty, x + ty\rangle = \langle x, x\rangle + t[\langle x, y\rangle + \langle y, x\rangle] + t^2\langle y, y\rangle \\ &= \langle x, x\rangle + 2t\,\mathrm{re}\,\langle x, y\rangle + t^2\langle y, y\rangle\,. \end{aligned}$$

We obtained a quadratic function of a variable t which is nonnegative and hence

$$0 \geq \Delta = 4(\mathrm{re}\,\langle x, y\rangle)^2 - 4\langle x, x\rangle\langle y, y\rangle\,,$$

$$(\mathrm{re}\,\langle x, y\rangle)^2 \leq \langle x, x\rangle\langle y, y\rangle\,.$$

If $|\alpha| = 1$, then replacing y by αy we obtain

$$\langle x, x\rangle\langle y, y\rangle = \langle x, x\rangle\langle \alpha y, \alpha y\rangle \geq (\mathrm{re}\,\langle x, \alpha y\rangle)^2 = (\mathrm{re}\,(\overline{\alpha}\langle x, y\rangle))^2\,.$$

In particular for $\alpha = \langle x, y\rangle/|\langle x, y\rangle|$ we have

$$\langle x, x\rangle\langle y, y\rangle \geq \left(\mathrm{re}\left(\frac{\overline{\langle x, y\rangle}}{|\langle x, y\rangle|}\langle x, y\rangle\right)\right)^2 = |\langle x, y\rangle|^2\,.$$

This completes the proof. □

Corollary 1.2. *If X is a space with an inner product $\langle\cdot,\cdot\rangle$, then*

$$\|x\| = \sqrt{\langle x, y\rangle}$$

is a norm.

Proof. The properties $\|\alpha x\| = |\alpha|\|x\|$ and $\|x\| = 0$ if and only if $x = 0$ are obvious. To prove the last property we need to apply the Schwarz inequality.

$$\begin{aligned} \|x + y\|^2 &= \langle x + y, x + y\rangle = \langle x, x\rangle + 2\,\mathrm{re}\,\langle x, y\rangle + \langle y, y\rangle \\ &\leq \|x\|^2 + 2\|x\|\,\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2\,. \end{aligned}$$

Definition. A space with an inner product $\langle\cdot,\cdot\rangle$ is called a *Hilbert space* if it is a Banach space with respect to the norm

$$\|x\| = \sqrt{\langle x, x\rangle}\,.$$

Proposition 1.3 (The Polarization Identity). *Let $\langle\cdot,\cdot\rangle$ be an inner product in X.*

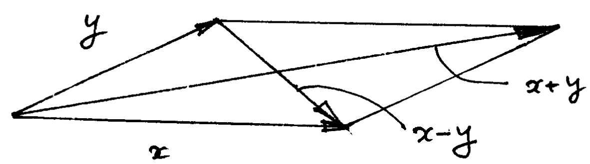
(1) *If* $\mathbb{K} = \mathbb{R}$, *then*
$$\langle x, y\rangle = \frac{1}{4}\left(\|x+y\|^2 - \|x-y\|^2\right),$$
for all $x, y \in X$.

(2) *If* $\mathbb{K} = \mathbb{C}$, *then*
$$\langle x, y\rangle = \frac{1}{4}\left(\|x+y\|^2 - \|x-y\|^2\right) - \frac{1}{4}\, i\left(\|ix+y\|^2 - \|ix-y\|^2\right)$$
for all $x, y \in X$.

Proof is left as an easy exercise.

Theorem 1.4. *Let* $(X, \|\cdot\|)$ *be a normed space over* $\mathbb{K}$ *(*$= \mathbb{R}$ *or* $\mathbb{C}$*). Then there is an inner product* $\langle\cdot,\cdot\rangle$ *such that*
$$\|x\| = \sqrt{\langle x, x\rangle}$$
if and only if the norm satisfies the Parallelogram Law, *i.e.*
$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2 \quad \text{for all } x, y \in X \ .$$

The Parallelogram Law has a nice geometric interpretation.



Proof. The implication $\Rightarrow$ follows from a direct computation. To prove the other implication $\Leftarrow$ we define the inner product using the Polarization Identities and we check that it has all the required properties. We leave the details as an exercise. □

1.1. **Examples. 1.** $\mathbb{C}^n$ with respect to each of the following norms is a Banach space
$$\|x\|_1 = \sum_{i=1}^{n} |x_i|,$$
$$\|x\|_\infty = \max_{i=1,2,\ldots,n} |x_i|,$$
$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}, \quad 1 \le p < \infty,$$
where $x = (x_1, \ldots, x_n) \in \mathbb{C}^n$. The Banach space $(\mathbb{C}^n, \|\cdot\|_p)$ is denoted by ℓ^p_n.

2. $\mathbb{C}^n$ with the inner product

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i \overline{y_i}$$

is a Hilbert space. Note that the corresponding norm is

$$\sqrt{\langle x, x\rangle} = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2} = \|x\|_2$$

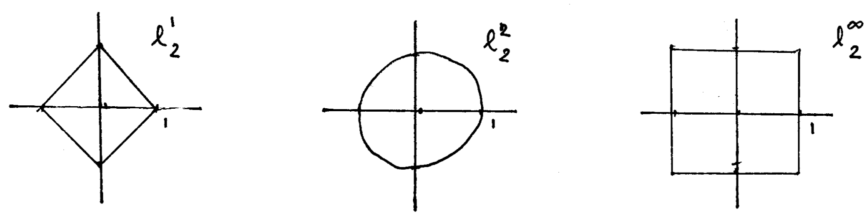
so ℓ^2_n is a Hilbert space. The metric associated with the norm is

$$d(x, y) = \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2}\,,$$

i.e. it is the Euclidean metric.

3. The spaces ℓ^p_n were defined over $\mathbb{K} = \mathbb{C}$, but we can also do the same construction for $\mathbb{K} = \mathbb{R}$ by replacing $\mathbb{C}^n$ by $\mathbb{R}^n$. The resulting space is also denoted by ℓ^p_n, but in each situation it will be clear whether we talk about the real or complex space ℓ^p_n so there is no danger of a confusion.

4. Consider ℓ^p_2 over $\mathbb{K} = \mathbb{R}$. Then the shape of the unit ball $\{x : \|x\| \leq 1\}$ is



5. A subset of the Euclidean space $\mathbb{R}^n$ is called an *ellipsoid* if it is the image of the unit ball in $\mathbb{R}^n$ under a nondegenerate linear mapping $L : \mathbb{R}^n \to \mathbb{R}^n$ (i.e. $\det L \neq 0$).

For every ellipsoid E in $\mathbb{R}^n$ there is an inner product in $\mathbb{R}^n$ such that E is the unit ball in the associated norm. Indeed, if $L : \mathbb{R}^n \to \mathbb{R}^n$ is an isomorphism such that $E = L(B^n(0, 1))$, then it suffices to define the inner product as

$$\langle x, y\rangle = (L^{-1}x) \cdot (L^{-1}y)$$

where $\cdot$ stands for the standard inner product in $\mathbb{R}^n$.

More precisely, if $L : \mathbb{R}^n \to \mathbb{R}^n$ is non-degenerate, then according to the polar decomposition theorem (P. Lax, *Linear Algebra*, p. 139)

$$L = RU$$

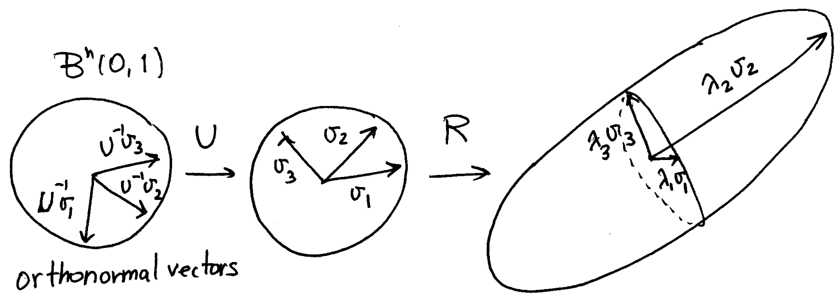
where U is unitary and R is positive self-adjoint. The mapping R can be computed explicitly

$$LL^T = RUU^TR^T = R^2, \quad R = \sqrt{LL^T}\,.$$

According to the spectral theorem there is an orthonormal basis $v_1, \dots v_n$ in $\mathbb{R}^n$ (with respect to the standard inner product) such that

$$R = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}$$

in this basis. That means the mapping L has the following structure. First we rotate (the mapping U) and then we apply R which has a simple geometric meaning of extending the length of vectors $v_1, \dots, v_n$ by factors $\lambda_1, \dots, \lambda_n$. Now if $B^n(0,1)$ is the unit ball, then $U(B^n(0,1))$ is also the unit ball, so $L(B^n(0,1)) = R(B^n(0,1))$ is an ellipsoid with semi-axes $\lambda_1 v_1, \dots, \lambda_n v_n$ of the lengths being eigenvalues of $R = \sqrt{LL^T}$. These numbers are called *singular values* of L.



Now the ellipsoid $E = L(B^n(0,1))$ is the unit ball for the inner product

$$\Big\langle \sum_{i=1}^n a_i v_i, \sum_{i=1}^n b_i v_i \Big\rangle = \sum_{i=1}^n \frac{a_i b_i}{\lambda_i^2}\,.$$

We will see later (Corollary 5.12) any real inner product space space H of dimension n is isometrically isomorphic to ℓ^2_n, i.e. $\mathbb{R}^n$ with the standard inner product, so if $L : \ell^2_n \to H$ is this isometric isomorphism, the unit ball in H is $L(B^n(0,1))$, so it is an ellipsoid. Thus we proved.

Theorem 1.5. *A convex set in $\mathbb{R}^n$ is a unit ball for a norm associated with an inner product if and only if it is an ellipsoid.*

6. ℓ^∞, the space of all bounded (complex, real) sequences $x = (a_n)_{n=1}^\infty$ with the norm

$$\|x\|_\infty = \sup_n |x_n|$$

is a Banach space. This is very easy to check.

7. c_1, the space of all (complex, real) convergent sequences with the norm $\|\cdot\|_\infty$ is a Banach space.

8. c_0, the space of all (complex, real) sequences that converge to zero with the norm $\|\cdot\|_\infty$ is a Banach space.

9. Note that $c_0 \subset c \subset \ell^\infty$ and both c_0 and c are closed linear subspaces of ℓ^∞ with respect to the metric generated by the norm.

EXERCISE. *Prove that ℓ^∞, c and c_0 are Banach spaces.*

EXERCISE. *Prove that the spaces c and c_0 are separable, while ℓ^∞ is not.*

10. ℓ^p, $1 \leq p < \infty$ is the space of all (complex, real) sequences $x = (x_n)_{n=1}^\infty$ such that

$$\|x\|_p = \left(\sum_{n=1}^\infty |x_n|^p\right)^{1/p} .$$

It follows from the Minkowski inequality for sequences that $\|\cdot\|_p$ is a norm and that ℓ^p is a linear space. We will prove now that ℓ^p is a Banach space, i.e. that it is complete. Let $x_n = (a_i^n)_{i=1}^\infty$ be a Cauchy sequence in ℓ^p, i.e. for every $\varepsilon > 0$ there is N such that for all $n, m > N$

$$\|x_n - x_m\|_p = \left(\sum_{i=1}^\infty |a_i^n - a_i^m|^p\right)^{1/p} < \varepsilon .$$

Hence for each i the sequence $(a_i^n)_{n=1}^\infty$ is a Cauchy sequence in $\mathbb{K}$ ($= \mathbb{C}$ or $\mathbb{R}$). Let $a_i = \lim_{n\to\infty} a_i^n$. Fix an integer k. Then for $n, m > N$ we have

$$\sum_{i=1}^k |a_i^n - a_i^m|^p < \varepsilon^p$$

and passing to the limit as $m \to \infty$ yields

$$\sum_{i=1}^k |a_i^n - a_i|^p \leq \varepsilon^p .$$

Now taking the limit as $k \to \infty$ we obtain

$$\sum_{i=1}^\infty |a_i^n - a_i|^p \leq \varepsilon^p ,$$

i.e.

$$\|x_n - x_m\|_p \leq \varepsilon \quad \text{where } x = (a_i)_{i=1}^\infty .$$

This proves that $x \in \ell^p$ and $x_n \to x$ in ℓ^p. The proof is complete.

In particular the space ℓ^2 is a Hilbert space because its norm is associated with the inner product

$$\langle x, y\rangle = \sum_{i=1}^\infty x_i \overline{y_i} .$$

EXERCISE. *Prove that ℓ^p, $1 \le p < \infty$ is separable.*

11. We will prove that ℓ^p for $p \neq 2$ is not an inner product space. Let $x = (1, 0, 0, \ldots)$, $y = (0, 1, 0, 0, \ldots)$. If $1 \le p < \infty$, then

$$\|x + y\|_p^2 + \|x - y\|_p^2 = 2^{1+2/p}\,, \quad 2\|x\|_p^2 + 2\|y\|_p^2 = 4\,,$$

and thence the Parallelogram Law is violated. The same example can also be used in the case $p = \infty$. In the real case this result can also be seen as a consequence of the fact that the two dimensional section of the unit ball in ℓ^p, $p \neq 2$, along the space generated by the first two coordinates is not an ellipse.

12. If X is equipped with a positive measure μ, then for $1 \le p < \infty$, $L^p(\mu)$ is a Banach space with respect to the norm

$$\|f\|_p = \left(\int_X |f|^p \, d\mu \right)^{1/p} .$$

Also $L^\infty(\mu)$ is a Banach space with the norm being the essential supremum of $|f|$. For the proofs see the notes from Analysis I.

For $p = 2$ the space $L^2(\mu)$ is a Hilbert space with respect to the inner product

$$\langle f, g \rangle = \int_X f \overline{g} \, d\mu\,.$$

13. If $X = \{1, 2, 3, \ldots\}$ and μ is the counting measure (i.e. $\mu(A) = \#A$), then $L^p(\mu) = \ell^p$. In particular this gives another proof that ℓ^p is a Banach space. However the proof given above is much more elementary.

14. If X is a compact metric space, then the space of continuous functions on X with respect to the norm

$$\|f\| = \sup_{x \in X} |f(x)|$$

is a Banach space. This space is denoted by $C(X)$. The resulting metric in $C(X)$ is the metric of uniform convergence.

15. Let H^2 be the class of all holomorphic functions on the unit disc $D = \{z \in \mathbb{C} : |z| < 1\}$ such that

$$\|f\|_{H^2} = \Big(\sup_{0<r<1} \frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\theta})|^2 \, d\theta \Big)^{1/2} < \infty\,.$$

This space is called *Hardy space* H^2. We will prove now that it is a Hilbert space and we will find an explicit formula for the inner product.

Since every holomorphic function in D can be represented as a Taylor polynomial $f(z)=\sum_{n=0}^\infty a_n z^n$, for $|z|<1$ we have

$$\begin{aligned}\frac{1}{2\pi}\int_0^{2\pi}|f(re^{i\theta})|^2\,d\theta &= \frac{1}{2\pi}\int_0^{2\pi}\Big(\sum_{n=0}^\infty a_n r^n e^{in\theta}\Big)\overline{\Big(\sum_{m=0}^\infty a_m r^m e^{im\theta}\Big)}\,d\theta \\ &= \sum_{n,m=0}^\infty a_n\overline{a_m}r^{n+m}\underbrace{\frac{1}{2\pi}\int_0^{2\pi}e^{i(n-m)\theta}\,d\theta}_{\text{1 if n=m, 0 if } n\neq m} \\ &= \sum_{n=0}^\infty r^{2n}|a_n|^2\,.\end{aligned}$$

Hence

$$\|f\|_{H^2}=\Big(\lim_{r\to 1^-}\sum_{n=0}^\infty r^{2n}|a_n|^2\Big)^{1/2}=\Big(\sum_{n=0}^\infty |a_n|^2\Big)^{1/2}.$$

This proves that the space H^2 is isometrically isomorphic with ℓ^2 and that the norm in H^2 is associated with the inner product

$$\langle f,g\rangle=\sum_{n=0}^\infty a_n\overline{b_n}=\lim_{r\to 1^-}\frac{1}{2\pi}\int_0^{2\pi}f(re^{i\theta})\overline{g(re^{i\theta})}\,d\theta\,,$$

where $f=\sum_{n=0}^\infty a_n z^n$, $g(z)=\sum_{n=0}^\infty b_n z^n$.

2. Linear operators

If X and Y are normed spaces, then linear functions $L:X\to Y$ will be called (linear) *operators* of (linear) *transforms*. Also we will often write Lx instead of $L(x)$. We say that a linear operator $L:X\to Y$ is *bounded* if there is a constant $C>0$ such that

$$\|Lx\|\leq C\|x\|\quad\text{for all } x\in X.$$

Theorem 2.1. *Let $L:X\to Y$ be a linear operator between normed spaces. Then the following conditions are equivalent:*

(1) *I is continuous;*
(2) *L is continuous at 0;*
(3) *L is bounded.*

Proof. The implication (1) $\Rightarrow$ (2) is obvious. (2) $\Rightarrow$ (3) For if not there would exist a sequence $x_n\in X$ such that $\|Lx_n\|>n\|x_n\|$. Then $\|L(x_n/(n\|x_n\|))\|>1$, but on the other hand $\|L(x_n/(n\|x_n\|))\|\to 0$, because $x_n/(n\|x_n\|)\to 0$ which is an obvious contradiction. (3) $\Rightarrow$ (1) Let $x_n\to x$. Then $Lx_n\to Lx$. Indeed,

$$\|Lx-Lx_n\|=\|L(x-x_n)\|\leq C\|x-x_n\|\to 0\,.$$

This completes the proof. □

For a linear operator $L : X \to Y$ we define its norm by

$$\|L\| = \sup_{\|x\|\leq 1} \|Lx\|\,.$$

Then

$$\|Lx\| \leq \|L\|\,\|x\| \quad \text{for all } x \in X.$$

Indeed,

$$\|Lx\| = \left\| L\Big(\|x\|\frac{x}{\|x\|}\Big)\right\| = \|x\|\left\| L\frac{x}{\|x\|}\right\| \leq \|x\|\,\|L\|\,.$$

Thus L is bounded if and only if $\|L\| < \infty$. Moreover $\|L\|$ is the smallest constant C for which the inequality

$$\|Lx\| \leq C\|x\| \quad \text{for all } x \in X$$

is satisfied. $\|L\|$ is called the *operator norm.*

The class of bounded operators $L : X \to Y$ is denoted by $B(X,Y)$. We also write $B(X) = B(X,X)$. Clearly $B(X,Y)$ has a structure of a linear space.

Lemma 2.2. *$B(X,Y)$ equipped with the operator norm is a normed space.*

Proof is very easy and left as an exercise.

Theorem 2.3. *If X is a normed space and Y is a Banach space, then $B(X,Y)$ is a Banach space.*

Proof. Let $\{L_n\}$ be a Cauchy sequence in $B(X,Y)$. Then

$$\|L_n x - L_m x\| \leq \|L_n - L_m\|\,\|x\|\,. \tag{2.1}$$

Since the right hand side converges to 0 as $n,m \to \infty$, we conclude that $\{L_n x\}$ is a Cauchy sequence for every $x \in X$ and hence $\{L_n x\}$ has a limit in Y. We denote it by $Lx = \lim_{n\to\infty} L_n x$. Because $\{L_n\}$ is a Cauchy sequence $\|L_n - L_m\| < \varepsilon$ for all sufficiently large n and m and passing to the limit in (2.1) as $n \to \infty$ yields

$$\|Lx - L_m x\| \leq \varepsilon\|x\| \quad \text{for all sufficiently large } m\,. \tag{2.2}$$

This gives

$$\|Lx\| \leq \|Lx - L_m x\| + \|L_m x\| \leq (\varepsilon + \|L_m\|)\|x\|\,.$$

Hence $L \in B(X,Y)$ and $L_m \to L$ in $B(X,Y)$ because of (2.2). □

Definition. By a (continuous linear) *functional* we mean an arbitrary bounded operator

$$L : X \to \mathbb{K}\,.$$

The space $X^* = B(X,\mathbb{K})$ is called the *dual space* of X. Elements of X^* will usually be denoted by x^* and we will write $\langle x^*, x\rangle$ instead of $x^*(x)$.

Corollary 2.4. *If X is a normed space,then X^* is a Banach space.*

Proof. Indeed, $\mathbb{K}$ is a Banach space and the result follows from Theorem 2.3 $\square$

Definition. We say that the two normed spaces X and Y are *isomorphic* if there is an algebraic isomorphism of linear spaces $L \in B(X, Y)$ (i.e. it is one-to-one and surjection) such that $L^{-1} \in B(Y, X)$. The mapping L is called an *isomorphism* of normed spaces X and Y. If X and Y are Banach spaces we call it isomorphism of Banach spaces X and Y.

We say that X and Y are *isometric* if there is an isomorphism $L \in B(X, Y)$ such that $\|Lx\| = \|x\|$ for all $x \in X$.

The following result immediately follows from the equivalence of continuity and boundedness of an operator.

Proposition 2.5. *$L \in B(X, Y)$ is an isomorphism if it is an algebraic isomorphism of linear spaces and there is $C > 0$ such that*

$$\|Lx\| \geq C\|x\| \quad \textit{for all } x \in X\,.$$

The next result is also very easy and left as an exercise.

Proposition 2.6. *If normed spaces X and Y are isomorphic and X is a Banach space, then Y is also a Banach space.*

Exercise. *Find two homeomorphic metric spaces X and Y such that X is complete, while Y is not.*

Example. We will construct now an example of an algebraic isomorphism of normed spaces $L \in B(X, Y)$ such that $L^{-1} \notin B(Y, X)$. Let X be the space of all real sequences $x = (a_1, a_2, \ldots)$ with only a finite number nonzero components, equipped with the norm $\| \cdot \|_\infty$. Let $L : X \to X$ be defined by

$$L(a_1, a_2, \ldots) = (a_1, \frac{a_2}{2}, \frac{a_3}{3}, \ldots)\,.$$

Clearly, L is a linear isomorphism with the inverse

$$L^{-1}(a_1, a_2, a_3, \ldots) = (a_1, 2a_2, 3a_3, \ldots)\,.$$

It is easy to see that L is continuous (because $\|Lx\|_\infty \leq \|x\|_\infty$), but $L^{-1} : X \to X$ is not. Indeed, if $x_n = (0, \ldots, 0, n^{-1}, 0, \ldots)$ where $a_n = n^{-1}$ and all other a_i's are equal zero, then $\|x_n\|_\infty \to 0$, but $\|L^{-1}(x_n)\|_\infty = 1$.

Theorem 2.7. *Let X be a Banach space. Then isomorphisms form an open subset in $B(X)$.*

Proof. If $A \in B(X)$, $\|A\| < 1$, then $I - A$ is an isomorphism.[1] Indeed, the series of operators

$$\sum_{n=0}^{\infty} A^n = I + A + A^2 + A^3 + \dots \tag{2.3}$$

converge absolutely, because

$$\sum_{n=0}^{\infty} \|A^n\| \leq \sum_{n=0}^{\infty} \|A\|^n < \infty\,.$$

Hence it easily follows that the sequence of partial sums of (2.3) is a Cauchy sequence in $B(X)$. Since $B(X)$ is a Banach space, it converges and thus (2.3) defines a bounded operator in $B(X)$. Now it is easy to check that this operator is an inverse of $I - A$, so $I - A$ is an isomorphism.

If $L \in B(X)$ is an isomorphism and $A \in B(X)$, $\|A\| < \|L^{-1}\|^{-1}$, then

$$L - A = L(I - L^{-1}A)$$

is an isomorphism as a composition of isomorphisms. We proved that a certain ball in $B(X)$ centered at L consists of isomorphisms. □

Example. We will prove that the real spaces ℓ_2^1 and ℓ_2^∞ are isometric. That is quite surprising because both spaces are $\mathbb{R}^2$ equipped with two different norms

$$\|(x,y)\|_1 = |x| + |y|, \quad \|(x,y)\|_\infty = \max\{|x|,|y|\}\,.$$

However the mapping

$$L : \ell_2^1 \to \ell_2^\infty, \quad L(x,y) = (x+y, x-y)$$

is an isometry, because $\max\{|x+y|,|x-y|\} = |x| + |y|$. Note that in both spaces the unit ball is a square and the mapping L maps one square onto another.

Exercise. Find all isometries between ℓ_2^1 and ℓ_2^∞.

Exercise. *Prove that the spaces ℓ_3^1 and ℓ_3^∞ are not isometric.*

Proposition 2.8. *The spaces c_0 and c are isomorphic.*

Proof. The mapping $T : c_0 \to c$,

$$L(a_1, a_2, a_3, \dots) = (a_1 + a_2, a_1 + a_3, a_1 + a_4, \dots)$$

is an algebraic isomorphism with the inverse

$$L^{-1}(a_1, a_2, a_3, \dots) = (a_0, a_1 - a_0, a_2 - a_0, \dots)\,,$$

where $a_0 = \lim_{n\to\infty} a_n$. Since $\|Lx\|_\infty \leq 2\|x\|_\infty$ for $x \in c_0$ and $\|L^{-1}x\|_\infty \leq 2\|x\|_\infty$ for $x \in c$ we conclude that L and L^{-1} are continuous, so L is an isomorphism of Banach spaces c_0 and c. □

[1] I stands for identity.

Exercise. *Prove that the spaces c_0 and c are not isometric.*

Definition. We say that two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on a linear space X are equivalent if there are constants $C_1, C_2 > 0$ such that

$$C_1\|x\|_1 \leq \|x\|_2 \leq C_2\|x\|_1 \quad \text{for all } x \in X\,.$$

Proposition 2.9. *The two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on a linear space X are equivalent if and only if the identity mapping $\mathrm{id} : (X, \|\cdot\|_1) \to (X, \|\cdot\|_2)$, $\mathrm{id}(x) = x$ is an isomorphism.*

Proof. It easily follows from the fact that continuity of a linear mapping is equivalent with its boundedness. □

2.1. **Examples of dual spaces.**

Theorem 2.10. *If $s = (s_i) \in \ell^1$, then*

$$\langle x^*, x\rangle = \sum_{i=1}^{\infty} \quad \textit{for } x = (x_i) \in c_0 \tag{2.4}$$

defines a bounded functional $x^* \in c_0^*$ with $\|x^*\| = \|s\|_1$. On the other hand if $x^* \in c_0^*$, then there is unique $x \in \ell^1$ such that x^* can be represented by (2.4). This proves that the space c_0^* is isometrically isomorphic to ℓ^1.

Proof. Let $s \in \ell^1$. Then x^* defined by (2.4) is a bounded linear functional on c_0. Indeed,

$$|\langle x^*, x\rangle| \leq \sup_i |x_i| \sum_{i=1}^{\infty} |s_i| = \|s\|_1\,\|x\|_\infty$$

proves that

$$\|x^*\| \leq \|s\|_1\,. \tag{2.5}$$

Now let $x^* \in c_0^*$. We will prove that there is $s \in \ell^1$ such that x^* satisfies (2.4). It is clear that two different elements of ℓ^1 define different functionals on c_0, so uniqueness is obvious. We will also prove that

$$\|s\|_1 \leq \|x^*\| \tag{2.6}$$

which together with (2.5) will give the equality $\|x^*\| = \|s\|_1$. This will complete the proof.

Let $e_i = (0, \ldots, 0, 1, 0, \ldots) \in c_0$ with 1 on ith coordinate and let $s_i = \langle x^*, e_i\rangle$. Define $z_i = \overline{s_i}/|s_i|$ if $s_i \neq 0$ and $z_i = 0$ if $s_i = 0$. Then

$$z^k = (z_1, \ldots, z_k, 0, 0, \ldots) = \sum_{i=1}^{k} z_i e_i \in c_0, \quad \|z^k\|_\infty \leq 1\,.$$

Hence

$$\|x^*\| \geq \langle x^*, z^k\rangle = \sum_{i=1}^{k} z_i \underbrace{\langle x^*, e_i\rangle}_{s_i} = \sum_{i=1}^{k} |s_i| \,.$$

Letting $k \to \infty$ we have

$$\|s\|_1 = \sum_{i=1}^{\infty} |s_i| \leq \|x^*\|$$

which proves that $s \in \ell^1$ along with the estimate (2.6). Now it easily follows that x^* satisfies (2.4). Indeed, if $x = (x_i) \in c_0$ and

$$x^k = (x_1, \ldots, x_k, 0, 0, \ldots) = \sum_{i=1}^{k} x_i e_i \,,$$

then $x^k \to x$ in c_0 and $\sum_{i=1}^{k} x_i s_i \to \sum_{i=1}^{\infty} x_i s_i$, because $s \in \ell^1$. Hence passing to the limit in the equality

$$\langle x^*, x^k\rangle = \sum_{i=1}^{k} x_i s_i$$

yields (2.4). The proof is complete. $\square$

Exercise. *Prove that the dual space c^* is isometrically isomorphic to ℓ^1.*

Theorem 2.11. *If $s = (s_i) \in \ell^\infty$, then*

$$\langle x^*, x\rangle = \sum_{i=1}^{\infty} s_i x_i \quad \textit{for } x = (x_i) \in \ell^1 \tag{2.7}$$

defines a bounded functional $x^* \in (\ell^1)^*$ with $\|x^*\| = \|s\|_\infty$. On the other hand if $x^* \in (\ell^1)^*$, then there is unique $s \in \ell^\infty$ such that x^* can be represented by (2.7). This proves that the space $(\ell^1)^*$ is isometrically isomorphic to ℓ^∞.

Proof. The proof is pretty similar to the previous one, so we will be short. If $s \in \ell^\infty$, then it is easily seen that x^* given by (2.4) defines a functional on ℓ^1 with $\|x^*\| \leq \|s\|_\infty$. Now let $x^* \in (\ell^1)^*$. It remains to prove that there is $s \in \ell^\infty$ such that x^* satisfied (2.7) and $\|s\|_\infty \leq \|x^*\|$ (uniqueness is obvious). Let $e_i = (0, \ldots, 0, 1, 0, \ldots) \in \ell^1$ and $s_i = \langle x^*, e_i\rangle$. Let $z_i = \overline{s_i}/|s_i|$ if $s_i \neq 0$ and $z_i = 0$ if $s_i = 0$. Put $z^i = z_i e^i \in \ell^1$, so $\|z^i\|_i \leq 1$. Then $\|x^*\| \geq \langle x^*, z^i\rangle = |s_i|$. Now taking supremum over all i yields $\|s\|_\infty \leq \|x^*\|$ and the result easily follows. $\square$

Exercise. *Prove that for every $s = (s_i) \in \ell^1$, $\langle x^*, x\rangle = \sum_{i=1}^{\infty} s_i x_i$ defines a bounded functional on ℓ^∞ with $\|x^*\| = \|s\|_1$.*

Later we will see that not every functional in $(\ell^\infty)^*$ can be represented by an element of ℓ^1, and the above exercise proves only that ℓ^1 is isometrically isomorphic to a closed subspace of $(\ell^\infty)^*$.

Theorem 2.12. *Let $1 < p < \infty$. If $s = (s_i) \in \ell^q$, where $q = p/(p-1)$, then*

$$\langle x^*, x\rangle = \sum_{i=1}^{\infty} s_i x_i \quad \text{for } x = (x_i) \in \ell^p \tag{2.8}$$

defines a bounded functional $x^* \in (\ell^p)^*$ with $\|x^*\| = \|s\|_q$. On the other hand if $x^* = (\ell^p)^*$, then there is unique $s \in \ell^q$ such that x^* can be represented by (2.8). This proves that the space $(\ell^p)^*$ is isometrically isomorphic to ℓ^q.

Proof. Again, the proof is very similar to those presented above, so we will explain only the step where a tiny difference in the argument occurs. Let $x^* \in (\ell^p)^*$. The crucial point is to show that for $s_i = \langle x^*, e_i\rangle$ we have $\|s\|_q \leq \|x^*\|_p$. To prove this we take $z_i = \overline{s_i}|s_i|^{q-2}$ and $z^k = (z_1, \ldots, z_k, 0, 0, \ldots) \in \ell^p$. Then

$$\|x^*\|\,\|z^k\|_p \geq \langle x^*, z^k\rangle = \sum_{i=1}^{k} |s_i|^q\,. \tag{2.9}$$

Since $\|z^k\|_p = (\sum_{i=1}^k |s_i|^q)^{1/p}$ inequality (2.9) yields $\|x^*\| \geq (\sum_{i=1}^k |s_i|^q)^{1/q}$ and the claim follows after passing to the limit as $k \to \infty$. $\square$

The last two results are special cases of the following deep result whose proof based on the Radon-Nikodym theorem is presented in notes from Analysis I and also will be proved in Section 5.5.

Theorem 2.13. *If μ is a σ-finite measure on X and $1 \leq p < \infty$, $1 < q \leq \infty$, $p^{-1} + q^{-1} = 1$, then for every function $g \in L^q(\mu)$,*

$$\Lambda f = \int_X fg\, d\mu \quad \text{for } f \in L^p(\mu) \tag{2.10}$$

defines a bounded functional $\Lambda \in (L^p(\mu))^*$ with

$$\|\Lambda\|_{(L^p(\mu))^*} = \|g\|_{L^q(\mu)}\,.$$

Moreover for every functional $\Lambda \in (L^p(\mu))^*$ there is unique $g \in L^q(\mu)$ such that Λ can be represented by (2.10). This proves that the space $(L^p(\mu))^*$ is isometrically isomorphic to $L^q(\mu)$.

Definition. Let X be a locally compact metric space and $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. We define

(1) $C_c(X)$ is the space of all continuous functions $u : X \to \mathbb{K}$ for which there is a compact set $E \subset X$ such that u vanishes outside E, i.e. $u(x) = 0$ for all $x \in X \setminus E$.
(2) $C_0(X)$ is the space of all continuous functions such that for every $\varepsilon > 0$ there is a compact set $E \subset X$ so that $|u(x)| < \varepsilon$ for all $x \in X \setminus E$.

EXERCISE. *Prove that* $u \in C_0(\mathbb{R}^n)$ *if and only if* u *is continuous and* $\lim_{|x|\to\infty} u(x) = 0$.

Theorem 2.14. *Let* X *be a locally compact metric space. Then* $C_0(X)$ *is a Banach space with the norm*

$$\|u\| = \sup_{x\in X} |u(x)| \, .$$

Moreover $C_c(X)$ *forms a dense subset in* $C_0(X)$.

We leave the proof as an exercise.

If μ is a signed measure, then we have unique Hahn decomposition

$$\mu = \mu^+ - \mu^- \, ,$$

where μ^+ and μ^- are positive measures concentrated on disjoint sets. We define the measure $|\mu|$ as

$$|\mu| = \mu^+ + \mu^- \, .$$

The number $|\mu|(X)$ is called *total variation* of μ.

Example. If $\mu(E) = \int_E f\, d\mu$ where $f \in L^1(\mu)$, then $|\mu|(E) = \int_E |f|\, d\mu$.

Theorem 2.15 (Riesz representation theorem)**.** *If* X *is a locally compact metric space and* $\Phi \in (C_0(X))^*$, *then there is unique Borel signed measure* μ *of finite total variation such that*

$$\Phi(f) = \int_X f\, d\mu \, .$$

Moreover $\|\Phi\| = |\mu|(X)$.

Thus the dual space of $C_0(X)$ is isometrically isomorphic to the space of signed measures of finite total variation.

3. Finitely dimensional spaces.

Theorem 3.1. *In a finitely dimensional linear space any two norms are equivalent. In particular every finitely dimensional normed space is a Banach space.*

Proof. Let $e_1, e_2, \ldots, e_n$ be a Hamel basis of the linear space X. For every $x \in X$ we define a norm

$$\|x\|' = \max_{i=1,2,\ldots,n} |x_i| \quad \text{where } x = \sum_{i=1}^n x_i e_i \, .$$

Note that the space X with respect to the norm $\|\cdot\|'$ is locally compact and complete. It suffices to prove that every norm in X is equivalent with $\|\cdot\|'$. Let $\|\cdot\|$ be an arbitrary norm in X. Then

$$\|x\| = \left\|\sum_{i=1}^n x_i e_i\right\| \leq \sum_{i=1}^n |x_i|\|e_i\| \leq \|x\|' \sum_{i=1}^n \|e_i\| = C\|x\|'\,.$$

Now it remains to prove the opposite inequality $\|x\|' \leq C\|x\|$. By contradiction suppose that for every n we can find $x_n \in X$ such that $\|x_n\|' \geq n\|x_n\|$. Let $y_n = x_n/\|x_n\|'$. Then $\|y_n\|' = 1$ and $\|y_n\| \leq 1/n \to 0$, so $y_n \to 0$ in the norm $\|\cdot\|$. Since the space $(X, \|\cdot\|')$ is locally compact, there is a subsequence $y_{n_i} \to y$ convergent with respect to the norm $\|\cdot\|'$. Since the convergence in the norm $\|\cdot\|'$ implies the convergence in the norm $\|\cdot\|$ (because $\|x\| \leq C\|x\|'$) we conclude that $y = 0$ which is a contradiction, because $\|y\|' = \lim_{i\to\infty} \|y_{n_i}\|' = 1$. The last part of the theorem is easy, because the space $(X, \|\cdot\|')$ is complete. $\Box$

As an application we will prove

Proposition 3.2. *For every polynomial P on $\mathbb{R}^n$ there is a constant $C > 0$ such that on every ball $B(x,r)$ in $\mathbb{R}^n$ we have*

$$\sup_{B(x,r)} |P| \leq C|B(x,r)|^{-1} \int_{B(x,r)} |P|\,. \tag{3.1}$$

Remark. The opposite inequality

$$|B(x,r)|^{-1} \int_{B(x,r)} |f| \leq \sup_{B(x,r)} |f|$$

is true for any locally integrable f. It is also clear that for every ball there is a constant C such that (3.1) is satisfied, however, we want to find a constant C that will be good for all the balls at the same time.

Proof. Let k be the degree of P. The space $\mathcal{P}^k$ of polynomials of degree less than or equal k has finite dimension, so all norms in that space are equivalent. In particular there is a constant $C > 0$ such that

$$\sup_{B(0,1)} |Q| \leq C|B(0,1)|^{-1} \int_{B(0,1)} |Q| \quad \text{for all } Q \in \mathcal{P}^k. \tag{3.2}$$

Now it remains to observe that a linear change of variables allows us to rewrite (3.1) in an equivalent form of (3.2) without changing the constant C, where Q is obtained from P by a linear change of variables. Since degree of Q equals k and the claim follows. $\Box$

While Theorem 3.1 proves that any two norms in a finitely dimensional vector space are comparable the theorem does not give an explicit estimate for the comparison. However the following result does.

Theorem 3.3 (F. John)**.** *Let $(X, \|\cdot\|)$ be an n-dimensional real normed space. Then there is a Hilbert norm $\|\cdot\|'$ in X such that*

$$\|x\| \leq \|x\|' \leq \sqrt{n}\, \|x\| \quad \textit{for all } x \in X.$$

Proof. The unit ball with respect to the norm $\|\cdot\|$ is a convex symmetric body K (symmetric means that $x \in K \Rightarrow -x \in K$) and any ellipsoid E centered at 0 is a unit balls for a Hilbert norm $\|\cdot\|'$ (we choose the inner product so that the semi-axes are orthonrmal, see Section 1.1, Example 5). Now it suffices to prove that there is an ellipsoid E such that $E \subset K \subset \sqrt{n}\, E$. Indeed, $E \subset K$ means that $\|x\|' \leq 1 \Rightarrow \|x\| \leq 1$ which easily implies $\|x\| \leq \|x\|'$ for any $x \in X$ and $K \subset \sqrt{n}\, E$ means that $\|x\| \leq 1 \Rightarrow \|x\|' \leq \sqrt{n}$ which easily implies $\|x\|' \leq \sqrt{n}\, \|x\|$ for any $x \in X$. Therefore we are left with the proof of the following result.

Theorem 3.4 (F. John ellipsoid theorem)**.** *If K is a closed convex symmetric body in an n-dimensional real vector space X, then there is a closed ellipsoid E centered at* 0 *such that*

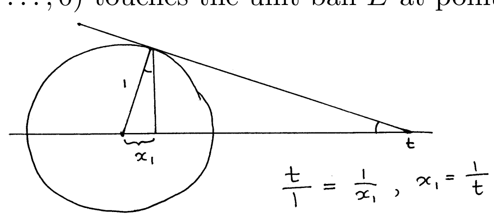
$$E \subset K \subset \sqrt{n} E\, .$$

Proof. Choose any Euclidean coordinate system in X and define E to be an ellipsoid centered at 0 contained in K that has maximal volume. A simple compactness argument shows that such an ellipsoid exists. Now we choose a new Euclidean coordinate system in which E becomes the unit ball (this system is obtained by rescaling semi-axes). Note that also in this new system E is an ellipsoid of maximal volume in K. It remains to prove that $K \subset \sqrt{n}\, E$, i.e. K is contained in the ball of radius $\sqrt{n}$.

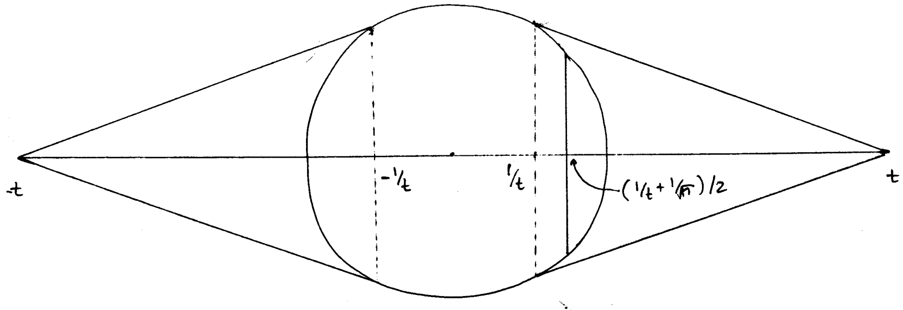
By contradiction suppose that there is an element in K whose distance to 0 is greater than $\sqrt{n}$. By rotating the coordinate system we can assume that

$$(t, 0, \ldots, 0) \in K, \quad t > \sqrt{n}\, .$$

An elementary geometric argument shows that the tangent cone to E with the vertex at $(t, 0, \ldots, 0)$ touches the unit ball E at points where $x_1 = 1/t$.



Since K is convex and symmetric it follows that the set W consisting of the ball E and the two tangent cones centered at $(\pm t, 0, \ldots, 0)$ is contained in K.



The ball E is described by the equation

$$x_1^2 + y^2 \leq 1, \quad y = (x_2^2 + \ldots + x_n^2)^{1/2}.$$

Now consider the family of ellipsoids E_ε for $0 < \varepsilon < 1$,

$$(1-\varepsilon)^{n-1}x_1^2 + y^2/(1-\varepsilon) \leq 1.$$

Since the x_1 semi-axis of E_ε has length $(1-\varepsilon)^{-(n-1)/2}$ and the length of each of the semi-axes $x_2, \ldots, x_n$ is $(1-\varepsilon)^{1/2}$, the ellipsoids E_ε have the same volume as E. It suffices to prove that for ε sufficiently small E_ε is contained in the interior of W. Indeed, it is then contained in the interior of K and hence even if we enlarge it slightly it is still contained in K which will contradict maximality of the volume of E.

Since

$$(1-\varepsilon)^{n-1} = 1-(n-1)\varepsilon + O(\varepsilon^2), \quad \frac{1}{1-\varepsilon} = 1+\varepsilon+\varepsilon^2 = \ldots = 1+\varepsilon+O(\varepsilon^2),$$

points of the ellipsoid E_ε satisfy

$$x_1^2 + y^2 - \varepsilon((n-1)x_1^2 - y^2) + O(\varepsilon^2) \leq 1$$

which is equivalent to

$$(x_1^2 + y^2 - 1)(1+\varepsilon) \leq \varepsilon(nx_1^2 - 1) + O(\varepsilon^2). \tag{3.3}$$

We consider the two parts of the ellipsoid E_ε: $|x_1| < (1/t + 1/\sqrt{n})/2$ and $|x_1| \geq (1/t + 1/\sqrt{n})/2$ separately. In the first case $|x_1| < (1/t+1/\sqrt{n})/2$ we have

$$nx_1^2 - 1 < n\Big(\frac{1}{4t^2} + \frac{1}{2t\sqrt{n}} + \frac{1}{4n}\Big) - 1 < n\Big(\frac{1}{4n} + \frac{1}{2n} + \frac{1}{4n}\Big) - 1 = 0,$$

where in the last inequality we used $t > \sqrt{n}$. Hence for ε sufficiently small the right hand side of (3.3) is negative and thus

$$x_1^2 + y^2 < 1,$$

i.e. points of E_ε satisfying $|x_1| < (1/t + 1/\sqrt{n})/2$ belong to the interior of E and hence interior of W. Regarding the other part of the ellipsoid E_ε with $|x_1| \geq (1/t + 1/\sqrt{n})/2$ note that the part of the ball E with $|x_1| \geq (1/t + 1/\sqrt{n})/2$ is a compact set contained in the interior of W, because

$|x_1| \geq (1/t+1/\sqrt{n})/2 > (1/t+1/t)/2 = 1/t$ (see the picture describing W). Now if ε is sufficiently small the ellipsoid E_ε is a tiny perturbation of E and hence also the part of E_ε with $|x_1| \geq (1/t + 1/\sqrt{n})/2$ is contained in the interior of W. The proof is complete. □

As an application of Theorem 3.1 we immediately obtain

Corollary 3.5. *In every finitely dimensional normed space a set is compact if and only if it is bounded and closed.*

That means finitely dimensional normed spaces are locally compact. We will prove that this characterizes finitely dimensional normed spaces, i.e no normed space of infinite dimension is locally compact.

Theorem 3.6 (The Riesz lemma)**.** *Let $X_0 \neq X$ be a closed linear subspace of a normed space X. Then for every $\varepsilon > 0$ there is $y \in X$ such that*

$$\|y\| = 1 \quad \textit{and} \quad \|y - x\| \geq 1 - \varepsilon \textit{ for all } x \in X_0 \,.$$

Proof. Fix $y_0 \in X \setminus X_0$ and define $\varrho = \inf_{x \in X_0} \|x - y_0\|$. Clearly $\varrho > 0$, because X_0 is closed. Choose $\eta > 0$ such that $\eta/(\varrho+\eta) \leq \varepsilon$, and then choose $x_0 \in X_0$ such that $\varrho \leq \|y_0 - x_0\| \leq \varrho + \eta$. We will prove that the vector $y = (y_0 - x_0)/\|y_0 - x_0\|$ has the properties that we need. Obviously $\|y\| = 1$. Moreover for $x \in X_0$ we have

$$\begin{aligned} \|y - x\| &= \frac{1}{\|y_0 - x_0\|} \Big\| y_0 - x_0 - \|y_0 - x_0\| x \Big\| \\ &= \frac{1}{\|y_0 - x_0\|} \Big\| y_0 - \underbrace{(x_0 + \|y_0 - x_0\| x)}_{\in X_0} \Big\| \\ &\geq \frac{\varrho}{\varrho + \eta} = 1 - \frac{\eta}{\varrho + \eta} \geq 1 - \varepsilon \,. \end{aligned}$$

The proof is complete. □

Corollary 3.7. *In a normed space of infinite dimension no closed ball is compact.*

Proof. It suffices to prove that the closed unit ball centered at 0 is not compact. To prove the lack of compactness of this ball it suffices to prove the existence of a sequence $\{x_i\} \subset X$ such that

$$\|x_i\| = 1, \quad \|x_i - x_j\| \geq \frac{1}{2} \text{ for } i \neq j \,.$$

We construct the sequence by induction. First we choose x_1 with $\|x_1\| = 1$ arbitrarily. Now suppose that the elements $x_1, \ldots, x_n$ have already been defined. Let $X_0 = \operatorname{span}\{x_1, \ldots, x_n\}$. Since X is infinitely dimensional, $X_0 \neq X$ and hence the Riesz lemma implies that there is $x_{n+1} \in X \setminus X_0$ such that

$$\|x_{n+1}\| = 1 \quad \text{and } \|x_{n+1} - x\| \geq \frac{1}{2}$$

for every $x \in X_0$. $\square$

Note that if in the Riesz lemma $\dim X_0 < \infty$, by local compactness of X_0 we can take in the proof $x_0 \in X_0$ such that $\|y_0 - x_0\| = \varrho$ and hence the same proof gives

Corollary 3.8. *Let $X_0 \neq X$ be a closed linear subspace of a normed space X with $\dim X_0 < \infty$. Then there is $y \in X$ such that*

$$\|y\| = 1 \quad \textit{and} \quad \|y - x\| \geq 1 \textit{ for all } x_0 \in X_0 \ .$$

In general Corollary 3.8 does not hold if $\dim X_0 = \infty$ as the next example shows, see, however, Theorem 14.11.

Example. Consider the closed linear subspace X of $C[0,1]$ consisting of functions vanishing at 0. Let

$$X_0 = \left\{ f \in X : \int_0^1 f(x)\, dx = 0 \right\}.$$

It is easy to see that X_0 is a proper closed linear subspace of X. We will prove that there is no function $f \in X$ such that

$$\|f\| = 1 \quad \text{and} \quad \|f - g\| \geq 1 \quad \text{for all } g \in X_0\,. \tag{3.4}$$

Assume that such a function $f \in X$ exists. Since f is continuous, $f(0) = 0$ and $\|f\| = \sup_{x\in[0,1]} |f(x)| = 1$, we conclude that

$$\int_0^1 |f(x)|\, dx < 1\,. \tag{3.5}$$

For every $h \in X \setminus X_0$ we set

$$g = f - ch, \quad c = \frac{\int_0^1 f(x)\, dx}{\int_0^1 h(x)\, dx}$$

and note that the denominator in nonzero, because $h \notin X_0$. Clearly $g \in X_0$ and (3.4) yields

$$1 \leq \|f - g\| = \|f - (f - ch)\| = |c|\, \|h\|\,,$$

i.e.

$$\Big| \int_0^1 h(x)\, dx \Big| \leq \Big| \int_0^1 f(x)\, dx \Big| \sup_{x\in[0,1]} |h(x)|\,.$$

Choosing $h(x) = x^{1/n} \in X \setminus X_0$ gives

$$\frac{n}{n+1} \leq \Big| \int_0^1 f(x)\, dx \Big| \quad \text{for all } n = 1, 2, 3, \ldots$$

Now passing to the limit as $n \to \infty$ we obtain

$$\int_0^1 |f(x)|\, dx = 1$$

which contradicts (3.5). $\square$

4. Operations on Banach spaces

4.1. **Subspace.** If $(X, \|\cdot\|)$ is a normed space and $Y \subset X$ is a linear subspace, then $(Y, \|\cdot\|)$ is a normed space.

Proposition 4.1. *A subspace Y of a Banach space X is a Banach space if and only if it is a closed subspace of X.*

This result is obvious.

Every metric space is isometric with a dense subspace of a complete metric space. The idea of the proof is to add to the space points which are abstract limits of Cauchy sequences. More precisely, in the space

$$X' = \big\{\{x_n\}_{n=1}^\infty \subset X : \{x_n\}_{n=1}^\infty \text{ is a Cauchy sequence}\big\}$$

we introduce the equivalence relation

$$\{x_n\} \sim \{y_n\} \quad \text{if and only if} \quad d(x_n, y_n) \to 0 \,.$$

Note that if $\{x_n\}$ and $\{y_n\}$ are convergent sequences, then $\{x_n\} \sim \{y_n\}$ if and only if they have the same limit. Thus the relation $\sim$ identifies those Cauchy sequences that should have the same limit. We define $\hat{X} = X'/\sim$ with the metric

$$d([\{x_n\}], [\{y_n\}]) = \lim_{n\to\infty} d(x_n, y_n) \,.$$

The space X can be identified with a subset of $\hat{X}$ through the embedding onto constant sequences

$$X \ni x \mapsto \{x, x, x, \ldots\}$$

After this identification X becomes a dense subset of $\hat{X}$. It is easy to verify that $\hat{X}$ is a complete metric space.

If X is equipped with a structure of a normed space, then also $\hat{X}$ has a natural normed space structure and hence it is a Banach space. Thus we sketched a proof of the following result.

Theorem 4.2. *Every normed space is isometrically isomorphic to a dense linear subspace of a Banach space.*

4.2. **Quotient space.** If X is a linear space and $Y \subset X$ is a linear subspace, then the equivalence relation in X

$$x \sim y \quad \text{if and only if} \quad x - y \in Y$$

defines the linear quotient space

$$X/Y = X/\sim \,.$$

Elements of the space X/Y can be identified with *cosets* $[x] = \{x + y : y \in Y\}$.

Definition. By a seminorm on a liner space X we mean any function $|\cdot|: X\to[0,\infty)$ such that

(1) $|x+y|\leq|x|+|y|$ for $x,y\in X$;
(2) $|\alpha x|=|\alpha|\,|x|$ for all $x\in X$, $\alpha\in\mathbb{K}$.

The only difference between the seminorm and the norm is that it can vanish on nonzero elements. A seminorm vanishes on a linear subspace if X.

If X is a normed space and $Y\subset X$ is a linear subspace, then we equip X/Y with a seminorm

$$\|[x]\|=\inf_{z\in[x]}\|z\|=\inf_{y\in Y}\|x-y\|=\operatorname{dist}(x,Y)\,.$$

Theorem 4.3. *The quotient space X/Y is a normed space if and only if Y is a closed subspace of X.*

Proof. $\Rightarrow$. Suppose that $\|\cdot\|$ is a norm in X/Y, but Y is not closed. Then there is $Y\ni y_n\to x_0\notin Y$. Since $x_0\notin Y$, $[x_0]\neq 0$ in X/Y and hence $\|[x_0]\|>0$ (because $\|\cdot\|$ is a norm). On the other hand $\|[x_0]\|\leq\inf_n\|x_0-y_n\|=0$ which is a contradiction.

$\Leftarrow$. If $\|[x]\|=0$, then there is a sequence $y_n\in Y$ such that $\|x-y_n\|\to 0$ and hence $x\in\overline{Y}=Y$ which yields $[x]=0$. Thus $\|\cdot\|$ is a norm. $\square$

Lemma 4.4. *A normed space X is a Banach space if and only if the absolute convergence of the series $\sum_{n=1}^\infty\|x_n\|<\infty$ implies its convergence $\sum_{n=1}^\infty x_n$ in X.*

Proof. $\Rightarrow$. This implication is obvious, because the absolute convergence implies that $\{x_n\}$ is a Cauchy sequence.

$\Leftarrow$. Let $\{a_n\}\subset X$ be a Cauchy sequence. In order to prove that $\{a_n\}$ is convergent to an element in X it suffices to prove that it contains a convergent subsequence (if a Cauchy sequence has a convergent subsequence, then the entire sequence is convergent). Let $\{a_{n_k}\}$ be a subsequence such that

$$\|a_{n_k}-a_{n_{k+1}}\|\leq 2^{-k}\,.$$

Then

$$\sum_{k=1}^\infty\|a_{n_k}-a_{n_{k+1}}\|<\infty$$

and by our assumption, the series

$$\sum_{k=1}^\infty(a_{n_{k+1}}-a_{n_k})$$

is convergent. That means the sequence whose $(k-1)$th element is

$$(a_{n_2}-a_{n_1})+\ldots+(a_{n_k}-a_{n_{k-1}})=a_{n_k}-a_{n_1}$$

is convergent, and hence the sequence $\{a_{n_k}\}$ is convergence as well. □

Theorem 4.5. *If Y is a closed linear subspace of a Banach space X, then the quotient space X/Y is a Banach space.*

Proof. It suffices to prove that

$$\sum_{n=1}^{\infty} \|[x_n]\| < \infty \quad \Rightarrow \quad \sum_{n=1}^{\infty}[x_n] \text{ is convergent}\,.$$

There are elements $y_n \in Y$ such that $\sum_{n=1}^{\infty}\|x_n+y_n\| < \infty$, so it follows from Lemma 4.4 that the series $\sum_{n=1}^{\infty}(x_n+y_n)$ converges to an element $x_0 \in X$, i.e.

$$\Big\|\sum_{n=1}^{k}(x_n+y_n) - x_0\Big\| \to 0 \quad \text{as } k \to \infty\,.$$

Hence

$$\Big\|\Big((\sum_{n=1}^{k} x_n) - x_0\Big) + \underbrace{\sum_{n=1}^{k} y_n}_{\in Y}\Big\| \to 0 \quad \text{as } k \to \infty$$

which yields

$$\Big\|\sum_{n=1}^{k}[x_n] - [x_0]\Big\| = \Big\|\Big[(\sum_{n=1}^{k} x_n) - x_0\Big]\Big\| \to 0 \quad \text{as } k \to \infty$$

and thus

$$\sum_{n=1}^{\infty}[x_n] = [x_0]\,.$$

The proof is complete. □

Exercise. *Let $x = (x_1, x_2, x_3, \ldots) \in \ell^\infty$. Prove that in the quotient space ℓ^∞/c_0*

$$\|[x]\| = \limsup_{n\to\infty} |x_n|\,.$$

Exercise. *Let Y be a closed subspace of c consisting of constant sequences. Prove that the quotient space c/Y is isomorphic to c_0.*

Now we will prove the following quite surprising result.

Theorem 4.6. *Every separable Banach space X is isometrically isomorphic to a quotient space ℓ^1/Y where Y is a closed subspace of ℓ^1.*

Proof. Let $x_1, x_2, x_3, \ldots$ be a countable and dense subset in the unit sphere $\{x \in X : \|x\| = 1\}$. Let $T : \ell^1 \to X$ be defined by

$$T(\lambda_1, \lambda_2, \ldots) = \sum_{n=1}^{\infty} \lambda_n x_n\,.$$

Then

$$\|T\lambda\| = \|\sum_{n=1}^{\infty}\lambda_n x_n\| \leq \|\lambda\|_1\,, \tag{4.1}$$

so T is continuous. We will prove that T is a surjection. For every $x\in X$, every positive integer k and every $\varepsilon>0$ there is $n>k$ such that

$$\big\|x-\|x\|x_n\big\|<\varepsilon\,.$$

Indeed, it is obvious for $x=0$; if $x\neq 0$ the inequality is equivalent to

$$\Big\|\frac{x}{\|x\|}-x_n\Big\|<\frac{\varepsilon}{\|x\|}$$

and the existence of x_n follows from the density in the unit sphere. Let n_1 be such that

$$\|x-\lambda_{n_1}x_{n_1}\|<\frac{\varepsilon}{2},\quad \text{where } \lambda_{n_1}=\|x\|\,.$$

Let $n_2>n_1$ be such that

$$\|(x-\lambda_{n_1}x_{n_1})-\lambda_{n_2}x_{n_2}\|<\frac{\varepsilon}{4},\quad \text{where } \lambda_{n_2}=\|x-\lambda_{n_1}x_{n_1}\|<\frac{\varepsilon}{2}\,.$$

Let $n_3>n_2$ be such that

$$\|(x-\lambda_{n_1}x_{n_1}-\lambda_{n_2}x_{n_2})-\lambda_{n_3}x_{n_3}\|<\frac{\varepsilon}{8},\quad \text{where } \lambda_{n_3}=\|x-\lambda_{n_1}x_{n_1}-\lambda_{n_2}x_{n_2}\|<\frac{\varepsilon}{4}$$

etc. We obtain a sequence $\lambda_{n_1},\lambda_{n_2},\ldots$ such that

$$\lambda_{n_{k+1}}=\|x-(\lambda_{n_1}x_{n_1}+\ldots+\lambda_{n_k}x_{n_k})\|<\frac{\varepsilon}{2^k}\,. \tag{4.2}$$

Let

$$\lambda_{x,\varepsilon}=(\lambda_1,\lambda_2,\lambda_3,\ldots)$$

where we put $\lambda_i=0$ if $i\not\in\{n_1,n_2,n_3,\ldots\}$. Clearly $\lambda_{x,\varepsilon}\in\ell^1$, because

$$\|\lambda_{x,\varepsilon}\|_1\leq\|x\|+\frac{\varepsilon}{2}+\frac{\varepsilon}{4}+\frac{\varepsilon}{8}+\ldots=\|x\|+\varepsilon\,.$$

Now continuity of T and (4.2) implies that $T(\lambda_{x,\varepsilon})=x$. This proves that T is a surjection onto X. Let

$$Y=\ker T=\{\lambda\in\ell^1:\,T(\lambda)=0\}\,.$$

Thus T induces an algebraic isomorphism of linear spaces ℓ^1/Y onto X. Continuity of T implies that Y is a closed subspace of ℓ^1, so ℓ^1/Y is a Banach space. Note that if $T(\lambda)=x$, then $\ell^1/Y\ni[\lambda]=\{\gamma:\,T(\gamma)=x\}$. We will prove that this algebraic isomorphism is actually an isometry. To this end we have to prove that if $T(\lambda)=x$, then $\|[\lambda]\|=\|x\|$. Let $T(\lambda)=x$. Since $T(\lambda_{x,\varepsilon})=x$ we have $\lambda_{x,\varepsilon}\in[\lambda]$ and hence

$$\|[\lambda]\|\leq\|\lambda_{x,\varepsilon}\|\leq\|x\|+\varepsilon$$

for every $\varepsilon>0$ and thus

$$\|[\lambda]\|\leq\|x\|\,. \tag{4.3}$$

On the other hand if $T(\lambda) = x$, then for every $\gamma \in [\lambda]$ inequality (4.1) implies that $\|x\| = \|T(\gamma)\| \le \|\gamma\|_1$ and hence

$$\|[\lambda]\| = \inf_{\gamma\in[\lambda]} \|\gamma\|_1 \ge \|x\| \,. \tag{4.4}$$

The two inequalities (4.3) and (4.4) imply that $\|[\lambda]\| = \|x\|$. This proves that T is an isometry of ℓ^1/Y onto X. □

Definition. For a closed subspace $M \subset X$ of a normed space X we define the *annihilator*

$$M^\perp = \{x^* \in X^* : \langle x^*, y\rangle = 0 \text{ for all } y \in M\} \,.$$

Clearly $M^\perp$ is a closed subspace of X^* and hence it is a Banach space.[2]

Theorem 4.7. *Let X be a normed space and $M \subset X$ a closed subspace. Then the dual space $(X/M)^*$ to the quotient space is isometrically isomorphic to $M^\perp$.*

Proof. Let $\pi : X \to X/M$, $\pi(x) = [x]$ be the quotient map. We define the map $T : (X/M)^* \to X^*$ by

$$\langle T(z^*), x\rangle = \langle z^*, [x]\rangle \quad \text{for } z^* \in (X/M)^* \text{ and } x \in X.$$

Since

$$|\langle T(z^*), x\rangle| \le \|z^*\| \, \|[x]\| \le \|z^*\| \, \|x\| \,,$$

T is bounded and

$$\|T(z^*)\| \le \|z^*\| \,. \tag{4.5}$$

We actually have equality in (4.5). To see this we need to prove opposite inequality. Given $\varepsilon > 0$ let $[x] \in X/M$ be such that

$$\|[x]\| = 1, \quad |\langle z^*, [x]\rangle| \ge \|z^*\| - \varepsilon \,.$$

Since

$$1 = \|[x]\| = \inf_{y\in[x]} \|y\| \,,$$

there is $y \in [x]$ such that $\|y\| < 1 + \varepsilon$ and obviously $[y] = [x]$. Hence

$$\begin{aligned} \|z^*\| - \varepsilon &\le |\langle z^*, [x]\rangle| = |\langle z^*, [y]\rangle| = |\langle T(z^*), y\rangle| \\ &\le \|T(z^*)\|(1+\varepsilon) \end{aligned}$$

and letting $\varepsilon \to 0$ yields $\|z^*\| \le \|T(z^*)\|$ which together with (4.5) proves

$$\|T(z^*)\| = \|z^*\| \,.$$

We proved that T is an isometric embedding of $(X/M)^*$ onto a closed subspace of X^*. Actually

$$T((X/M)^*) \subset M^\perp \,.$$

[2] Even if X is only a normed space.

Indeed, for $z^* \in (X/M)^*$, $T(z^*) \in M^\perp$, because for $y \in M$

$$\langle T^*(z^*), y\rangle = \langle z^*, [y]\rangle = \langle z^*, 0\rangle = 0\,.$$

It remains to show that T is a surjection onto $M^\perp$. Let $x^* \in M^\perp$. It is easy to see that

$$\langle z^*, [y]\rangle = \langle x^*, y\rangle \quad \text{for } y \in X \tag{4.6}$$

is a well defined[3] and bounded[4] functional $z^* \in (X/M)^*$ such that $T(z^*) = x^*$, so T is surjective. □

Theorem 4.8. *If $L \in B(X,Y)$ is a bounded mapping between Banach spaces such that* $\dim(Y/L(X)) < \infty$*, then $L(X)$ is a closed subspace of Y.*

Proof. Let $\pi : Y \to Y/L(X)$ be a quotient mapping.[5] Let $\{z_1, \ldots, z_n\}$ be a Hamel basis of $Y/L(X)$ and let $y_1, \ldots, y_n \in Y$ be such that $\pi(y_i) = z_i$. The space $V = \operatorname{span}\{y_1, \ldots, y_n\}$ satisfies

$$L(X) + V = Y, \quad L(X) \cap V = \{0\}\,.$$

Note that V is a closed subspace as finitely dimensional, so Y/V is a Banach space and the quotient mapping

$$\eta : Y \to Y/V$$

is bounded. The mapping

$$\tilde{L} : X/\ker L \to Y, \quad \tilde{L}([x]) = Lx$$

is bounded, $\|\tilde{L}\| \le \|L\|$ and it is a bijection onto $\tilde{L}(X) = L(X)$. Now the composition

$$X/\ker L \xrightarrow{\tilde{L}} Y \xrightarrow{\eta} Y/V$$

is a bijection, so it is an isomorphism of Banach spaces. Hence

$$\tilde{L} \circ (\eta \circ \tilde{L})^{-1} : Y/V \to L(X)$$

is an isomorphism[6] of normed spaces and thus $L(X)$ is a Banach space by Proposition 2.6. Therefore $L(X)$ is closed in Y. □

---

[3]If $[y_1] = [y_2]$, then $y_1 - y_2 \in M$ and hence $\langle x^*, y_1\rangle = \langle x^*, y_2 + (y_1 - y_2)\rangle = \langle x^*, y_2\rangle$ and hence (4.6) is well defined.

[4]For $\tilde{y} \in [y]$ we have $|\langle z^*, [y]\rangle| = |\langle x^*, \tilde{y}\rangle| \le \|x^*\|\|\tilde{y}\|$ and taking infimum over $\tilde{y} \in [y]$ yields $|\langle z^*, [y]\rangle| \le \|x^*\|\,\|[y]\|$ which proves boundedness of z^*.

[5]We consider it only as an algebraic mapping. Since we do not know yet that $L(X)$ is closed, so we do not know that $Y/L(X)$ is a normed space.

[6]It is an isomorphism, because the inverse is η restricted to $L(X)$. Indeed, $\eta \circ \tilde{L} \circ (\eta \circ \tilde{L})^{-1} = \mathrm{id}$ on $L(X)$.

4.3. **Direct sum.**

DEFINITION. If X and Y are normed spaces with the norms $\|\cdot\|_1$ and $\|\cdot\|_2$ respectively, then we define $X \oplus Y$ as $X \times Y$ equipped with the norm $\|(x,y)\| = \|x\|_1 + \|y\|_2$. The following proposition is obvious.

Proposition 4.9. *If X and Y are Banach spaces, then $X \oplus Y$ is a Banach space as well.*

Example. The space $C^n[a,b]$ of real functions whose derivatives of order up to n are continuous on $[a,b]$ is a Banach space with respect to the norm

$$\|f\| = \sup_{x\in[a,b]} |f(x)| + \sup_{x\in[a,b]} |f'(x)| + \ldots + \sup_{x\in[a,b]} |f^{(n)}(x)| \,.$$

This space is isomorphic to $\mathbb{R}^n \oplus C[a,b]$. Indeed, the mapping

$$C^n[a,b] \ni f \mapsto (f(a), f'(a), \ldots, f^{(n-1)}(a), f^{(n)}) \in \mathbb{R}^n \oplus C[a,b]$$

is an isomorphism. This follows from the Taylor formula.

EXERCISE. *Provide a detailed proof of the above fact.*

In the case of Hilbert spaces we define a direct sum in a slightly different way than in the case of Banach spaces.

DEFINITION. If $(H_1, \langle\cdot,\cdot\rangle_1)$, $(H_2, \langle\cdot,\cdot\rangle_2)$ are Hilbert spaces, then their direct sum is the space $H_1 \oplus H_2 = H_1 \times H_2$ with the inner product

$$\langle (x_1,x_2),(y_1,y_2)\rangle = \langle x_1,y_1\rangle_1 + \langle x_2,y_2\rangle_2 \,.$$

It is easy to see that $(H_1 \oplus H_2, \langle\cdot,\cdot\rangle)$ is a Hilbert space.

Observe, however, that the corresponding norm is

$$\|(x_1,x_2)\| = \left(\|x_1\|_1^2 + \|x_2\|_2^2\right)^{1/2}$$

which is different, (but equivalent) than the norm defined in the case of a direct sum of Banach spaces.

We can also define a direct sum of an infinite sequence of Hilbert spaces.

DEFINITION. If $(H_i, \langle\cdot,\cdot\rangle_i)$, $i = 1,2,\ldots$ are Hilbert spaces, then the direct sum is the space

$$\bigoplus_{i=1}^{\infty} H_i = H_1 \oplus H_2 \oplus \ldots = \left\{ (x_i)_{i=1}^{\infty} :\ x_i \in H_i,\ i = 1,2,\ldots,\ \sum_{i=1}^{\infty} \|x_i\|^2 < \infty \right\}$$

equipped with the inner product

$$\langle (x_i),(y_i)\rangle = \sum_{i=1}^{\infty} \langle x_i, y_i\rangle_i$$

and hence
$$\|(x_i)\| = \Big(\sum_{i=1}^\infty \|x_i\|_i^2\Big)^{1/2}.$$

Theorem 4.10. *$H_1 \oplus H_2 \oplus \ldots$ is a Hilbert space.*

EXERCISE. *Prove the above theorem.*

Example. $\mathbb{R} = \ell^2_1$ is a one dimensional Hilbert space and it is easy to see that
$$\mathbb{R} \oplus \mathbb{R} \oplus \ldots = \ell^2 .$$

5. Hilbert spaces

In this section we will develop a basic theory of Hilbert spaces. In particular we will prove the existence of an orthonormal basis. We will also show many interesting applications.

DEFINITION. A subset E of a linear space V is *convex* if
$$x, y \in E,\ t \in [0,1] \quad \Rightarrow \quad z_t = (1-t)x + ty \in E .$$

Note that if E is convex, then its translation
$$E + x = \{y + x : \, y \in E\}$$
is also convex.

DEFINITION. Let H be a Hilbert space. We say that the vectors $x, y \in H$ are *orthogonal* if $\langle x, y\rangle = 0$. We denote orthogonal vectors by $x \perp y$.

It is easy to see that if $x \perp y$, then $\|x+y\|^2 = \|x\|^2 + \|y\|^2$. This fact is known as the *Pythagorean theorem.*

For $x \in H$ we define
$$x^\perp = \{y \in H : \, y \perp x\}$$
and for a linear subspace $M \subset H$
$$M^\perp = \{y \in H : \, y \perp x \text{ for all } x \in M\} .$$
It is easy to see that $x^\perp$ is a closed subspace of H and hence
$$M^\perp = \bigcap_{x\in M} x^\perp$$
is also a closed subspace of H, even if M is not closed.

Theorem 5.1. *Every nonempty, convex and closed set E in a Hilbert space H contains a unique element of smallest norm.*

Proof. Recall the Parallelogram Law:

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2 \quad \text{for all } x, y \in H\,.$$

Let $\delta = \inf\{\|x\| : x \in E\}$. Applying the Parallelogram Law to $x/2$ and $y/2$ we have

$$\frac{1}{4}\|x-y\|^2 = \frac{1}{2}\|x\|^2 + \frac{1}{2}\|y\|^2 - \left\|\frac{x+y}{2}\right\|^2.$$

Since E is convex and $x, y \in E$ we also have $(x+y)/2 \in E$. Accordingly

$$\|x-y\|^2 \leq 2\|x\|^2 + 2\|y\|^2 - 4\delta^2. \tag{5.1}$$

If $\|x\| = \|y\| = \delta$, then $x = y$ which implies uniqueness of the element with the smallest norm. Now let $y_n \in E$, $\|y_n\| \to \delta$. Then (5.1) yields

$$\|y_n - y_m\|^2 \leq 2\|y_n\|^2 + 2\|y_m\|^2 - 4\delta^2 \to 0 \quad \text{as } n, m \to \infty\,.$$

Hence $\{y_n\}$ is a Cauchy sequence in H. Thus $y_n \to x_0 \in H$. Since E is closed, $x_0 \in E$ and continuity of the norm yields

$$\|x_0\| = \lim_{n\to\infty} \|y_n\| = \delta.$$

The proof is complete. □

The next example shows that the property described in the above theorem is not true in every Banach space. However, it is true in reflexive spaces, see Theorem 14.10.

Example. For $y = (y_1, y_2, \ldots) \in \ell^\infty$ define $\varphi : \ell^1 \to \mathbb{C}$ by

$$\varphi(x) = \sum_{n=1}^{\infty} x_n y_n \quad \text{where } x = (x_1, x_2, \ldots).$$

Since $\varphi \in (\ell^1)^*$

$$W = \{x \in \ell^1 : \varphi(x) = 1\}$$

is convex and closed. If $y = (1, 1, 1, \ldots)$, then

$$W = \{x \in \ell^1 : \sum_{n=1}^{\infty} x_i = 1\}$$

so for $x \in W$, $\|x\| = \sum_{n=1}^{\infty} |x_n| \geq 1$ and hence every vector $x \in W$ with $x_n \geq 0$ for all n has the smallest norm equal 1. Thus we have infinitely many vectors of smallest norm.

On the other hand if $y = (1/2, 2/3, 3/4, 4/5, \ldots)$, then

$$W = \left\{x \in \ell^1 : \sum_{n=1}^{\infty} \frac{n}{n+1}\, x_n = 1\right\}.$$

In particular $e_n(n+1)/n \in W$, where $e_n = (0,\ldots,0,1,0,\ldots)$ with 1 on nth coordinate and $\|e_n(n+1)/n\|_1 = (n+1)/n$, so $\operatorname{dist}(0,W) \leq 1$, but also $\operatorname{dist}(0,W) \geq 1$ because for $x \in W$

$$\|x\|_1 = \sum_{n=1}^{\infty} |x_n| > \sum_{n=1}^{\infty} \frac{n}{n+1}\,|x_n| \geq \sum_{n=1}^{\infty} \frac{n}{n+1}\,x_n = 1\,.$$

Thus $\operatorname{dist}(0,W) = 1$. This also proves that there are no vectors in W of smallest norm because $\|x\|_1 > 1$ for any $x \in W$. $\square$

Another example is provided in the next exercise.

EXERCISE. *Let $M \subset C[0,1]$ be a subset consisting of all functions f such that*

$$\int_0^{1/2} f(t)\,dt - \int_{1/2}^{1} f(t)\,dt = 1\,.$$

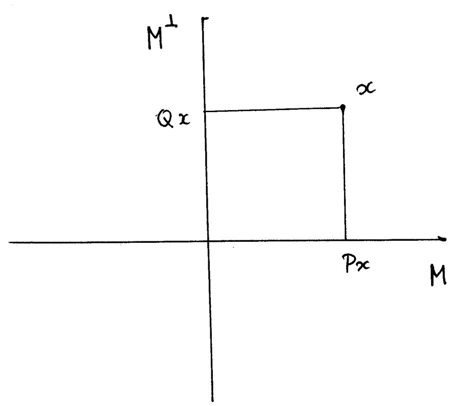
Prove that M is a closed convex subset of $C[0,1]$ that has no element of minimal norm.

Theorem 5.2. *Let M be a closed subspace of Hilbert space H. Then*

(a) *Every $x \in H$ has unique decomposition*

$$x = Px + Qx \quad \text{where} \quad Px \in M,\ Qx \in M^{\perp}\ .$$

(b) *Px and Qx are nearest points to x in M and $M^{\perp}$ respectively.*
(c) *The mappings $P : H \to M$, $Q : H \to M^{\perp}$ are linear.*
(d) $\|x\|^2 = \|Px\|^2 + \|Qx\|^2$.



The mappings P and Q are called *orthogonal projections* of H onto M and $M^{\perp}$.

Corollary 5.3. *If $M \neq H$ is a closed subspace, then there is $y \in H$, $y \neq 0$, $y \perp M$.*

Corollary 5.4. *If M is a closed subspace of a Hilbert space H, then H/M is isometrically isomorphic to $M^{\perp}$. In particular it is a Hilbert space.*

Observe that this is a very different situation than in the case of Theorem 4.6.

Proof of Theorem 5.2. (a) Uniqueness is easy. If

$$x = x' + y' = x'' + y'', \quad x', x'' \in M, \quad y', y'' \in M^{\perp},$$

then

$$M \ni x' - x'' = y'' - y' \in M^{\perp}.$$

Since $M \cap M^{\perp} = \{0\}$ (because $\langle x, x\rangle = 0$ implies $x = 0$) we conclude that $x' = x''$ and $y' = y''$. Thus we are left with the proof of the existence of the decomposition. The set

$$x + M = \{x + y : \ y \in M\}$$

is convex and closed. Let Qx be the element of the smallest norm in $x + M$ and let $Px = x - Qx$. Clearly $x = Px + Qx$. Since $Qx \in x + M$, it follows that $Px \in M$. We still need to prove that $Qx \in M^{\perp}$. To this end we have to prove that $\langle Qx, y\rangle = 0$ for all $y \in M$. We can assume that $\|y\| = 1$. Denote $z = Qx$. The minimizing property of Qx shows that

$$\langle z, z\rangle = \|z\|^2 \leq \|z - \alpha y\|^2 = \langle z - \alpha y, z - \alpha y\rangle$$

for all $\alpha \in \mathbb{K}$. Hence

$$\langle z, z\rangle \leq \langle z, z\rangle + |\alpha|^2 \underbrace{\langle y, y\rangle}_{1} - \alpha\langle y, z\rangle - \overline{\alpha}\langle z, y\rangle,$$

$$0 \leq |\alpha|^2 - \alpha\langle y, z\rangle - \overline{\alpha}\langle z, y\rangle.$$

Taking $\alpha = \langle z, y\rangle$ we have

$$0 \leq -|\langle z, y\rangle|^2,$$

i.e. $\langle z, y\rangle = 0$. This proves that $z = Qx \in M^{\perp}$.

(b) We know that $Px \in M$. If $y \in M$, then

$$\|x - y\|^2 = \|Qx + (Px - y)\|^2 = \|Qx\|^2 + \|Px - y\|^2.$$

The last equality follows from $Qx \perp Px - y$. Thus the minimal value of $\|x - y\|^2$ is attained when $y = Px$, so Px is the nearest point to x in M. Similarly one can prove that Qx is the nearest point to x in $M^{\perp}$.

(c) If we apply (a) to x, y and $\alpha x + \beta y$, we obtain

$$\alpha x + \beta y = P(\alpha x + \beta y) + Q(\alpha x + \beta y),$$

$$\alpha(Px + Qx) + \beta(Py + Qy) = P(\alpha x + \beta y) + Q(\alpha x + \beta y),$$

$$\underbrace{\alpha Px + \beta Py - P(\alpha x + \beta y)}_{\in M} = \underbrace{Q(\alpha x + \beta y) - \alpha Qx - \beta Qy}_{\in M^{\perp}}.$$

Since $M \cap M^{\perp} = \{0\}$ both sides are equal zero which proves linearity both of P and Q.

(d) This is a direct consequence of the orthogonality of Px and Qx. $\square$

Example. This will be an illustration for a striking application of Corollary 5.3 in the finitely dimensional case. We will prove that for every integer $m \geq 0$ there is $\varphi \in C_0^\infty(B^n(0,1))$ such that

$$\int_{\mathbb{R}^n} \varphi(x)\, dx = 1, \quad \int_{\mathbb{R}^n} \varphi(x) x^\alpha\, dx = 0 \text{ for } 0 < |\alpha| \leq m\,.$$

Indeed, let N be the number of multiindices $\alpha = (\alpha_1, \ldots, \alpha_n)$ such that $|\alpha| \leq m$. Define a linear map

$$T : C_0^\infty(B^n(0,1)) \to \mathbb{R}^N\,, \quad T(\varphi) = (\xi_\alpha)_{|\alpha|\leq m}$$

where

$$\xi_\alpha = \int_{\mathbb{R}^n} \varphi(x) x^\alpha\, dx\,.$$

It suffices to prove that the mapping T is surjective. For if not the image $\operatorname{im} T \neq \mathbb{R}^N$ is a closed subspace of $\mathbb{R}^N$ and hence there is a nonzero vector $0 \neq (\eta_\alpha)_{|\alpha|\leq m} \perp \operatorname{im} T$, i.e.

$$0 = \sum_{|\alpha|\leq m} \eta_\alpha \int_{\mathbb{R}^n} \varphi(x) x^\alpha\, dx = \int_{\mathbb{R}^n} \varphi(x) \sum_{|\alpha|\leq m} \eta_\alpha x^\alpha\, dx \tag{5.2}$$

for all $\varphi \in C_0^\infty(B^n(0,1))$. In particular if $\psi \in C_0^\infty(B^n(0,1))$, then

$$\varphi(x) = \psi(x) \sum_{|\alpha|\leq m} \eta_\alpha x^\alpha \in C_0^\infty(B^n(0,1))$$

and hence (5.2) yields

$$0 = \int_{\mathbb{R}^n} \psi(x) \Big| \sum_{|\alpha|\leq m} \eta_\alpha x^\alpha \Big|^2\, dx$$

for all $\psi \in C_0^\infty(B^n(0,1))$. Hence $\sum_{|\alpha|\leq m} \eta_\alpha x^\alpha = 0$ in $B^n(0,1)$ and since it is a polynomial, $\eta_\alpha = 0$ for all $|\alpha| \leq m$ which is a contradiction with $(\eta_\alpha)_{|\alpha|\leq m} \neq 0$. □

Recall that for a function $\varphi \in C_0^\infty(\mathbb{R}^n)$ and $\varepsilon > 0$ we define $\varphi_\varepsilon(x) = \varepsilon^{-n}\varphi(x/\varepsilon)$.

Exercise. *Use the above example to prove that for every integer $m > 0$ there is $\varphi \in C_0^\infty(B^n(0,1))$ such that for every $\varepsilon > 0$*

$$\varphi_\varepsilon * P = P \quad \textit{for all polynomials } P \textit{ of degree} \leq m\,.$$

Obviously for every $y \in H$, $x \mapsto \langle x, y\rangle$ is a bounded linear functional on H. It turns out that every functional on H can be represented in that form.

Theorem 5.5 (Riesz representation theorem)**.** *If L is a continuous linear functional on H, then there is unique $y \in H$ such that*

$$Lx = \langle x, y\rangle \quad \textit{for } x \in H\,.$$

Proof. If $L = 0$, then we take $y = 0$. Otherwise

$$M = \{x : \; Lx = 0\}$$

is a closed linear subspace with $M \neq H$. Thus there is $z \in M^{\perp}$, $\|z\| = 1$ (see Corollary 5.3). For any $x \in H$ let

$$u = (Lx)z - (Lz)x\,.$$

We have

$$Lu = (Lx)(Lz) - (Lz)(Lx) = 0\,,$$

so $u \in M$ and hence $\langle u, z\rangle = 0$, i.e.

$$(Lx)\underbrace{\langle z, z\rangle}_{1} - (Lz)\langle x, z\rangle = 0\,,$$

$$Lx = (Lz)\langle x, z\rangle = \langle x, \underbrace{(\overline{Lz})z}_{y}\rangle\,.$$

The uniqueness is easy. If $\langle x, y\rangle = \langle x, y'\rangle$ for all $x \in H$, then $\langle x, y - y'\rangle = 0$ for all $x \in H$ and in particular for $x = y - y'$

$$\|y - y'\|^2 = \langle y - y', y - y'\rangle = 0, \quad y = y'\,.$$

The proof is complete. □

5.1. **Orthonormal basis.** As set $\{u_\alpha\}_{\alpha\in A} \subset H$ is called *orthonormal* if

$$\langle u_\alpha, u_\beta\rangle = \begin{cases} 1 & \text{if } \alpha = \beta, \\ 0 & \text{if } \alpha \neq \beta. \end{cases}$$

If $\{u_\alpha\}_{\alpha\in A}$ is orthonormal, then with each $x \in H$ we associate *Fourier coefficients* defined by

$$\hat{x}(\alpha) = \langle x, u_\alpha\rangle\,.$$

Proposition 5.6. *Suppose that $\{u_\alpha\}_{\alpha\in A}$ is an orthonormal set in H and F is a finite subset of A. Let*

$$M_F = \operatorname{span}\{u_\alpha\}_{\alpha\in F}\,.$$

(a) *If φ is a complex function on A, equal to zero outside F, then the vector*

$$y = \sum_{\alpha\in F} \varphi(\alpha) u_\alpha \in M_F$$

satisfies $\hat{y}(\alpha) = \varphi(\alpha)$ for $\alpha \in F$ and

$$\|y\|^2 = \sum_{\alpha\in F} |\varphi(\alpha)|^2\,.$$

(b) *If* $x \in H$, *and*

$$s_F(x) = \sum_{\alpha\in F} \hat{x}(\alpha) u_\alpha \,,$$

then

$$\|x - s_F(x)\| \leq \|x - s\| \tag{5.3}$$

for every $s \in M_F$ *with the equality only for* $s - s_F(x)$. *Moreover*

$$\sum_{\alpha\in F} |\hat{x}(\alpha)|^2 \leq \|x\|^2 \,. \tag{5.4}$$

Proof. Part (a) is obvious. To prove (b) denote $s_F = s_F(x)$ and observe that $\hat{s_F}(\alpha) = \hat{x}(\alpha)$ for $\alpha \in F$. That means $(x - s + F) \perp M_F$. In particular $\langle x - s_F, s_F - s\rangle = 0$ and hence

$$\|x - s\|^2 = \|(x - s_F) + (s_F - s)\|^2 = \|x - s_F\|^2 + \|s_F - s\|^2 \tag{5.5}$$

which implies (5.3). Now (5.5) with $s = 0$ gives

$$\|x\|^2 = \|s - s_F\|^2 + \|s_F\|^2 \geq \|s_F\|^2$$

which is (5.4). $\Box$

Remark. The part (b) says that $s_F(x)$ is the best unique approximation of x in M_F, i.e. (see Theorem 5.2) $s_F(x)$ is the orthogonal projection of x onto M_F.

If A is any set, then we define $\ell^\infty(A)$ to be the Banach space of *all* bounded functions on A (no measurability condition), and $\ell^p(A)$, $1 \leq p < \infty$ is the Banach space of p-integrable functions with respect to the counting measure. Thus $\varphi \in \ell^p(A)$, $1 \leq p < \infty$ if $\varphi(\alpha) \neq 0$ for at most countably many α and

$$\sum_{\alpha\in A} |\varphi(\alpha)|^p = \sum_{\substack{\alpha\in A\\ \varphi(\alpha)\neq 0}} |\varphi(\alpha)|^p < \infty \,.$$

Note that

$$\sum_{\alpha\in A} |\varphi(\alpha)|^p = \sup_F \sum_{\alpha\in F} |\varphi(\alpha)|^p$$

where the supremum if over all finite subsets $F \subset A$.

If $A = \mathbb{Z}$, then $\ell^p(A) = \ell^p$. Clearly, the functions φ that are equal zero except on a *finite* subset of A are dense in $\ell^p(A)$, $1 \leq p < \infty$. It is also obvious that $\ell^2(A)$ is a Hilbert space with respect to the inner product

$$\langle \varphi, \psi\rangle = \sum_{\alpha\in A} \varphi(\alpha)\overline{\psi(\alpha)} \,.$$

We will need the following elementary result.

Lemma 5.7. *Suppose that*

(a) X *and* Y *are metric spaces and* X *is complete;*

(b) *$f : X \to Y$ is continuous;*
(c) *X has a dense subset X_0 on which f is an isometry;*
(d) *$f(X_0)$ is dense in Y.*

Then f is an isometry of X onto Y.

An important part of the lemma is that f is a surjection. Recall that f is an isometry if it preserves the distances, i.e.

$$d_Y(f(x), f(y)) = d_X(x, y) \quad \text{for all } x, y \in X \ .$$

Proof. Clearly f is an isometry on X by continuity. To prove that f is a surjection let $y \in Y$. Since $f(X_0)$ is dense in Y, there is a sequence $x_n \in X_0$ such that $d_Y(y, f(x_n)) \to 0$ as $n \to \infty$. This implies that $f(x_n)$ and hence x_n (by isometry) are Cauchy sequences. Hence x_n is convergent, $x_n \to x_0$ (because X is complete) and thus $f(x_0) = y$. □

Theorem 5.8. *Let $\{u_\alpha\}_{\alpha \in A}$ be an orthonormal set in a Hilbert space H, and let P be the space of all finite linear combinations of the vectors u_α, i.e. $P = \operatorname{span}\{u_\alpha\}_{\alpha \in A}$. Then the* Bessel inequality

$$\sum_{\alpha \in A} |\hat{x}(\alpha)|^2 \leq \|x\|^2$$

holds for all $x \in H$ and $x \mapsto \hat{x}$ is a continuous linear mapping of H onto $\ell^2(A)$ whose restriction to the closure $\overline{P}$ is an isometry of $\overline{P}$ onto $\ell^2(A)$.

Proof. The Bessel inequality follows from the fact that

$$\sum_{\alpha \in F} |\hat{x}(\alpha)|^2 \leq \|x\|^2$$

for every finite subset $F \subset A$. Let

$$f : H \to \ell^2(A) , \quad f(x) = \hat{x} \, .$$

Linearity of f is obvious. Since

$$\|f(x) - f(y)\|^2 = \|f(x - y)\|^2_{\ell^2(A)} \leq \|x - y\|^2 ,$$

f is continuous. Moreover f is an isometry of P onto a dense subset of $\ell^2(A)$ consisting of functions in $\ell^2(A)$ which are equal to zero except on a finite set. Hence the theorem follows from the lemma. □

Theorem 5.9. *Let $\{u_\alpha\}_{\alpha \in A}$ be an orthonormal set in a Hilbert space H. Then the following conditions are equivalent.*

(a) *$\{u_\alpha\}_{\alpha \in A}$ is a maximal orthonormal set in H.*
(b) *The space $P = \operatorname{span}\{u_\alpha\}_{\alpha \in A}$ is dense in H.*

(c) *The* Parseval identity
$$\sum_{\alpha\in A} |\hat{x}(\alpha)|^2 = \|x\|^2$$
holds for all $x \in H$.

(d)
$$\sum_{\alpha\in A} \hat{x}(\alpha)\overline{\hat{y}(\alpha)} = \langle x, y\rangle \quad \textit{for all } x, y \in H \ .$$

Proof. (a) $\Rightarrow$ (b). Suppose P is not dense, i.e. $\overline{P} \neq H$. Let $u \in \overline{P}^{\perp}$, $\|u\| = 1$. Then the set $\{u\} \cup \{u_\alpha\}_{\alpha\in A}$ is orthonormal which contradicts maximality of $\{u_\alpha\}_{\alpha\in A}$.

(b) $\Rightarrow$ (c). It follows from the previous theorem
$$\hat{\ }: \overline{P} = H \to \ell^2(A)$$
is an isometry, so the Parseval identity follows.

(c) $\Rightarrow$ (d) this implication follows from the Polarization identity, Proposition 1.3.

(d) $\Rightarrow$ (a). If (a) were false, then there would be $0 \neq u \in H$ such that $\langle u, u_\alpha\rangle = 0$ for all $\alpha \in A$. Let $x = y = u$. We have
$$0 < \langle x, y\rangle = \sum_{\alpha\in A} \underbrace{\hat{x}(\alpha)}_{0}\underbrace{\overline{\hat{y}(\alpha)}}_{0} = 0$$
which is a contradiction. $\square$

Definition. Any maximal orthonormal set in H is called an *orthonormal basis.*

Therefore the above theorem provides several equivalent conditions for an orthonormal set to be a basis.

A direct application of the Hausdorff maximality theorem (equivalent to the axiom of choice) gives

Theorem 5.10. *Every orthonormal set is contained in an orthonormal basis.*

Definition. We say that two Hilbert spaces H_1 and H_2 are *isomorphic* if there is a linear bijection $\Lambda : H_1 \to H_2$ such that
$$\langle \Lambda x, \Lambda y\rangle = \langle x, y\rangle \quad \text{for all } x, y \in H \ .$$

Corollary 5.11 (Riesz-Fisher)**.** *If $\{u_\alpha\}_{\alpha\in A}$ is an orthonormal basis in H, then the mapping $x \mapsto \hat{x}$ is an isometry of H onto $\ell^2(A)$.*

It is easy to see that orthonormal vectors are linearly independent.

Corollary 5.12. *If a Hilbert space H has finite dimension n, then H is isomorphic to ℓ^2_n. If a Hilbert space H is separable and infinitely dimensional, then it is isomorphic to ℓ^2.*

Proof. If $\dim H = n$, then the maximal orthonormal set consists on n vectors. If $\dim H = \infty$ and H is separable, then the maximal orthonormal set is countable, so it can be indexed by integers. □

Theorem 5.13. *If H is a separable Hilbert space and $\{e_n\}_{n\in\mathbb{Z}}$ is an orthonormal basis, then for every $x \in H$*

$$x = \sum_{n=-\infty}^{\infty} \hat{x}(n) e_n$$

in the sense of convergence in H, i.e.

$$\Big\|x - \sum_{|n|\le k} \hat{x}(n) e_n\Big\| \to 0 \quad \text{as } k \to \infty \ .$$

Proof. As we know $\|x - \sum_{|n|\le k} \hat{x}(n)e_n\|$ equals to the distance of x to the space $\mathrm{span}\{e_n\}_{|n|\le k}$ (see Theorem 5.6(b)). Since $\mathrm{span}\,\{e_n\}_{n\in\mathbb{Z}}$ is a dense subset of H this distance converges to 0 as $k \to \infty$. □

As a variant of Theorems 5.9 and 5.13 we obtain

Theorem 5.14. *Let $H_1, H_2, \ldots$ be closed subspaces of a Hilbert space H such that $H_i \perp H_j$ for $i \neq j$ and linear combinations of elements of subspaces H_i are dense in H. Then H is isometrically isomorphic to the direct sum of Hilbert spaces $H_1 \oplus H_2 \oplus \ldots$. More precisely, for very $x \in H$ there are unique elements $x_i \in H_i$ such that*

$$x = \sum_{i=1}^{\infty} x_i$$

in the sense of convergence in H and

$$\|x\|^2 = \sum_{i=1}^{\infty} \|x_i\|^2 \, .$$

Observe that if $\{e_{ji}\}_i$ is an orthonormal basis in each H_i, then $\{e_{ji}\}_{i,j}$ is an orthonormal basis in H.

Exercise. *Prove Theorem 5.14.*

Theorem 5.15. *Let μ and ν be positive σ-finite measures on X and Y respectively. If $\{\varphi_i\}_{i=1}^{\infty}$ is an orthonormal basis in $L^2(\mu)$ and $\{\psi_i\}_{i=1}^{\infty}$ an orthonormal basis in $l^2(\nu)$, then $h_{ij}(x,y) = \varphi_i(x)\psi_j(y)$, $i, j = 1, 2, \ldots$ is an orthonormal basis in $L^2(\mu \times \nu)$.*

Proof. It easily follows from the Fubini theorem that the functions h_{ij} form an orthonormal family in $L^2(\mu\times\nu)$ and it remains to prove that

$$\|f\|_{L^2(\mu\times\nu)} = \sum_{i,j=1}^\infty |\langle f,h_{ij}\rangle|^2 \quad \text{for all } f\in L^2(\mu\times\nu)\ ,$$

see Theorem 5.9(c). If $f\in L^2(\mu\times\nu)$, then for ν-a.e. $y\in Y$, $x\mapsto f_y(x)=f(x,y)\in L^2(\mu)$ by the Fubini theorem. Denoting

$$\widehat{f_y}(i)=\langle f_y,\varphi_i\rangle_{L^2(\mu)} = \int_X f(x,y)\overline{\varphi_i(x)}\, d\mu(x)$$

we have

$$\int_X |f(x,y)|^2\, d\mu(x) = \sum_{i=1}^\infty |\widehat{f_y}(i)|^2$$

for ν-a.e. $y\in Y$. Note that for $i=1,2,\ldots$

$$y\mapsto \widehat{f_y}(i)\in L^2(\nu)$$

because

$$\begin{aligned}\int_Y |\widehat{f_y}(i)|^2\, d\nu(y) &= \int_Y |\langle f_y,\varphi_i\rangle_{L^2(\mu)}|^2\, d\nu(y)\\ \leq &\int_Y \|f_y\|^2_{L^2(\mu)} \underbrace{\|\varphi_i\|^2_{L^2(\mu)}}_{1}\, d\nu(y) = \int_Y\int_X |f(x,y)|^2\, d\mu(x)\, d\nu(y)<\infty\,.\end{aligned}$$

Hence

$$\begin{aligned}\|f\|^2_{L^2(\mu\times\nu)} &= \int_Y\left(\int_X |f(x,y)|^2\, d\mu(x)\right)\, d\nu(y)\\ &= \int_Y \sum_{i=1}^\infty |\widehat{f_y}(i)|^2\, d\nu(y) = \sum_{i=1}^\infty \int_Y |\widehat{f_y}(i)|^2\, d\nu(y)\\ &= \sum_{i=1}^\infty\sum_{j=1}^\infty |\langle \widehat{f_y}(i),\psi_j\rangle_{L^2(\nu)}|^2\\ &= \sum_{i,j=1}^\infty \left|\int_Y\left(\int_X f(x,y)\overline{\varphi_i(x)}\, d\mu(x)\right)\overline{\psi_j(y)}\, d\nu(y)\right|^2\\ &= \sum_{i,j=1}^\infty |\langle f,h_{ij}\rangle|^2\,.\end{aligned}$$

the proof is complete. □

5.2. **Gramm-Schmidt orthogonalization.** If H is a separable Hilbert space, then an orthonormal basis can be constructed through the *Gramm-Schmidt orthogonalization* that we describe next. Let $\{f_1,f_2,\ldots\}$ be a dense subset of H. We remove (by induction) the function f_n if it is a linear combination of $f_1, \ldots, f_{n-1}$. Denote the remaining functions by $g_1, g_2, \ldots$. Now the orthonormal basis can be defined as follows

$$e_1 = \frac{g_1}{\|g_1\|}, \quad e_n = \frac{g_n - \sum_{k=1}^{n-1} \langle g_n, e_k \rangle e_k}{\left\| g_n - \sum_{k=1}^{n-1} \langle g_n, e_k \rangle e_k \right\|}$$

This construction has a clear geometric interpretation that we lave to the reader.

Example. The *Legendre polynomials* L_n, $n = 0, 1, 2, \ldots$ are defined by

$$L_n(t) = \frac{1}{n! 2^n} \frac{d^n}{dt^n} \left((t^2 - 1)^n \right).$$

Observe that $L_0 = 1$ and L_n is a polynomial of degree n. We will prove that the functions

$$\frac{\sqrt{2n+1}}{\sqrt{2}} L_n(t) \quad n = 0, 1, 2, \ldots$$

form an orthonormal basis in $L^2[-1, 1]$ obtained from the functions $1, t, t^2, t^3, \ldots$ through the Gramm-Schmidt orthogonalization. To this end it suffices to show that

(1) $\langle L_n, L_m \rangle = 0$ for $n \neq m$,
(2) $\langle L_n, L_n \rangle = 2/(2n+1)$ for $n = 0, 1, 2, \ldots$,
(3) $\langle P, L_n \rangle = 0$ for all polynomials P of degree less than n,
(4) $\langle t^n, L_n \rangle > 0$ for $n = 0, 1, 2, \ldots$

If $n > 1$ and P is a polynomial of degree less than n, then

$$\langle P, L_n \rangle = \frac{(-1)^n}{n! 2^n} \int_{-1}^{1} \frac{d^n P(t)}{dt^n} (t^2 - 1)^n \, dt = 0$$

which proves (3)[7]. Since L_n is a polynomial of degree n it also proves (1). Recall that $(2n)!! = 2 \cdot 4 \cdot 6 \cdot \ldots \cdot (2n) = n! 2^n$ and $(2n-1)!! = 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2n-1)$. Since $(t^2 - 1)^n = t^{2n} + \ldots$ we have

$$\frac{d^n}{dt^n} L_n(t) = \frac{1}{(2n)!!} \frac{d^{2n}}{dt^{2n}} \left((t^2 - 1)^n \right) = \frac{(2n)!}{(2n)!!} = (2n-1)!!$$

and hence

$$\begin{aligned}
\langle L_n, L_n \rangle &= \frac{1}{(2n)!!} \int_{-1}^{1} L_n(t) \frac{d^n}{dt^n} \left((t^2 - 1)^n \right) dt \\
&= \frac{(-1)^n}{(2n)!!} \int_{-1}^{1} (2n-1)!! \, (t^2 - 1)^n \, dt = \frac{2 \cdot (2n-1)!!}{(2n)!!} \int_0^1 (1 - t^2)^n \, dt \\
&= \frac{2 \cdot (2n-1)!!}{(2n)!!} \int_0^{\pi/2} \cos^{2n+1} u \, du = \frac{2}{2n+1}
\end{aligned}$$

[7]We have no boundary terms resulting from the integration by parts, because derivatives of $(t^2 - 1)^n$ of order $\leq n - 1$ vanish at ± 1.

which proves (2). Finally

$$\langle t^n, L_n\rangle = \frac{1}{(2n)!!}\int_{-1}^{1} t^n \frac{d^n}{dt^n}\big((t^2-1)^n\big)\,dt = \frac{n!}{(2n)!!}\int_{-1}^{1}(1-t^2)^n\,dt > 0$$

which proves (4). The proof is complete. □

5.3. **Distance to a subspace.** We will show now how to compute the distance of an element of a Hilbert spaces to a finitely dimensional subspace.

Definition. The *Gramm determinant* of vectors $x_1, x_2, \ldots, x_n$ in an inner product space is

$$G(x_1, x_2, \ldots, x_n) = \det\{\langle x_i, x_j\rangle\}_{i,j=1}^{n}\,.$$

Lemma 5.16. *Let $x_1, x_2, \ldots, x_n$ be vectors in an inner product space H and let $H_{n-1} = \operatorname{span}\{x_1, x_2, \ldots, x_{n-1}\}$. Then*

$$G(x_1, \ldots, x_n) = G(x_1, \ldots, x_{n-1}) \cdot \operatorname{dist}(x_n, H_{n-1})^2\,.$$

Proof. Consider the Gramm matrix

$$\begin{bmatrix} \langle x_1, x_1\rangle & \ldots & \langle x_1, x_{n-1}\rangle & \langle x_1, x_n\rangle \\ \vdots & \ddots & \vdots & \vdots \\ \langle x_{n-1}, x_1\rangle & \ldots & \langle x_{n-1}, x_{n-1}\rangle & \langle x_{n-1}, x_n\rangle \\ \langle x_n, x_1\rangle & \ldots & \langle x_n, x_{n-1}\rangle & \langle x_n, x_n\rangle \end{bmatrix}$$

whose determinant equals $G(x_1, \ldots, x_n)$. Let $w = \lambda_1 x_1 + \lambda_2 x_2 = \ldots + \lambda_{n-1}x_{n-1}$ be the orthogonal projection of x_n onto H_{n-1}. Subtract from the last row the first $(n-1)$ rows multiplied by $\lambda_1, \ldots, \lambda_{n-1}$ respectively. This operation does not change the determinant and in the last row we have

$$[\langle x_n - w, x_1\rangle, \ldots, \langle x_n - w, x_{n-1}\rangle, \langle x_n - w, x_n\rangle] = [0, \ldots, 0, \langle x_n - w, x_n\rangle]$$

because $x_n - w \perp H_{n-1}$, see Theorem 5.2. Now we subtract from the last column the first $(n-1)$ columns multiplied by $\overline{\lambda_1}, \ldots, \overline{\lambda_{n-1}}$ respectively. We obtain the matrix

$$\begin{bmatrix} \langle x_1, x_1\rangle & \ldots & \langle x_1, x_{n-1}\rangle & 0 \\ \vdots & \ddots & \vdots & \vdots \\ \langle x_{n-1}, x_1\rangle & \ldots & \langle x_{n-1}, x_{n-1}\rangle & 0 \\ 0 & \ldots & 0 & \langle x_n - w, x_n\rangle \end{bmatrix}$$

Hence

$$G(x_1, \ldots x_{n-1}, x_n) = G(x_1, \ldots, x_{n-1}) \cdot \langle x_n - w, x_n\rangle\,,$$

so the result follows because

$$\langle x_n - w, x_n\rangle = \langle x_n - w, x_n - w\rangle + \underbrace{\langle x_n - w, w\rangle}_{0} = \operatorname{dist}(x_n, H_{n-1})^2\,.$$

The proof is complete. □

Corollary 5.17. *For any vectors $x_1,\dots,x_n$ in an inner product space $G(x_1,\dots,x_n)\geq 0$ and $G(x_1,\dots,x_n)=0$ if and only if the vectors are linearly dependent.*

Theorem 5.18. *Let $x_1,\dots,x_n$ be linearly independent vectors in an inner product space H and let*

$$H_n = \operatorname{span}\{x_1,\dots,x_n\}\,.$$

Then for any $x\in H$

$$\operatorname{dist}(x,H) = \sqrt{\frac{G(x_1,\dots,x_{n-1},x)}{G(x_1,\dots,x_{n-1})}}\,.$$

As another corollary we obtain that if $x_1,\dots x_n$ are linearly independent vectors in an inner product space, then the volume of the parallelpiped spanned by the vectors $x_1,\dots,x_n$ equals

$$V(x_1,\dots,x_n)=\sqrt{G(x_1,\dots,x_n)}\,.$$

Example. We will prove that if $B_1,B_2,\dots,B_k$ is a family of balls in $\mathbb{R}^n$ and $B_i\neq B_j$ for $i\neq j$, then

$$\det\{|B_i\cap B_j|\}_{i,j=1}^k > 0\,.$$

Indeed, let $f_i=\chi_{B_i}\in L^2(\mathbb{R}^n)$ for $i=1,2,\dots$ The functions $f_1,f_2,\dots f_k$ are linearly independent in $L^2(\mathbb{R}^n)$ and hence

$$G(f_1,\dots,f_k)=\det\{|B_i\cap B_j|\}_{i,j=1}^k>0\,.$$

Observe that the corresponding result is not true is we replace balls by cubes. Why?

We close the section on Hilbert spaces with two elegant applications.

5.4. **Müntz theorem.** As a profound application of the formula for the distance to a subspace we will we will prove

Theorem 5.19 (Müntz)**.** *Let $0<p_1<p_2<\dots$, $\lim_{i\to\infty}p_i=\infty$. Then every continuous function in $C[0,1]$ can be uniformly approximated by functions of the form*

$$\lambda_0+\lambda_1 t^{p_1}+\dots+\lambda_n t^{p_n} \tag{5.6}$$

if and only if

$$\sum_{i=1}^{\infty}\frac{1}{p_i}=\infty\,. \tag{5.7}$$

Proof. A crucial step in the proof is the following lemma.

Lemma 5.20. *Let* $W = \mathrm{span}\,\{t^{p_1}, \dots, t^{p_n}\} \subset L^2[0,1]$*, where* $-1/2 < p_1 < \dots < p_n$*. Then for any* $q \geq 0$

$$\mathrm{dist}_{L^2}(t^q, W) = \sqrt{\frac{1}{2q+1}\prod_{i=1}^n \frac{|q-p_i|}{q+p_i+1}}\,.$$

Before we prove the lemma we will show how to complete the proof of the Müntz theorem. First we will prove an L^2 version of the Müntz theorem.

Theorem 5.21. *Let* $-1/2 < p_1 < p_2 < p_3 < \dots$*,* $\lim_{i\to\infty} p_i = \infty$*. Then the functions* $t^{p_1}, t^{p_2}, t^{p_3}, \dots$ *are linearly dense in* $L^2[0,1]$*, i.e.*

$$\overline{\mathrm{span}\,\{t^{p_1}, t^{p_2}, t^{p_3}, \dots\}} = L^2[0,1] \tag{5.8}$$

if and only if

$$\sum_{i=1}^{\infty}{}' \frac{1}{p_i} = \infty \tag{5.9}$$

where in the sum $\sum'$ *we omit the term with* $p_i = 0$*.*

Proof. Since polynomials are dense in $L^2[0,1]$ a necessary and sufficient condition for (5.8) is that for any integer $q \geq 0$ and any $\varepsilon > 0$ there are coefficients $\lambda_1, \lambda_2, \dots, \lambda_n$ such that

$$\int_0^1 \Big|t^q - \sum_{i=1}^n \lambda_i t^{p_i}\Big|^2 < \varepsilon\,.$$

If $q = p_i$ for some i it is obvious, so we can assume that $q \neq p_i$ for any i. According to Lemma 5.20 the condition is equivalent to

$$\prod_{i=1}^{\infty} \frac{|q-p_i|}{q+p_i+1} = 0$$

or

$$\prod_{p_i > q} \frac{1 - q/p_i}{1 + (1+q)/p_i} = 0\,. \tag{5.10}$$

Indeed, since $p_i \to \infty$ we neglect in the second product only a finite number of nonzero factors which does not affect divergence to 0.

We will also need the following elementary lemma from Advanced Calculus.

Lemma 5.22. *If* $b_i > 0$*, then the product* $\prod_{n=1}^\infty (1+b_n)$ *converges to a finite positive limit if and only if the series* $\sum_{n=1}^\infty b_n$ *converges. If* $0 < b_n < 1$*, then the product* $\prod_{n=1}^\infty (1-b_n)$ *converges to a finite positive limit if and only if the series* $\sum_{n=1}^\infty b_n$ *converges.*

Proof. Since $e^x \geq 1 + x$ we have

$$1 + b_1 + \ldots + b_n \leq (1 + b_1) \ldots (1 + b_n) \leq e^{b_1 + \ldots + b_n}$$

which easily implies the first part of the lemma. If $0 < b_n < 1$, then

$$(1 - b_1) \ldots (1 - b_n) = \frac{1}{(1 + a_1) \ldots (1 + a_n)} \tag{5.11}$$

where $a_i = b_i/(1 - b_i)$. Now the expression on the right hand side of (5.11) converges to a positive limit if and only if $\prod_{n=1}^{\infty}(1 + a_n)$ converges and due to the first part of the lemma it remains to observe that $\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} b_n/(1 - b_n)$ converges if and only if $\sum_{n=1}^{\infty} b_n$ converges. □

It remains to prove that (5.10) is equivalent to (5.9). Assume first that $q > 0$. If (5.9) is satisfied, then

$$\prod_{p_i > q} (1 - q/p_i) = 0, \quad \prod_{p_1 > q} (1 + (1 + q)/p_i) = \infty$$

by Lemma 5.22 and hence (5.10) follows. On the other hand if the sum at (5.9) is finite, then both products

$$\prod_{p_i > q} (1 - q/p_i), \quad \prod_{p_1 > q} (1 + (1 + q)/p_i)$$

have finite and positive limits, so (5.10) is not satisfied.

If $q = 0$, (5.10) reads as

$$\prod_{i=1}^{\infty} \frac{1}{1 + p_i} = 0$$

which according to Lemma 5.22 is equivalent to (5.9). The proof of the L^2 version of the Müntz theorem is complete. □

We are ready now to complete the proof of the Müntz theorem in the continuous case. First observe that the condition (5.7) is necessary. Indeed, density of the functions (5.6) in $C[0,1]$ imply density in $L^2[0,1]$ and hence necessity of (5.7) follows from Theorem 5.21. It remains to prove that the condition (5.7) is sufficient.

Since (5.6) contains all constant functions, it remains to prove that for any integer $m \geq 1$, t^m can be uniformly approximated by functions of the form (5.6). Let i_0 be such that $p_i > 1$ for $i \geq i_0$. Then for any $t \in [0,1]$ we have

$$\begin{aligned} \left| t^m - \sum_{i=i_0}^{n} \mu_i m p_i^{-1}\, t^{p_i} \right| &= m \left| \int_0^t \left(\tau^{m-1} - \sum_{i=i_0}^{n} \mu_i \tau^{p_i - 1} \right) d\tau \right| \\ &\leq m \sqrt{\int_0^1 \left| \tau^{m-1} - \sum_{i=i_0}^{n} \mu_i \tau^{p_i - 1} \right|^2 d\tau}\,. \end{aligned} \tag{5.12}$$

Since $\sum_{i=i_0}^{\infty} 1/(p_i-1)=\infty$ we conclude from Theorem 5.21 that for any $\varepsilon>0$ there are numbers $\mu_{i_0},\ldots,\mu_n$ such that the right hand side of (5.12) is less than ε and hence for $\lambda_0=\lambda_1=\ldots=\lambda_{i_0-1}=0$, $\lambda_i=\mu_i m p_i^{-1}$, $i=i_0,\ldots,n$ we have

$$\sup_{t\in[0,1]}\left|t^m-\big(\lambda_0+\lambda_1 t^{p_1}+\ldots+\lambda_n t^{p_n}\big)\right|<\varepsilon\,.$$

The proof is complete. □

Proof of Lemma 5.20. According to Theorem 5.18

$$\mathrm{dist}_{L^2}(t^q,W)^2=\frac{G(t^q,t^{p_1},\ldots,t^{p_n})}{G(t^{p_1},\ldots,t^{p_n})}\,. \tag{5.13}$$

Since

$$\langle t^\alpha,t^\beta\rangle=\int_0^1 t^{\alpha+\beta}\,dt=\frac{1}{\alpha+\beta+1}$$

we have

$$G(t^{p_1},t^{p_2},\ldots,t^{p_n})=\det\left\{\frac{1}{p_i+p_j+1}\right\}_{i,j=1}^n=\det\left\{\frac{1}{a_i+b_j}\right\}_{i,j=1}^n\,,$$

where $a_i=p_i$, $b_j=p_j+1$ and similarly

$$G(t^q,t^{p_1},\ldots,t^{p_n})=\det\left\{\frac{1}{\alpha_i+\beta_j}\right\}_{i,j=1}^{n+1}\,,$$

where $\alpha_1=q$, $\alpha_i=p_{i-1}$, $i=2,3,\ldots,n+1$, $\beta_1=q+1$, $\beta_j=p_{j-1}+1$, $j=2,3,\ldots,n+1$.

Thus a crucial step is to compute the Cauchy determinant

$$D_m=\det\left\{\frac{1}{a_i+b_j}\right\}_{i,j=1}^m\,.$$

Observe that computing the determinant using the algebraic definition that involves permutations and then taking the common denominator gives

$$D_m=\frac{P_m(a_1,\ldots,a_m,b_1,\ldots,b_m)}{\prod_{i,j=1}^m(a_i+b_j)}\,.$$

where P_m is a polynomial of degree m^2-m.

Exercise. *Prove that if $P(x,y)$ is a polynomial of two variables that vanishes when $x=y$, then $P(x,y)=(x-y)Q(x,y)$, where Q is another polynomial. Generalize this result to polynomials of higher number of variables.*

The polynomial P_m vanishes when $a_i=a_j$ for some $i\neq i$ (because the matrix has identical two rows, so its determinant equals 0) or when $b_i=b_j$

for some $i \neq j$. Therefore

$$P_m = \underbrace{\prod_{1\le i<j\le m}(a_j-a_i)}_{A_m}\ \underbrace{\prod_{1\le i<j\le m}(b_j-b_i)}_{B_m}\ Q(a_1,\dots,a_m,b_1,\dots,b_m)\,.$$

The degree of each of the polynomials A_m and B_m equals $m(m-1)/2$, so the degree of the polynomial A_mB_m is the same as the degree of P_m and hence Q is a constant

$$P_m = \gamma_m A_m B_m\,.$$

We will prove that $\gamma_m = 1$. A direct computation shows that $\gamma_1 = 1$. Multiplying the last row of the matrix in the determinant D_m by a_m, letting $a_m \to \infty$ and then letting $b_m \to \infty$ we obtain

$$\begin{aligned} a_m D_m &= \begin{vmatrix} \frac{1}{a_1+b_1} & \cdots & \frac{1}{a_1+b_{m-1}} & \frac{1}{a_1+b_m} \\ \vdots & \ddots & \vdots & \vdots \\ \frac{a_m}{a_m+b_1} & \cdots & \frac{a_m}{a_m+b_{m-1}} & \frac{a_m}{a_m+b_m} \end{vmatrix} \\ &\longrightarrow \begin{vmatrix} \frac{1}{a_1+b_1} & \cdots & \frac{1}{a_1+b_{m-1}} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ \frac{1}{a_{m-1}+b_1} & \cdots & \frac{1}{a_{m-1}+b_{m-1}} & 0 \\ 1 & \cdots & 1 & 1 \end{vmatrix} = D_{m-1}. \end{aligned}$$

On the other hand the same limiting process gives

$$\begin{aligned} a_m\gamma_m^{-1}D_m &= \frac{a_m\prod_{1\le i<j\le m}(a_j-a_i)(b_j-b_i)}{\prod_{i,j=1}^m(a_i+b_j)} \\ &\longrightarrow \frac{\prod_{1\le i<j\le m-1}(a_j-a_i)(b_j-b_i)}{\prod_{i,j=1}^{m-1}(a_i+b_i)} = \gamma_{m-1}^{-1}D_{m-1}\,, \end{aligned}$$

so $\gamma_m = \gamma_{m-1}$ and by induction $\gamma_m = 1$. We proved that

$$D_m = \frac{\prod_{1\le i<j\le m}(a_j-a_i)(b_j-b_i)}{\prod_{i,j=1}^m(a_i+b_j)}\,.$$

Taking $a_i = p_i$, $b_j = p_j + 1$ we have

$$G(t^{p_1},\dots,t^{p_n}) = \frac{\prod_{1\le i<j\le n}(p_j-p_i)^2}{\prod_{i,j=1}^n(p_i+p_j+1)}$$

and similarly

$$G(t^q,t^{p_1},\dots,t^{p_n}) = \frac{\prod_{i=1}^n(q-p_i)^2\prod_{1\le i<j\le n}(p_i-p_j)^2}{(2q+1)\prod_{i=1}^n(q+p_i+1)^2\prod_{i,j=1}^n(p_i+p_j+1)}\,.$$

The two above formulas together with (5.13) give the result. $\square$

5.5. **Radon-Nikodym-Lebesgue theorem.** As an elegant application of the Riesz representation theorem (Theorem 5.5) we will prove the Radon-Nikodym-Lebesgue theorem.

Recall that if μ and ν are two positive measures on $\mathfrak{M}$, then ν is called *absolutely continuous* with respect to ν if

$$E \in \mathfrak{M},\ \mu(E) = 0 \quad \Rightarrow \quad \nu(E) = 0$$

and we write $\nu \ll \mu$. If there are disjoint sets $A, B \in \mathfrak{M}$ such that μ is concentrated on A and ν is concentrated on B we say that the measures are *mutually singular* and we write $\mu \perp \nu$.

Theorem 5.23 (Radon-Nikodym-Lebesgue)**.** *Let μ be a σ-finite measure on $\mathfrak{M}$ and ν a finite measure on $\mathfrak{M}$. Then*

(a) *There is unique pair of measures ν_a, ν_s on $\mathfrak{M}$ such that*

$$\nu = \nu_a + \nu_s, \quad \nu_a \ll \mu, \quad \nu_s \perp \mu\,.$$

(b) *There is a nonnegative function $h \in L^1(\mu)$ such that*

$$\nu_a(E) = \int_E h\, d\mu \quad \textit{for all } E \in \mathfrak{M}.$$

Proof. Uniqueness of ν_a and ν_s is easy and left as an exercise, so we need to prove the existence ν_a, ν_s and h. We will prove all this at the same time. We need

Lemma 5.24. *If μ is a positive σ-finite measure on a σ-algebra $\mathfrak{M}$ in X, then there is a function $w \in L^1(\mu)$ such that $0 < w(x) < 1$ for every $x \in X$.*

Proof. Since μ is σ-finite it is the union of countably many sets of finite measure $E_n \in \mathfrak{M}$. We may also assume that the sets are pairwise disjoint. Now the function $w(x) = 2^{-n}(1 + \mu(E_n))$ if $x \in E_n$ satisfies the claim. □

Define a new measure $d\varphi = d\nu + w\, d\mu$. It is a finite measure and for every measurable nonnegative function f

$$\int_X f\, d\varphi = \int_X f\, d\nu + \int_X fw\, d\mu\,.$$

The Schwarz inequality for $f \in L^2(\varphi)$ yields

$$\left| \int_X f\, d\nu \right| \leq \int_X |f|\, d\varphi \leq \|f\|_{L^2(\varphi)}\, \varphi(X)^{1/2}\,.$$

Accordingly

$$f \mapsto \int_X f\, d\nu$$

is a bounded functional on $L^2(\varphi)$ and hence the Riesz representation theorem gives the existence of $g \in L^2(\varphi)$ such that

$$\int_X f\,d\nu = \int_X fg\,d\varphi \quad \text{for all } f \in L^2(\varphi)\,. \tag{5.14}$$

For $f = \chi_E$, $\varphi(E) > 0$ we obtain

$$\nu(E) = \int_E g\,d\varphi\,.$$

Hence

$$0 \le \frac{1}{\varphi(E)} \int_E g\,d\varphi = \frac{\nu(E)}{\varphi(E)} \le 1\,.$$

Since E can be chosen arbitrarily we conclude that $g(x) \in [0,1]$, φ-a.e. Therefore we may alternate g on a set of φ-measure zero so that $g(x) \in [0,1]$ for all X without affecting (5.14). We can rewrite (5.14) as

$$\int_X (1-g)f\,d\nu = \int_X fgw\,d\mu\,. \tag{5.15}$$

Let

$$A = \{x :\, 0 \le g(x) < 1\} \quad B = \{x :\, g(x) = 1\}$$

and define the measures ν_a and ν_s by

$$\nu_a(E) = \nu(E \cap A), \quad \nu_s(E) = \nu(E \cap B)$$

for $E \in \mathfrak{M}$. If we take $f = \chi_B$, (5.15) gives

$$0 = \int_B w\,d\mu\,.$$

Since $w(x) > 0$ for all x, we conclude that $\mu(B) = 0$ and hence $\nu_s \perp \mu$.[8] Applying (5.15) to

$$f = (1 + g + \ldots + g^n)\,\chi_E$$

we obtain

$$\int_E (1 - g^{n+1})\,d\nu = \int_E g(1 + g + \ldots + g^n)w\,d\mu\,.$$

If $x \in B$, then $g(x) = 1$ and if $x \in A$, then $g^{n+1}(x)$ decreases monotonically to zero. Hence letting $n \to \infty$ the left hand side converges to $\nu(A \cap E) = \nu_a(E)$. On the right hand side the function that we integrate increases to a measurable function h and the monotone convergence theorem gives

$$\nu_a(E) = \int_E h\,d\mu\,.$$

Hence $\nu_a \ll \mu$. Finally if we take $E = X$ we obtain that $h \in L^1(\mu)$, because $\nu_a(X) < \infty$. □

[8]Indeed, $\mu(B) = 0$ implies that μ is concentrated on $X \setminus B$. Since ν_s is concentrated on B and $B \cap (X \setminus B) = \emptyset$ we obtain that $\mu \perp \nu_s$.

6. Fourier Series

In this section we will present a basic theory of Fourier series. It is one of the most important applications of the theory of Hilbert spaces.

We denote the space of continuous, L^p,...etc functions on $\mathbb{R}$ with the period 1 by $C(S^1)$, $L^p(S^1)$...etc. This notation comes from the fact that functions with period 1 can be identified with functions on the unit circle through the mapping $x \mapsto e^{2\pi i x}$. The space $L^2(S^1)$ is a Hilbert space with respect to the inner product

$$\langle f, g\rangle = \int_0^1 f(x)\overline{g(x)}\, dx\,. \tag{6.1}$$

The $C(S^1)$ and $L^p(S^1)$ norms will be denoted by $\|\cdot\|_\infty$ and $\|\cdot\|_p$ respectively, but the $L^2(S^1)$ norm will be simply denoted by $\|\cdot\|$. As a direct application of Theorem 5.13 we have.

Theorem 6.1. *The functions $e_n(x) = e^{2\pi i n x}$, $n \in \mathbb{Z}$ form an orthonormal basis in $L^2(S^1)$. Hence any $f \in L^2(S^1)$ can be represented as a series*

$$f = \sum_{n=-\infty}^{\infty} \hat{f}(n)\, e_n$$

where

$$\hat{f}(n) = \langle f, e_n\rangle = \int_0^1 f(x) e^{-2\pi i n x}\, dx\,,$$

in the sense that the series converges to f in the L^2 norm, i.e.

$$\Big\| f - \sum_{|n|\leq k} \hat{f}(n)\, e_n \Big\| \to 0 \quad \text{as } k \to \infty.$$

Moreover the Plancherel identity

$$\|f\|_2 = \sum_{n=-\infty}^{\infty} |\hat{f}(n)|^2$$

is satisfied for all $f \in L^2(S^1)$ and the mapping $f \mapsto \{\hat{f}(n)\}_{n\in\mathbb{Z}}$ is an isometry of Hilbert spaces

$$\wedge : L^2(S^1) \to \ell^2(\mathbb{Z})\,.$$

Proof. We only need to prove that the family $\{e_n\}_{n\in\mathbb{Z}}$ is an orthonormal basis in $L^2(S^1)$. It is easy to see that the functions e_n are orthonormal, so we are left with the proof that linear combinations of the e_n's are dense in $L^2(S^1)$ (see Theorem 5.9(b)). Since linear combinations of the functions e_n are exactly trigonometric polynomials of period 1, they are dense in $C(S^1)$ by the Weierstrass theorem, and the claim follows from the density of $C(S^1)$ in $L^2(S^1)$. $\square$

Another orthonormal basis in $L^2(S^1)$ is given by the functions 1, $\sqrt{2}\cos(2\pi nx)$, $\sqrt{2}\sin(2\pi nx)$, $n = 1, 2, 3\ldots$ and Theorem 6.1 has an obvious counterpart in this case as well. However, it is more convenient to work with the basis e_k and we will always use it in what follows.

As a first step we will derive a useful integral representation for the partial sum of the Fourier series. Note that the partial sums of the Fourier series are well defined for $f \in L^1(S^1)$. We have

$$\begin{aligned} s_n(f,x) &= \sum_{|k|\le n} \hat{f}(k)e^{2\pi ikx} = \sum_{|k|\le n} e^{2\pi ikx} \int_0^1 f(y)e^{-2\pi iky}\,dy \\ &= \int_0^1 f(y) \underbrace{\sum_{|k|\le n} e^{2\pi ik(x-y)}}_{D_n(x-y)}\,dy\,, \end{aligned}$$

where

$$D_n(x) = \sum_{|k|\le n} e_k(x) = \sum_{|k|\le n} e^{2\pi ikx}\,.$$

To evaluate the sum $D_n(x)$ note that

$$D_n(x)e^{\pi ix} = \sum_{k=-n}^{n} e^{\pi i(2k+1)x}\,,$$

$$D_n(x)e^{-\pi ix} = \sum_{k=-n}^{n} e^{\pi i(2k-1)x} = \sum_{k=-n-1}^{n-1} e^{\pi i(2k+1)x}$$

and hence

$$D_n(x)\big(e^{\pi ix} - e^{-\pi ix}\big) = e^{\pi i(2n+1)x} - e^{-\pi i(2n+1)x}\,,$$

so

$$D_n(x) = \frac{e^{\pi i(2n+1)x} - e^{-\pi i(2n+1)x}}{e^{\pi ix} - e^{-\pi ix}} = \frac{\sin \pi(2n+1)x}{\sin \pi x}\,.$$

We have $D_n(0) = 2n+1$ and

$$\int_0^1 D_n(x)\,dx = \sum_{|k|\le n} \int_0^1 e_k(x)\,dx = 1\,.$$

Now

$$\begin{aligned}
s_n(f,x) &= \int_0^1 f(y)D_n(x-y)\,dy \\
&\overset{(y-x=t)}{=} \int_{-x}^{1-x} f(x+t)D_n(-t)\,dt \\
&\overset{(D_n(-t)=D_n(t))}{=} \int_{-x}^{1-x} f(x+t)D_n(t)\,dt \\
&\overset{\substack{t\mapsto f(x+t)D_n(t)\\ \text{has period 1}}}{=} \int_{-1/2}^{1/2} f(x+t)D_n(t)\,dt\,.
\end{aligned}$$

We proved

Proposition 6.2. *If $f\in L^1(S^1)$, then*

$$s_n(f,x)=\int_{-1/2}^{1/2} f(x+y)D_n(y)\,dy\,,$$

where the Dirichlet kernel

$$D_n(x)=\frac{\sin\pi(2n+1)x}{\sin\pi x}$$

has the properties

$$D_n(0)=2n+1,\qquad \int_0^1 D_n(x)\,dy=1\,.$$

As a first application we will prove

Theorem 6.3. *For any $1\le m<\infty$ and $f\in C^m(S^1)$ the partial sum*

$$s_n=s_n(f)=\sum_{|k|\le n}\hat f(k)e_k$$

converges uniformly to f as $n\to\infty$. In fact

$$\|f-s_n(f)\|_\infty\le C(m)n^{-m+1/2}\,\|f^{(m)}\|\,.$$

Proof. For $f\in C^m(S^1)$ we have

$$\begin{aligned}
(f^{(m)})^\wedge(n) &= \int_0^1 f^{(m)}(y)e^{-2\pi iny}\,dy \\
&\overset{\text{(by parts)}}{=} (2\pi in)^m\int_0^1 f(y)e^{-2\pi iny}\,dy\,,
\end{aligned}$$

i.e.

$$(f^{(m)})^\wedge(n)=(2\pi in)^m\hat f(n)\,. \tag{6.2}$$

Hence for $n \le n' < \infty$ we can estimate

$$\begin{aligned}
\|s_n - s_{n'}\|_\infty &\le \sum_{|k|>n} |\hat{f}(k)| \\
&= \sum_{|k|>n} |(f^{(m)})^\wedge(k)|(2\pi k)^{-m} \\
&\overset{\text{(Schwarz)}}{\le} \Big(\sum_{|k|>n} \big|(f^{(m)})^\wedge(k)\big|^2\Big)^{1/2} \Big(\sum_{|k|>n} (2\pi k)^{-2m}\Big)^{1/2}.
\end{aligned}$$

Bessel's inequality gives

$$\Big(\sum_{|k|>n} \big|(f^{(m)})^\wedge(k)\big|^2\Big)^{1/2} \le \|f^{(m)}\|$$

and we also have

$$\Big(\sum_{|k|>n} (2\pi k)^{-2m}\Big)^{1/2} \le C(m) n^{-m+1/2},$$

because

$$\sum_{k=n+1}^{\infty} \frac{1}{k^{2m}} \le \int_n^\infty \frac{dx}{x^{2m}} = \frac{n^{-2m+1}}{2m-1}.$$

Accordingly

$$\|s_n - s_{n'}\|_\infty \le C(m) n^{-m+1/2} \|f^{(m)}\|. \tag{6.3}$$

This implies the uniform convergence of the sequence of partial sums s_n. Since $s_n \to f$ in L^2 and f is continuous, $s_n \rightrightarrows f$ uniformly. Now letting $n' \to \infty$ in (6.3) yields

$$\|s_n - f\|_\infty \le C(m) n^{-m+1/2} \|f^{(m)}\|.$$

The proof is complete. □

The above result allows us to characterize smooth functions on S^1 in terms of a rapid decay of the Fourier coefficients.

Theorem 6.4. *Let $f \in L^2(S^1)$. Then $f \in C^\infty(S^1)$ if and only if for every positive integer m, $|n|^m \hat{f}(n) \to 0$ as $|n| \to \infty$.*[9]

Proof. If $f \in C^\infty(S^1)$, then according to (6.2)

$$|n^m \hat{f}(n)| = C(m)|n|^{-1}\big|(f^{(m+1)})^\wedge(n)\big| \le C(m)|n|^{-1}\|f^{(m+1)}\|_1 \to 0$$

[9]More precisely, if $f \in L^2(S^1)$ and $|n|^m \hat{f}(n) \to 0$ as $|n| \to \infty$ for any m, then there is $\tilde{f} \in C^\infty(S^1)$ such that $f = \tilde{f}$ a.e.

as $|n| \to \infty$. To prove the opposite implication observe that for any integer $m \geq 0$, $|n|^{m+2}\hat{f}(n) \to 0$ as $|n| \to 0$ and hence $|n^m \hat{f}(n)| \leq C(m)|n|^{-2}$ for all n. Thus the series of term by term derivatives

$$\sum_{n=-\infty}^{\infty} \hat{f}(n)e_n^{(m)} = (2\pi i)^m \sum_{n=-\infty}^{\infty} n^m \hat{f}(n)e^{2\pi i n x}$$

converges uniformly by the M-test. That, however, implies that the Fourier series of f defines a C^∞ function

$$\tilde{f} = \sum_{n=-\infty}^{\infty} \hat{f}(n)e_n \tag{6.4}$$

with

$$\tilde{f}^{(m)} = \sum_{n=-\infty}^{\infty} \hat{f}(n)e_n^{(m)}$$

for any m. Since the Fourier series at (6.4) converges to f in $L^2(S^1)$, we conclude that $f = \tilde{f}$ a.e. □

The following result is surprisingly difficult and we will not prove it.

Theorem 6.5 (Carleson)**.** *If $f \in L^2(S^1)$, then $s_n(f) \to f$ a.e.*

In particular if $f \in C(S^1)$, then $s_n(f) \to f$ a.e. and it is natural to expect the everywhere convergence. Unfortunately it is not always true, see Section 9.4.

Theorem 6.6 (Fejer)**.** *If $f \in C(S^1)$, then*

$$\frac{s_0 + s_1 + \ldots + s_{n-1}}{n} \rightrightarrows f$$

uniformly on S^1.

Proof. We have

$$\frac{s_0 + s_1 + \ldots + s_{n-1}}{n} = \int_{-1/2}^{1/2} f(x+y) \underbrace{\frac{D_0(y) + \ldots + D_{n-1}(y)}{n}}_{F_n(y)} \, dy\,.$$

The function F_n is called *Fejer kernel.* Clearly

$$\int_{-1/2}^{1/2} F_n = \frac{1}{n}\sum_{k=0}^{n-1}\int_{-1/2}^{1/2} D_k = 1$$

and hence

$$\frac{s_0 + s_1 + \ldots + s_{n-1}}{n} - f = \int_{-1/2}^{1/2} (f(x+y) - f(x))F_n(y)\, dy\,.$$

It suffices to prove that

$$I(x) := \int_{-1/2}^{1/2} |f(x+y) - f(x)||F_n(y)|\, dy \rightrightarrows 0 \quad \text{as } n \to 0.$$

Note that

$$F_n(x) = \frac{1}{n}\sum_{k=0}^{n-1} \frac{\sin \pi(2k+1)x}{\sin \pi x} = \frac{1}{n}\Big(\frac{\sin n\pi x}{\sin \pi x}\Big)^2 \geq 0\,.$$

Indeed, by the formula for the sum of the geometric sequence

$$\sum_{k=0}^{n-1} e^{\pi(2k+1)xi} = e^{\pi x i}\sum_{k=0}^{n-1} e^{2k\pi x i} = \frac{(1-e^{\pi x n i})(1+e^{\pi x n i})}{-2i\sin \pi x}$$

and hence

$$\sum_{k=0}^{n-1} \sin \pi(2k+1)x = \operatorname{im} \sum_{k=0}^{n-1} e^{\pi(2k+1)xi} = \frac{\sin^2 \pi x n}{\sin \pi x}\,.$$

For a small $0 < \delta < 1/2$ we split the integral

$$I(x) = \int_{|y|<\delta} + \int_{|y|\geq\delta} := I_1(x) + I_2(x)\,.$$

We have

$$\begin{aligned} I_1(x) &= \int_{-\delta}^{\delta} |f(x+y) - f(x)| F_n(y)\, dy \\ &\leq \max_{|x|\leq 1/2}\ \max_{|y|\leq\delta} |f(x+y) - f(x)| \underbrace{\int_{|y|<\delta} F_n(y)\, dy}_{<1} \\ &\leq \max_{|y|\leq\delta} \|f_y - f\|_\infty\,, \end{aligned}$$

where $f_y(x) = f(x+y)$. For every $\varepsilon > 0$ we can find $0 < \delta < 1/2$ such that $I_1(x) < \varepsilon/2$ for all x by uniform continuity of f. Now

$$\begin{aligned} I_2(x) &= \int_{|y|\geq\delta} \leq \frac{4}{n}\|f\|_\infty \int_{\delta}^{1/2} \Big(\frac{\sin n\pi y}{\sin \pi y}\Big)^2 dy \\ &\leq \frac{2}{n}(\sin \pi\delta)^{-2}\|f\|_\infty\,, \end{aligned}$$

and to given δ we can find n_0 so large that $I_2(x) < \varepsilon/2$ for $n > n_0$ and all $x \in [0,1]$. The proof is complete $\square$

Fourier coefficients can be defined for $f \in L^1(S^1)$ by the same integral formula. Clearly

$$\wedge : L^1(S^1) \to \ell^\infty(\mathbb{Z})$$

because

$$|\hat{f}(n)| = \Big|\int_0^1 f(x)e^{-2\pi i n x}\, dx\Big| \leq \|f\|_1\,.$$

The following result is a variant of the Fejer theorem.

Theorem 6.7. *If $f \in L^1(S^1)$, then*

$$\Big\| \frac{s_0 + s_1 + \ldots + s_{n-1}}{n} - f \Big\|_1 \to 0 \quad \text{as } n \to \infty. \tag{6.5}$$

In particular the map $^\wedge : L^1(S^1) \to \ell^\infty(\mathbb{Z})$ *is one-to-one.*

Proof. As in the proof of the Fejer theorem we have

$$\frac{s_0 + s_1 + \ldots + s_{n-1}}{n} - f = \int_{-1/2}^{1/2} (f(x+y) - f(x)) F_n(y)\, dy\,.$$

Thus

$$\Big\| \frac{s_0 + s_1 + \ldots + s_{n-1}}{n} - f \Big\|_1 \leq \int_{-1/2}^{1/2} \|f_y - f\|_1\, F_n(y)\, dy := I\,.$$

Recall that by a well known fact from Analysis I, $\|f_y - f\|_1 \to 0$ as $|y| \to 0$ for any $f \in L^1(S^1)$. We write

$$I = \int_{|y|<\delta} + \int_{|y|\geq\delta} := I_1 + I_2\,.$$

For $\varepsilon > 0$ we choose $\delta > 0$ such that $\|f_y - f\|_1 < \varepsilon/2$ whenever $|y| < \delta$. Hence for $|y| < \delta$ we can estimate I_1 by

$$I_1 \leq \frac{\varepsilon}{2} \int_{|y|<\delta} F_n(y)\, dy < \frac{\varepsilon}{2}\,.$$

For the second integral note that $\|f_y - f\|_1 \leq 2\|f\|_1$ and hence

$$\begin{aligned} I_2 &\leq 2\|f\|_1 \int_{|y|\geq\delta} F_n(y)\, dy \\ &= 4\|f\|_1 \int_{\delta}^{1/2} \frac{1}{n} \Big(\frac{\sin n\pi y}{\sin \pi y} \Big)^2 dy \\ &\leq \frac{2}{n} \|f\|_1 (\sin \pi\delta)^{-2} \end{aligned}$$

which is less than $\varepsilon/2$ provided n is sufficiently large. Now it is easy to see that the convergence (6.5) implies that the mapping $^\wedge : L^1(S^1) \to \ell^\infty(\mathbb{Z})$ is one-to-one. The proof is complete. □

Recall that c_0 is a closed subspace of $\ell^\infty(\mathbb{Z})$ consisting of sequences convergent to 0 at $\pm\infty$.

Theorem 6.8 (The Riemann-Lebesgue lemma)**.**

$$^\wedge : L^1(S^1) \to c_0\,,$$

i.e. $|\hat{f}(n)| \to 0$ *as* $|n| \to \infty$.

First proof. Since $e^{\pi i} = -1$ we have

$$\begin{aligned}\hat{f}(n) &= \int_0^1 f(x)e^{-2\pi inx}\,dx = -\int_0^1 f(x)e^{-2\pi inx}e^{\pi i}\,dx \\ &= -\int_0^1 f(x)e^{-2\pi in(x-\frac{1}{2n})}\,dx = -\int_0^1 f\Big(x+\frac{1}{2n}\Big)e^{-2\pi inx}\,dx\,.\end{aligned}$$

Hence

$$\hat{f}(n) = \frac{1}{2}\int_0^1 \Big(f(x) - f\Big(x+\frac{1}{2n}\Big)\Big)\,e^{-2\pi inx}\,dx\,,$$

so

$$|\hat{f}(n)| \le \frac{1}{2}\|f - f_{\frac{1}{2n}}\|_1 \to 0$$

as $|n| \to \infty$. □

Second proof. If $f \in C^1(S^1)$, then

$$(f')^\wedge = 2\pi in\hat{f}(n)$$

and hence

$$\hat{f}(n) = (2\pi in)^{-1}(f')^\wedge(n)\,.$$

Since $\|(f')^\wedge\|_\infty < \infty$ it follows that $|\hat{f}(n)| \to 0$ as $|n| \to \infty$, i.e.

$$\wedge : C^1(S^1) \to c_0\,.$$

Since[10] $\|\hat{f}\|_\infty \le \|f\|_1$ the claim follows from the density of $C^1(S^1)$ in $L^1(S^1)$. □

As mentioned above (cf. Section 9.4), in general, for $f \in C(S^1)$ the Fourier series need not converge to f, however, the following results provides sufficient conditions for the convergence.

Theorem 6.9 (Dini's criterion)**.** *If $f \in L^1(S^1)$ and for some x with $|x| \le 1/2$*

$$\int_{-1/2}^{1/2} \Big|\frac{f(x+y) - f(x)}{y}\Big|\,dy < \infty\,,$$

then

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

Theorem 6.10 (Jordan's criterion)**.** *If f is a function of bounded variation in a neighborhood of x, then*

$$\lim_{n\to\infty} s_n(f,x) = \frac{1}{2}(f(x^+) + f(x^-))$$

where $f(x^\pm)$ denote the right and left hand side limits of f at x.

[10] $\|\cdot\|_\infty$ is the norm in the Banach space c_0.

We will not prove these theorems.

The following result tells us that the rate of convergence of $s_n(f)$ near $x_0 \in S^1$ depends only on properties of f in a neighborhood of x_0.

Theorem 6.11 (Riemann's localization principle)**.** *If $f \in L^1(S^1)$ vanishes in a neighborhood of x_0, then $s_n(f) \rightrightarrows 0$ uniformly in a (perhaps smaller) neighborhood of x_0.*

Remark. If $f = g$ in a neighborhood of x_0, then

$$s_n(f) - s_n(g) = s_n(f-g) \rightrightarrows 0$$

near x_0 and hence $s_n(f)$ is uniformly close to $s_n(g)$ in a neighborhood of x_0. In particular if $s_n(g) \rightrightarrows g$ near x_0, then also $s_n(f) \rightrightarrows f$ near x_0 and as an application we obtain the following result: if $f \in L^1(S^1)$ is C^1 in a neighborhood of x_0, then $s_n(f) \rightrightarrows f$ in a neighborhood of x_0.

Proof. We can assume that $x_0 = 0$. Suppose that $f(x) = 0$ for $|x| \leq \delta$. Then $f(x+y) = 0$ for $|x| \leq \delta/2$ and $|y| \leq \delta/2$. Note that if $|x| \leq \delta/2$, the function

$$y \mapsto \frac{f(x+y)}{\sin \pi y}$$

vanishes for $|y| \leq \delta/2$ and hence has no singularity at $y = 0$. Therefore the function is integrable on S^1 and for $|x| \leq \delta/2$ we have

$$\begin{aligned} s_n(f,x) &= \int_{-1/2}^{1/2} f(x+y)\, \frac{\sin(2n+1)\pi y}{\sin \pi y}\, dy \\ &= \int_{-1/2}^{1/2} \frac{f(x+y)}{\sin \pi y}\, \frac{e^{\pi i y} e^{2n\pi i y} - e^{-\pi i y} e^{-2n\pi i y}}{2i}\, dy \\ &= \frac{1}{2i}\big((Q^+)^\wedge(-n) - (Q^-)^\wedge(n)\big)\,, \end{aligned}$$

where

$$Q^{\pm}(y) = \frac{f(x+y)}{\sin \pi y}\, e^{\pm \pi i y}\,.$$

Now it follows from the Riemann-Lebesgue lemma that $s_n(f,x) \to 0$ as $n \to \infty$. We still need to prove that the convergence is uniform with respect to $x \in [-\delta/2, \delta/2]$.

For $x_1, x_2 \in [-\delta/2, \delta/2]$ we have

$$\begin{aligned}
|s_n(f,x_1) - s_n(f,x_2)| &\leq \int_{-1/2}^{1/2} |f(x_1+y) - f(x_2+y)| \Big| \frac{\sin(2n+1)\pi y}{\sin \pi y} \Big| \, dy \\
&= \int_{\delta/2 \leq |y| \leq 1} |f(x_1+y) - f(x_2+y)| \Big| \frac{\sin(2n+1)\pi y}{\sin \pi y} \Big| \, dy \\
&\leq \big(\sin \frac{\pi\delta}{2} \big)^{-1} \|f_{x_1} - f_{x_2}\|_1 \\
&= \big(\sin \frac{\pi\delta}{2} \big)^{-1} \|f_z - f\|_1 \, ,
\end{aligned}$$

where $z = x_2 - x_1$. This implies that the family $\{s_n(f)\}_{n=1}^\infty$ is equicontinuous on the interval $[-\delta/2, \delta/2]$. Since it converges at every point of the interval equicontinuity implies uniform convergence. □

6.1. **Convolution.** $L^1(S^1)$ is an algebra with respect to the convolution

$$f * g(x) = \int_0^1 f(x-y) g(y) \, dy, \quad \text{for } f, g \in L^1(S^1).$$

Although the function $y \mapsto f(x-y)g(y)$ need not be integrable for every x and hence $f * g(x)$ is not defined for every x, we will prove[11] that $f * g \in L^1(S^1)$ and

$$\|f * g\|_1 \leq \|f\|_1 \|g\|_1.$$

Obviously, if

$$\int_0^1 |f(x-y)g(y)| \, dy < \infty \, , \tag{6.6}$$

then $f * g(x)$ is finite for such x. Fubini's theorem yields

$$\int_0^1 \Big(\int_0^1 |f(x-y)g(y)| \, dy \Big) \, dx = \int_0^1 |f(x-y)| \, dx \int_0^1 |g(y)| \, dy = \|f\|_1 \|g\|_1 \, .$$

Hence (6.6) holds for a.e. x. Moreover

$$\begin{aligned}
\|f * g\|_1 &= \int_0^1 \Big| \int_0^1 f(x-u) g(y) \, dy \Big| \, dx \leq \int_0^1 \int_0^1 |f(x-y)g(y)| \, dy \, dx \\
&= \|f\|_1 \|g\|_1 \, .
\end{aligned}$$

Note that with the convolution notation we have

$$s_n(f,x) = \int_0^1 f(y) D_n(x-y) \, dy = D_n * f(x)$$

and similarly

$$\frac{s_0 + s_1 + \ldots s_{n-1}}{n} = F_n * f(x) \, .$$

Theorem 6.12. *For $f, g \in L^1(S^1)$ we have*

[11] Just like in Analysis I.

(a) $f * g(x) = g * f(x)$
(b) $(f * g) * h(x) = f * (g * h)(x)$

We leave the proof as an easy exercise. □

Theorem 6.13. *For* $f, g \in L^1(S^1)$

$$(f * g)^\wedge(n) = \hat{f}(n)\hat{g}(n)$$

Proof. We have

$$\begin{aligned}
(f * g)^\wedge(n) &= \int_0^1 \Big(\int_0^1 f(x-y)g(y)\, dy \Big) e^{-2\pi i n x}\, dx \\
&= \int_0^1 \Big(\int_0^1 f(x-y)e^{-2\pi i n(x-y)}\, dy \Big) g(y) e^{-2\pi i n y}\, dy \\
&= \hat{f}(n)\hat{g}(n)
\end{aligned}$$

The proof is complete. □

Now we will show several applications of the Furier series.

6.2. **Riemann zeta funciton.**

Theorem 6.14.

$$\zeta(2) = \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}\, .$$

Proof. Let $f(x) = x$ for $x \in [0, 1)$. Then $f \in L^2(S^1)$ and

$$\hat{f}(n) = \int_0^1 x e^{-2\pi i n x}\, dx = \begin{cases} \frac{1}{2} & \text{for } n = 0, \\ -\frac{1}{2\pi i n} & \text{for } n \neq 0. \end{cases}$$

Hence the Plancherel identity yields

$$\frac{1}{3} = \int_0^1 x^2\, dx = \|f\|_2^2 = \|\hat{f}\|_2^2 = \frac{1}{4} + \sum_{n\neq 0} \Big(\frac{1}{2\pi n}\Big)^2 = \frac{1}{4} + \frac{1}{2\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2}$$

and the claim easily follows. □

6.3. **Writinger's inequality.**

Theorem 6.15 (Writinger's inequality)**.** *If* $f \in C^1([a, b])$*,* $f(a) = f(b) = 0$*, then*

$$\int_a^b |f|^2 \leq \frac{(b-a)^2}{\pi^2} \int_a^b |f'|^2$$

and the constant $(b-a)^2/\pi^2$ *cannot be replaced by any smaller number.*

Proof. It is enough to prove the theorem for $a = 0$, $b = 1/2$ as the general case follows from a linear change of variables, i.e. it suffices to prove the inequality

$$\int_0^{1/2} |f|^2 \leq \frac{1}{4\pi^2} \int_0^{1/2} |f'|^2 .$$

The function f can be extended from $[0, 1/2]$ to $[-1/2, 1/2]$ as an odd function in $C^1(S^1)$. By the oddness of f we have

$$\hat{f}(0) = \int_{-1/2}^{1/2} f = 0$$

and hence the Planchelel identity yields

$$\begin{aligned} \int_{-1/2}^{1/2} |f'|^2 &= \sum_{n=-\infty}^{\infty} |(f')^\wedge(n)|^2 = \sum_{n=-\infty}^{\infty} |2\pi in\hat{f}(n)|^2 \\ &\geq 4\pi^2 \sum_{n\neq 0} |\hat{f}(n)|^2 = 4\pi^2 \int_{-1/2}^{1/2} |f|^2 . \end{aligned}$$

The equality is achieved for the function $f(x) = 2i \sin 2\pi x = e^{2\pi ix} - e^{-2\pi ix}$.
□

6.4. **The isoperimetric problem.**

Theorem 6.16 (Isoperimetric theorem)**.** *Among all Jordan curves of fixed length, the one that encloses the largest area is the circle. All other curves enclose smaller area.*

Proof. We can assume that the fixed length of the Jordan curve is 1. Denoting the enclosed area by A we can write the theorem in the form of the inequality

$$A \leq \frac{1}{4\pi} \tag{6.7}$$

with the equality if and only if the curve is a circle. We will assume that the Jordan curve $\gamma(t) = (x(t), y(t))$ is of class $C^1(S^1)$, i.e. γ is

closed:

$$x(0) = x(1), \quad y(0) = y(1);$$

smooth:

$$x, y \in C^1(S^1), \quad \gamma(t) \neq 0 \text{ for all } t;$$

Jordan:

$$\gamma(t_1) \neq \gamma(t_2) \quad \text{ for } t_1 \neq t_2,\ t_1, t_2 \in [0, 1);$$

unit length:

$$\ell = \int_0^1 \left(\dot{x}(t)^2 + \dot{y}(t)^2\right)^{1/2} dt = 1 .$$

Since every *rectifiable curve* (i.e. of finite length) admits an *arc-length parametrization* (i.e. of speed 1) we can assume that

$$\dot{x}(t)^2 + \dot{y}(t)^2 = 1 \quad \text{for all } t.$$

As a consequence of the Plancherel theorem we have

$$\begin{aligned} 1 &= \int_0^1 |\dot{x}(t)|^2 + |\dot{y}(t)|^2\, dt = \sum_{n=-\infty}^{\infty} |\hat{\dot{x}}(n)|^2 + |\hat{\dot{y}}(n)|^2 \\ &= 4\pi^2 \sum_{n=-\infty}^{\infty} n^2(|\hat{x}(n)|^2 + |\hat{y}(n)|^2) \end{aligned}$$

and hence

$$\sum_{n=-\infty}^{\infty} n^2(|\hat{x}(n)|^2 + |\hat{y}(n)|^2) = (4\pi^2)^{-1}\,.$$

It easily follow from the Green theorem that

$$\begin{aligned} A &= \int_\gamma x\, dy = \int_0^1 x(t)\dot{y}(t)\, dt \overset{\dot{y}\in\mathbb{R}}{=} \int_0^1 x(t)\overline{\dot{y}(t)}\, dt \\ &= \sum_{n=-\infty}^{\infty} \hat{x}(n)\overline{\hat{\dot{y}}(n)} \overset{A\in\mathbb{R}}{=} \operatorname{re} \sum_{n=-\infty}^{\infty} \hat{x}(n)\overline{\hat{\dot{y}}(n)} \\ &= \pi \sum_{n=-\infty}^{\infty} n\cdot 2\operatorname{re}\hat{x}(n)\overline{i\hat{y}(n)}\,. \end{aligned}$$

Hence

$$\sum_{n=-\infty}^{\infty} n\cdot 2\operatorname{re}\hat{x}(n)\overline{i\hat{y}(n)} = A\pi^{-1}\,.$$

Therefore

$$\begin{aligned} (4\pi^2)^{-1} - A\pi^{-1} &= \sum_{n=-\infty}^{\infty} n^2(|\hat{x}(n)|^2 + |\hat{y}(n)|^2) - \sum_{n=-\infty}^{\infty} n\cdot 2\operatorname{re}\hat{x}(n)\overline{i\hat{y}(n)} \\ &= \sum_{n=-\infty}^{\infty} (n^2 - |n|)(|\hat{x}(n)|^2 + |\hat{y}(n)|^2) \\ &\quad + \sum_{n=-\infty}^{\infty} |n|(|\hat{x}(n)|^2 - \operatorname{sgn}(n) 2\operatorname{re}\hat{x}(n)\overline{i\hat{y}(n)} + |\hat{y}(n)|^2) \\ &= \sum_{n=-\infty}^{\infty} (n^2 - |n|)(|\hat{x}(n)|^2 + |\hat{y}(n)|^2) + \sum_{n=-\infty}^{\infty} |n|\, |\hat{x}(n) - \operatorname{sgn}(n) i\hat{y}(n)|^2 \qquad (6.8) \end{aligned}$$

because

$$|a \pm b|^2 = |a|^2 \pm 2\operatorname{re} a\overline{b} + |b|^2\,.$$

Since the right hand side of (6.8) is nonnegative we conclude that

$$(4\pi^2)^{-1} - A\pi^{-1} \geq 0, \quad \text{so} \quad A \leq \frac{1}{4\pi}$$

which is the isoperimetric inequality (6.7). The equality in the isoperimetric inequality holds if and only if the right hand side of (6.8) equals 0, so

$$\hat{x}(n) = \hat{y}(n) = 0 \quad \text{for } |n| \geq 2 \tag{6.9}$$

and

$$|\hat{x}(n) - \operatorname{sgn}(n) i \hat{y}(n)|^2 = 0 \quad \text{for } |n| = 1. \tag{6.10}$$

Now (6.9) yields

$$x(t) = \hat{x}(-1)e^{-2\pi i t} + \hat{x}(0) + \hat{x}(1)e^{2\pi i t}\,,$$
$$y(t) = \hat{y}(-1)e^{-2\pi i t} + \hat{y}(0) + \hat{y}(1)e^{2\pi i t}\,.$$

Hence

$$(x(t) - \hat{x}(0))^2 + (y(t) - \hat{y}(0))^2 =$$
$$(\underbrace{\hat{x}(-1)^2 + \hat{y}(-1)^2}_{0 \text{ by } (6.10)})e^{-4\pi i t} + 2\hat{x}(-1)\hat{x}(1) + 2\hat{y}(-1)\hat{y}(1) + (\underbrace{\hat{x}(1)^2 + \hat{y}(1)^2}_{0 \text{ by } (6.10)})e^{4\pi i t}\,.$$

Therefore

$$(x(t) - \hat{x}(0))^2 + (y(t) - \hat{y}(0))^2 = \text{const.}$$

The proof is complete. □

6.5. **Equidistribution of arithmetic sequences.** For a number $0 < \gamma < 1$ and $x \in [0, 1)$ define[12]

$$x_n = x + n\gamma - [x + n\gamma] \in [0, 1)\,.$$

If we identify $[0, 1)$ with S^1 via the exponential mapping

$$[0, 1) \ni t \mapsto e^{2\pi i t} \in S^1\,,$$

then x_n is identified with a point $e^{2\pi i(x+n\gamma)}$ on the circle and hence $x \mapsto x_n$ is a rotation of S^1 by the angle $2\pi n\gamma$.

Theorem 6.17 (Weyl)**.** *If $0 < \gamma < 1$ is irrational*[13] *and $f \in C(S^1)$, then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(x_k) = \int_0^1 f(x)\, dx\,. \tag{6.11}$$

Proof. First we will prove the theorem in the case in which

$$f(x) = e_m(x) = e^{2\pi i m x} \quad \text{for some } m \in \mathbb{Z}.$$

If $m = 0$, then both sides of (6.11) are equal 1. If $m \neq 0$, then

$$\int_0^1 e_m(x)\, dx = 0$$

[12] $[x]$ stands for the largest integer $\leq x$.
[13] The result is not true if γ is rational (exercise).

and

$$\frac{1}{n}\sum_{k=0}^{n-1} e_m(x_k) = \frac{1}{n}\sum_{k=0}^{n-1} e^{2\pi i m(x+k\gamma)} \overset{\binom{\text{geom.}}{\text{sum}}}{=} \frac{1}{n}\frac{e^{2\pi i m x}(1-e^{2\pi i m\gamma n})}{1-e^{2\pi i m\gamma}} \to 0 \quad \text{as } n\to\infty\,.$$

(Note that sice γ is irrational the denominator is nonzero.) Thus (6.11) is satisfied. Next (6.11) is satisfied by trigonometric polynomials which are finite sums of the form

$$f(x) = \sum_{|m|\leq k} a_m e_m(x) \tag{6.12}$$

For a general $f\in C(S^1)$ and $\varepsilon>0$ we can find a trigonometric polynomial f_ε of the form (6.12) such that[14]

$$\|f_\varepsilon - f\|_\infty < \varepsilon/3\,.$$

Since the function f_ε satisfies (6.11) for sufficiently large n we have

$$\Big|\frac{1}{n}\sum_{k=0}^{n-1} f(x_k) - \int_0^1 f\Big| \leq \Big|\frac{1}{n}\sum_{k=0}^{n-1} f_\varepsilon(x_k) - \int_0^1 f_\varepsilon\Big| + \underbrace{\Big|\frac{1}{n}\sum_{k=0}^{n-1}(f_\varepsilon(x_k)-f(x_k))\Big|}_{<\varepsilon/3} + \underbrace{\Big|\int_0^1 (f_\varepsilon - f)\Big|}_{<\varepsilon/3}$$

$$\leq \Big|\frac{1}{n}\sum_{k=0}^{n-1} f_\varepsilon(x_k) - \int_0^1 f_\varepsilon\Big| + \frac{2\varepsilon}{3} < \varepsilon\,.$$

The proof is complete. $\square$

Corollary 6.18. *If $0<\gamma<1$ is irrational and $0\leq a<b\leq 1$, then*

$$\lim_{n\to\infty}\frac{\#\{k<n:\ a\leq x_k\leq b\}}{n} = b-a\,.$$

Proof. Let $f^\pm$ be continuous functions that approximate the characteristic function of the interval $[a,b]$ is the following sense[15]

$$f^- \leq \chi_{[a,b]} \leq f^+\,,$$

$$b-a-\frac{\varepsilon}{2} < \int_0^1 f^- \leq \int_0^1 f^+ < b-a+\frac{\varepsilon}{2}\,.$$

Since

$$\frac{\#\{k<n:\ a\leq x_k\leq b\}}{n} = \frac{1}{n}\sum_{k=0}^{n-1}\chi_{[a,b]}(x_k)$$

[14]By the Stone-Weierstrass theorem or simply by the Fejer theorem.

[15]A picture will help to see that such functions exist.

we conclude from the previous theorem that

$$\begin{aligned}
(b-a)-\varepsilon \quad &< \quad \int_0^1 f^-(x)\,dx - \frac{\varepsilon}{2} \\
&\overset{\text{large } n}{<} \quad \frac{1}{n}\sum_{k=0}^{n-1} f^-(x_k) \\
&\leq \quad \frac{\#\{k<n:\, a\leq x_k \leq b\}}{n} \\
&\leq \quad \frac{1}{n}\sum_{k=0}^{n-1} f^+(x_k) \\
&\overset{\text{large } n}{<} \quad \int_0^1 f^+(x)\,dx + \frac{\varepsilon}{2} < (b-a)+\varepsilon\,.
\end{aligned}$$

Hence

$$\left|\frac{\#\{k<n:\, a\leq x_k \leq b\}}{n} - (b-a)\right| < \varepsilon$$

for all sufficiently large n. □

EXERCISE. *The sequence* 2^n *starts with* $\mathbf{2}, \mathbf{4}, \mathbf{8}, \mathbf{1}6, \mathbf{3}2, \mathbf{6}4, \mathbf{1}28, \mathbf{2}56, \ldots$

(a) *Investigate how many times* 7 *and* 8 *will appear as a first digit of the decimal representation of* 2^n *for* $n \leq 45$.
(b) *The digit* c *appears as a first digit of* 2^n *with the frequency*

$$\lim_{n\to\infty} \frac{\#\{k<n : 2^k = \mathbf{c}\ldots\}}{n}\,.$$

Prove that this limit exists for any $c \in \{1,2,\ldots,9\}$ *and show that* 7 *appears more often than* 8.

7. Spherical harmonics

The theory of Fourier series describes functions defined on S^1. It is very tempting to develop an analog of the theory of Fourier series for functions defined on the sphere S^{n-1} in $\mathbb{R}^n$, and this is what we will do now.

Trigonometric functions 1, $\sqrt{2}\cos(2\pi nt)$, $\sqrt{2}\sin(2\pi nt)$, $n=1,2,\ldots$ form an orthonormal basis in $L^2(S^1)$. Our purpose now is to find an orthonormal basis in $L^2(S^{n-1})$, the space of square integrable functions on a sphere in $\mathbb{R}^n$. The trigonometric functions listed above are defined on $[0,1]$ and they correspond to functions on S^1 through the parametrization $[0,1]\ni t\mapsto e^{2\pi it}\in S^1$. The corresponding inner product in S^1 is

$$\langle f,g\rangle = \frac{1}{2\pi}\int_{S^1} f\overline{g}\,d\sigma$$

where we integrate with respect to the length element on S^1. The factor $1/(2\pi)$ comes from the fact that in the definition (6.1) of the inner product we integrate over the interval of length 1, but the length of the circle is 2π. However, until the end of the section we will consider another inner product in $L^2(S^1)$

$$\langle f,g\rangle = \int_{S^1} f\overline{g}\, d\sigma$$

and hence the orthonormal basis in $L^2(S^1)$ corresponds to functions $1/\sqrt{2\pi}$, $\pi^{-1/2}\cos(2\pi nt)$, $\pi^{-1/2}\sin(2\pi nt)$, $n=1,2,\ldots$ through the parametrization $[0,1]\ni t\mapsto e^{2\pi t}\in S^1$.

More precisely if $z=e^{2\pi it}\in S^1$, then

$$\frac{1}{\sqrt{\pi}}\, z^n = \frac{1}{\sqrt{\pi}}\, \cos 2\pi nt + \frac{i}{\sqrt{\pi}}\, \sin 2\pi nt\, .$$

Hence the orthonormal basis on S^1 consists of functions $(2\pi)^{-1/2}$, $\pi^{-1/2}\mathrm{re}\, z^n$, $\pi^{-1/2}\mathrm{im}\, z^n$.

The real and imaginary parts of z^n are polynomials in two variables (x,y) that are harmonic. Moreover they are homogeneous of degree n, i.e. they are of the form

$$P(x,y) = \sum_{k=0}^{n} a_k x^k y^{n-k}\, .$$

Therefore it should not come as a surprise that we will search for an orthonormal basis in $L^2(S^{n-1})$ among homogeneous harmonic polynomials.

We equip the space $L^2(S^{n-1})$ with the inner product

$$\langle f,g\rangle = \int_{S^{n-1}} f(x)\overline{g(x)}\, d\sigma(x)\, .$$

Let $\mathcal{P}_k$ be the linear space of homogeneous polynomials P in $\mathbb{R}^n$ of degree k, i.e.

$$P(x) = \sum_{|\alpha|=k} a_\alpha x^\alpha$$

where we use the multiindex notation.

Let $\mathcal{H}_k$ be the linear subspace of $\mathcal{P}_k$ consisting of harmonic polynomials, i.e. $\Delta P = 0$. The elements of $\mathcal{H}_k$ are called *solid spherical harmonics of degree k*.

By H_k we denote the liner space of restrictions of solid spherical harmonics of degree k to S^{n-1}, so H_k is a finitely dimensional linear subspace of $L^2(S^{n-1})$. The elements of H_k are called *surface spherical harmonics of degree k*.

Lemma 7.1. *The finitely dimensional subspaces $\{H_k\}_{k=0}^{\infty}$ of $L^2(S^{n-1})$ are mutually orthogonal.*

Proof. Let $\tilde{P} \in H_k$ and $\tilde{Q} \in H_l$, $k \neq l$ be restictions of $P \in \mathcal{H}_k$ and $Q \in \mathcal{H}_l$ to the unit sphere. Since $P(rx) = r^k P(x)$ it easily follows that

$$\frac{\partial P}{\partial \nu} = \frac{d}{dr}|_{r=1}\, P(rx) = kr^{k-1}P(x)|_{r=1} = kP(x)\,,$$

where ν is the outward normal to S^{n-1}. Similarly $\partial Q/\partial\nu = lQ(x)$. Hence Green's formula yields

$$\begin{aligned} (k-l)\langle \tilde{P}, \tilde{Q}\rangle &= (k-l)\int_{S^{n-1}} P\overline{Q}\, d\sigma = \int_{S^{n-1}} \Big(\overline{Q}\,\frac{\partial P}{\partial \nu} - P\,\frac{\partial \overline{Q}}{\partial \nu}\Big) \\ &= \int_{|x|\leq 1} (\overline{Q}\Delta P - P\Delta Q)\, dx = 0 \end{aligned}$$

by harmonicity of P and Q. □

Theorem 7.2. *Every polynomial $P \in \mathcal{P}_k$ can be uniquely represented as*

$$P(x) = P_0(x) + |x|^2 P_1(x) + \ldots + |x|^{2l} P_l(x)$$

where $P_j \in \mathcal{H}_{k-2j}$, $j = 0, 1, 2, \ldots, l$.

Proof. It suffices to prove that every $P \in \mathcal{P}_k$ can be uniquely represented as

$$P = Q_1 + |x|^2 Q_2, \quad Q_1 \in \mathcal{H}_k,\, Q_2 \in \mathcal{P}_{k-2}\ . \tag{7.1}$$

The result will follow then from a routine inductive argument. With each polynomial

$$P(x) = \sum_{|\alpha|=k} a_\alpha x^\alpha \in \mathcal{P}_k$$

we associate a differential operator

$$P\left(\frac{\partial}{\partial x}\right) = \sum_{|\alpha|=k} a_\alpha \frac{\partial^{|\alpha|}}{\partial x^\alpha}$$

and we define a new inner product in $\mathcal{P}_k$ as follows

$$\langle P, Q\rangle = P\left(\frac{\partial}{\partial x}\right)\overline{Q} = \sum_{|\alpha|=k} a_\alpha \frac{\partial^{|\alpha|}}{\partial x^\alpha}\Big(\sum_{|\beta|=k} \overline{b_\beta} x^\beta\Big) = \sum_{|\alpha|=k} a_\alpha \overline{b_\alpha}\alpha! \tag{7.2}$$

where the last equality follows from the fact that

$$\frac{\partial^{|\alpha|}x^\beta}{\partial x^\alpha} = 0 \text{ if } \alpha \neq \beta\,, \quad \frac{\partial^{|\alpha|}x^\alpha}{\partial x^\alpha} = \alpha!$$

Note that the right hand side of (7.2) proves that this is a true inner product. Let

$$|x|^2\mathcal{P}_{k-2} = \big\{|x|^2 P(x) : \ P \in \mathcal{P}_{k-2}\big\} \subset \mathcal{P}_k\,.$$

It suffices to prove that the orthogonal complement of the subspace $|x|^2\mathcal{P}_{k-2}$ is $\mathcal{H}_k$, i.e.

$$\mathcal{P}_k = \mathcal{H}_k \oplus |x|^2\mathcal{P}_{k-2} \tag{7.3}$$

because it readily implies unique representation (7.1). To prove (7.3) observe that $P_1 \in \mathcal{P}_k$ is in the orthogonal complement of $|x|^2\mathcal{P}_{k-2}$ if and only if for every $P_2 \in \mathcal{P}_{k-2}$

$$\langle |x|^2 P_2, P_1\rangle = 0\,.$$

We have

$$0 = \langle |x|^2 P_2, P_1\rangle = \Delta\Big(P_2\left(\frac{\partial}{\partial x}\right)\overline{P_1}\Big) = P_2\left(\frac{\partial}{\partial x}\right)\Delta\overline{P_1} = \langle P_2, \Delta P_1\rangle$$

for every $P_2 \in P_{k-2}$, where in the right hand side we have the inner product in $\mathcal{P}_{k-2}$. This is, however, true if and only if $\Delta P_1 = 0$, i.e. $P_1 \in \mathcal{H}_k$. □

Corollary 7.3. *The restriction of any polynomial to the unit sphere S^{n-1} is a finite sum of surface spherical harmonics.*

Proof. Any polynomial is a sum of homogeneous polynomials of possibly different degrees. If P is homogeneous, then the representation from Theorem 7.2 gives

$$P(x) = P_0(x) + P_1(x) + \ldots + P_l(x), \quad \text{for } x \in S^{n-1},$$

where P_i are surface spherical harmonics. □

By the Stone-Weierstrass theorem restrictions of polynomials to S^{n-1} are dense in $C(S^{n-1})$ and hence are dense in $L^2(S^{n-1})$. Therefore linear combinations of surface spherical harmonics are dense in $L^2(S^{n-1})$, see Corollary 7.3. Now Theorem 5.14 and Lemma 7.1 give

Theorem 7.4. *For every $f \in L^2(S^{n-1})$ there are unique elements $Y_k \in H_k$, $k = 0, 1, 2, \ldots$ such that*

$$f = \sum_{k=1}^{\infty} Y_k$$

in the sense of convergence in $L^2(S^{n-1})$. Moreover

$$\|f\|^2 = \sum_{k=1}^{\infty} \|Y_k\|^2\,.$$

Accordingly,

$$L^2(S^{n-1}) = H_0 \oplus H_1 \oplus H_2 \oplus \ldots$$

In particular if $\{Y_1^k, \ldots, Y_{a_k}^k\}$ is an orthonormal basis in H_k, then $\{Y_i^k\}_{i,k}$ is an orthonormal basis in $L^2(S^{n-1})$ consisting of surface spherical harmonics.

On every Riemannian manifold M, in particular on every smooth submanifold of the Euclidean space, one can define a natural Laplace operator

Δ_M called the *Lapalce-Beltrami* operator. The general construction is quite involved, but in the case of the sphere it can be easily done as follows.

DEFINITION. If $f \in C^2(S^{n-1})$, then we extend f to a neighborhood of S^{n-1} as a 0-homogeneous function, i.e. $\tilde{f}(x) = f(x/|x|)$ and we define the *Laplace-Beltrami* operator on S^{n-1} (*spherical Lapalcean*) as follows

$$\Delta_S f = \Delta \tilde{f}|_{S^{n-1}}.$$

We will prove now that the surface spherical harmonics are eigenfunctions of the spherical Laplacean Δ_S.

Theorem 7.5. *If $Y \in H_k$, then*

$$\Delta_S Y(x) = -k(k+n-2)Y(x)\,.$$

Proof. If Y is a restriction to the sphere of $P \in \mathcal{H}_k$, then P is k-homogeneous and hence the 0-homogeneous extension of Y is $\tilde{Y}(x) = |x|^{-k}P(x)$. Since P is harmonic we have[16]

$$\Delta\big(|x|^{-k}P(x)\big) = \big(\Delta|x|^{-k}\big)\,P(x) + 2\nabla|x|^{-k}\cdot\nabla P(x) + |x|^{-k}\underbrace{\Delta P(x)}_{0}\,.$$

Because $|x|^{-k} = (x_1^2 + \ldots + x_n^2)^{-k/2}$ an easy computation shows that

$$\nabla|x|^{-k} = -k|x|^{-(k+2)}\langle x_1,\ldots,x_n\rangle\,,$$

$$\Delta|x|^{-k} = \operatorname{div}\nabla|x|^{-k} = -k(n-k-2)|x|^{-(k+2)}\,.$$

Hence

$$\begin{aligned}
\Delta_S Y(x) &= \Delta\big(|x|^{-k}P(x)\big)|_{|x|=1} \\
&= -k(n-k-2)P(x) - 2k\underbrace{\sum_{i=1}^n x_i\,\frac{\partial P}{\partial x_i}\,(x)}_{=\,kP(x)} \\
&= -k(k+n-2)P(x)|_{|x|=1} \\
&= -k(k+n-2)Y(x)\,.
\end{aligned}$$

We used here Euler's theorem which asserts that a k-homogeneous function f satisfies $\sum_{i=1}^n x_i\partial f(x)/\partial x_i = kf(x)$. The proof is complete. □

According to this theorem H_k is the eigenspace of the Laplace-Beltrami operator Δ_S corresponding to the eigenvalue $-k(k+n-2)$. Moreover eigenspaces corresponding to different eigenvalues are mutually orthogonal. Therefore there is an orthonormal basis of $L^2(S^{n-1})$ consisting of smooth eigenfunctions of Δ_S (surface spherical harmonics). Analogous result holds also for the Laplace-Beltrami operator on any smooth compact Riemannian

[16]We use here $\Delta(fg) = (\Delta f)g + 2\nabla f\cdot\nabla g + f\Delta g$.

manifold, but the proof is to complicated to be presented here[17]. As we will see later results of this type are very general: existence of an orthonormal basis in a Hilbert space consisting of eigenfunctions is a general consequence of the spectral theorem.

We proved in Analysis II that $f \in C^\infty(S^1)$ if and only if $n^p \hat{f}(n) \to 0$ as $|n| \to \infty$ for every $p < \infty$. As a striking application of Theorem 7.5 we will generalize this result to the case of functions on S^{n-1}.

Theorem 7.6. *Let $f \in L^2(S^{n-1})$ has the spherical harmonics expansion*

$$f = \sum_{k=0}^{\infty} Y_k, \quad Y_k \in H_k\,. \tag{7.4}$$

Then $f \in C^\infty(S^{n-1})$ (possibly after correction on a set of measure zero) if and only if for every $N > 0$

$$\int_{S^{n-1}} |Y_k|^2\, d\sigma = O(k^{-N}) \quad \text{as } k \to \infty\ . \tag{7.5}$$

Proof. We will need the following lemma.[18]

Lemma 7.7. *If $f, g \in C^2(S^{n-1})$, then*

$$\int_{S^{n-1}} f \Delta_S\, g\, d\sigma = \int_{S^{n-1}} g \Delta_S f\, d\sigma\,.$$

Proof. Let $\tilde{f}$ and $\tilde{g}$ be 0-homogeneous extensions of f and g to the annulus

$$A_\varepsilon = B^n(0, 1+\varepsilon) \setminus B^n(0, 1-\varepsilon)\,.$$

Then the classical Green formula gives

$$\int_{A_\varepsilon} \big(\tilde{f}\Delta\tilde{g} - \tilde{g}\Delta\tilde{f}\big)\, dx = \int_{\partial A_\varepsilon} \Big(\tilde{f}\,\frac{\partial\tilde{g}}{\partial\nu} - \tilde{g}\,\frac{\partial\tilde{f}}{\partial\nu}\Big)\, d\sigma = 0\,,$$

because $\tilde{f}$ and $\tilde{g}$ are constant along the direction of ν and hence $\partial\tilde{g}/\partial\nu = \partial\tilde{f}/\partial\nu = 0$. Finally the fundamental theorem of calculus gives

$$\int_{S^{n-1}} \big(f\Delta_S\, g - g\Delta_S f\big)\, d\sigma = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \int_{A_\varepsilon} \big(\tilde{f}\Delta\tilde{g} - \tilde{g}\Delta\tilde{f}\big)\, dx = 0$$

and the lemma follows. □

Suppose $f \in C^\infty(S^{n-1})$. We will prove (7.5). If we rewrite (7.4) with normalized spherical harmonics

$$f = \sum_{k=0}^{\infty} a_k Y_k^0, \quad \text{where } Y_k^0 = Y_k/\|Y_k\|_{L^2}\,,$$

[17] F.W. Warner, *Foundations of differentiable manifolds and Lie groups*, Chapter 6.

[18] The same lemma is true for the Laplace-Beltrami operator on any compact Riemannian manifold without boundary, see F.W. Warner, *Foundations of differentiable manifolds and Lie groups.*

then (7.5) is equivalent to $|a_k| = O(k^{-N/2})$ as $k \to \infty$. Applying Lemma 7.7 m times we have

$$\begin{aligned} \big\|(\Delta_S)^m f\big\|_{L^2} &\geq \int_{S^{n-1}} (\Delta_S)^m f \cdot \overline{Y_k^0}\, d\sigma = \int_{S^{n-1}} f (\Delta_S)^m \overline{Y_k^0}\, d\sigma \\ &= (-k(k+n-2))^m \underbrace{\int_{S^{n-1}} f\, \overline{Y_k^0}\, d\sigma}_{a_k}, \end{aligned}$$

so

$$|a_k| \leq C(k(k+n-2))^{-m} = O(k^{-2m})$$

and hence (7.5) follows because m can be arbitrarily large.

Suppose now that $f \in L^2(S^{n-1})$ has the expansion (7.4) satisfying (7.5). It remains to prove that the series at (7.4) converges uniformly to a smooth function on the sphere.

Recall that if the functions $g_i \in C^\infty(\Omega)$ are such that for every multiindex $\sum_{i=1}^\infty \|D^\alpha g_i\|_\infty < \infty$, then $\sum_{i=1}^\infty D^\alpha g_i$ converges uniformly on Ω, and hence $g = \sum_{i=1}^\infty g_i \in C^\infty(\Omega)$. Thus a good estimate of derivatives of any order of Y_k will imply smoothness of f. Actually if $\tilde{Y}_k = Y_k(x/|x|)$ in a neighborhood of the sphere, it suffices to prove the estimate[19]

$$\sup_{|x|=1} \big|D^\alpha \tilde{Y}_k(x)\big| \leq C_\alpha k^{n/2+|\alpha|} \|Y_k\|_{L^2}\,. \tag{7.6}$$

Since $P_k = |x|^k \tilde{Y}_k \in \mathcal{H}_k$ is harmonic, the estimate will follow from suitable estimates for harmonic functions.

Recall that harmonic functions have the mean value property, i.e. if u is harmonic on $\mathbb{R}^n$, then

$$u(x) = \fint_{S^{n-1}(x,r)} u\, d\sigma$$

for every $x \in \mathbb{R}^n$ and every $r > 0$. This easily implies that if $\varphi \in C_0^\infty(B^n(0,1))$ is radial (i.e. constant on spheres centered at 0), $\int_{\mathbb{R}^n} \varphi(x)\, dx = 1$ and $\varphi_\varepsilon(x) = \varepsilon^{-n}\varphi(x/\varepsilon)$, then

$$u(x) = (\varphi_\varepsilon * u)(x)$$

for every $\varepsilon > 0$ and every $x \in \mathbb{R}^n$. Thus

$$D^\alpha u(x) = (D^\alpha \varphi_\varepsilon * u)(x)$$

[19] If F_k is a representation of Y_k in a local coordinate system (parametrization of a neighborhood of a point on the sphere), then it follows from the chain rule that on any compact set K, $\sup_K |D^\alpha F_k| \leq C \sum_{\beta \leq \alpha} \sup_{|x|=1} |D^\beta \tilde{Y}_k| \leq C(\alpha, K) k^{n/2+|\alpha|} \|Y_k\|_{L^2}$ and hence $\sum_k \sup_K |D^\alpha F_k| < \infty$ by (7.5). This shows that F_k converges to a smooth function, so does $\sum_k Y_k$.

and the Schwarz inequality easily yields

$$|D^\alpha u(x)| \leq \varepsilon^{-n/2-|\alpha|} \underbrace{\|D^\alpha \varphi\|_{L^2}}_{C_\alpha} \Big(\int_{B(x,\varepsilon)} |u|^2 \Big)^{1/2}. \tag{7.7}$$

Now

$$\int_{|x|\leq 1+\varepsilon} |P_k(x)|^2\, dx = \Big(\int_{S^{n-1}} |Y_k|^2\, d\sigma \Big) \int_0^{1+\varepsilon} t^{2k+n-1}\, dt \leq (1+\varepsilon)^{2k+n} \|Y_k\|^2_{L^2}.$$

Now for any $x \in S^{n-1}$ inequality (7.7) gives

$$|D^\alpha P_k(x)| \leq C_\alpha \varepsilon^{-n/2-|\alpha|} \Big(\int_{B(x,\varepsilon)} |P_k|^2 \Big)^{1/2} \leq C_\alpha \varepsilon^{-n/2-|\alpha|} (1+\varepsilon)^{k+n/2} \|Y_k\|_{L^2}.$$

Taking $\varepsilon = 1/k$ we see that $(1+\varepsilon)^{k+n/2}$ is bounded by a constant and hence

$$\sup_{|x|=1} \big| D^\alpha P_k(x) \big| \leq C'_\alpha k^{n/2+|\alpha|} \|Y_k\|^2_{L^2}$$

Since $\tilde{Y}_k(x) = |x|^{-k} P_k(x)$, the Leibnitz rule implies (7.6). □

8. Baire category theorem

The Baire category theorem proved below plays an important role in many areas of mathematics. In this section we will show its applications outside functional analysis and in the next two sections we will use it in the proofs of two fundamental theorems in functional analysis, the Banach-Stenihaus theorem and the Banach open mapping theorem.

Theorem 8.1 (Baire). *The intersection of a countable family of open and dense sets in a complete metric space is a dense set.*

Proof. Let $V_1, V_2, V_3, \ldots$ be open and dense sets in a complete metric space. We need to prove that their intersection $V_1 \cap V_2 \cap \ldots$ is dense. To this end it suffices to prove that for every open set $W \neq \emptyset$

$$W \cap \bigcap_{i=1}^\infty V_i \neq \emptyset. \tag{8.1}$$

The density of V_1 yields the existence of a ball

$$B(x_1, r_1) = B_1 \subset \overline{B_1} \subset V_1 \cap W, \quad r_1 < 1.$$

The density of V_2 implies the existence of a ball

$$B(x_2, r_2) = B_2 \subset \overline{B_2} \subset B_1 \cap V_2, \quad r_2 < 1/2.$$

Similarly there is a ball

$$B(x_3, r_3) = B_3 \subset \overline{B_3} \subset B_2 \cap V_3, \quad r_3 = 1/3$$

etc. We obtain a decreasing sequence of balls $B(x_i, r_i)$. The centers $\{x_i\}$ form a Cauchy and hence convergent sequence $x_i \to x \in X$. Since the balls $\overline{B_j}$ are closed and $x_i \in \overline{B_j}$ for $i \geq j$ we conclude

$$x \in \bigcap_{j=1}^{\infty} \overline{B_j} \subset W \cap \bigcap_{j=1}^{\infty} V_j$$

which proves (8.1). □

Definition. We say that a subset E of a metric space X is *nowhere dense* if the closure $\overline{E}$ has no interior points. Sets which are countable unions of nowhere dense sets are *first category.* All other sets are are *second category.*

Therefore the Baire theorem can be restated as follows.

Theorem 8.2 (Baire)**.** *A nonempty complete metric space is second category.*

As a first striking application we will prove

Proposition 8.3. *If $f \in C^\infty(\mathbb{R})$ and for every $x \in \mathbb{R}$ there is a nonnegative integer n such that $f^{(n)}(x) = 0$, then f is a polynomial.*

The following exercise shows that the result cannot be to easy.

Exercise. *Prove that there is a function $f \in C^{1000}(\mathbb{R})$ which is not a polynomial, but has the property described in the above proposition.*

Proof of the proposition. Let $\Omega \subset \mathbb{R}$ be the union of all open intervals $(a, b) \subset \mathbb{R}$ such that $f|_{(a,b)}$ is a polynomial. The set Ω is open, so

$$\Omega = \bigcup_{i=1}^{\infty} (a_i, b_i)\,, \tag{8.2}$$

where $a_i < b_i$ and $(a_i, b_i) \cap (b_i, b_j) \neq \emptyset$ for $i \neq j$. Observe that $f|_{(a_i,b_i)}$ is a polynomial (Why?).[20] We want to prove that $\Omega = \mathbb{R}$. First we will prove that $\overline{\Omega} = \mathbb{R}$. To this end it suffices to prove that for any interval $[a, b]$, $a < b$ we have $[a, b] \cap \Omega \neq \emptyset$. Let

$$E_n = \{x \in \mathbb{R} : \, f^{(n)}(x) = 0\}\,.$$

The sets $E_n \cap [a, b]$ are closed and

$$[a, b] = \bigcup_{n=0}^{\infty} E_n \cap [a, b]\,.$$

[20]It suffices to prove that f is a polynomial on every compact subinterval $[c, d] \subset (a_i, b_i)$. This subinterval has a finite covering by open intervals on which f is a polynomial. Taking an integer n larger than the maximum of the degrees of these polynomials, we see that $f^{(n)} = 0$ on $[c, d]$ and hence f is a polynomial of degree $< n$ on $[c, d]$.

Since $[a,b]$ is complete, it follows from the Baire theorem that for some n the set $E_n\cap[a,b]$ has nonempty interior (in the topology of $[a,b]$), so there is $(c,d)\subset E_n\cap[a,b]$ such that $f^{(n)}=0$ on (c,d). Accordingly f is a polynomial on (c,d) and hence

$$(c,d)\subset\Omega\cap[a,b]\neq\emptyset.$$

The set $X=\mathbb{R}\setminus\Omega$ is closed and hence complete. It remains to prove that $X=\emptyset$. Suppose not. Observe that every point $x\in X$ is an accumulation point of the set, i.e. there is a sequence $x_i\in X$, $x_i\neq x$, $x_i\to x$. Indeed, otherwise x would be an isolated point, i.e. there would be two intervals

$$(a,x),\ (x,b)\subset\Omega,\ \ x\notin\Omega\,. \tag{8.3}$$

The function f restricted to each of the two intervals is a polynomial, say of degrees n_1 and n_2. If $n>\max\{n_1,n_2\}$, then $f^{(n)}=0$ on $(a,x)\cup(x,b)$. Since $f^{(n)}$ is continuous on (a,b), it must be zero on the entire interval and hence f is a polynomal of degree $\leq n-1$ on (a,b), so $(a,b)\subset\Omega$ which contradicts (8.3).

The space $X=\mathbb{R}\setminus\Omega$ is complete. Since

$$X=\bigcup_{n=1}^{\infty}X\cap E_n\,,$$

the second application of the Baire theorem gives that $X\cap E_n$ has a nonempty interior in the topology of X, i.e. there is an interval (a,b) such that

$$X\cap(a,b)\subset X\cap E_n\neq\emptyset\,. \tag{8.4}$$

Accordingly $f^{(n)}(x)=0$ for all $x\in X\cap(a,b)$. Since for every $x\in X\cap(a,b)$ there is a sequence $x_i\to x$, $x_i\neq x$ such that $f^{(n)}(x_i)=0$ it follows from the definition of the derivative that $f^{(n+1)}(x)=0$ for every $x\in X\cap(a,b)$, and by induction $f^{(m)}(x)=0$ for all $m\geq n$ and all $x\in X\cap(a,b)$.

We will prove that $f^{(n)}=0$ on (a,b). This will imply that $(a,b)\subset\Omega$ which is a contradiction with (8.4). Since $f^{(n)}=0$ on $X\cap(a,b)=(a,b)\setminus\Omega$ it remains to prove that $f^{(n)}=0$ on $(a,b)\cap\Omega$. To this end it suffices to prove that for any interval (a_i,b_i) that appears in (8.2) such that $(a_i,b_i)\cap(a,b)\neq\emptyset$, $f^{(n)}=0$ on (a_i,b_i). Since (a,b) is not contained in (a_i,b_i) one of the endpoints belongs to (a,b), say $a_i\in(a,b)$. Clearly $a_i\in X\cap(a,b)$ and hence $f^{(m)}(a_i)=0$ for all $m\geq n$. If f is a polynomial of degree k on (a_i,b_i), then $f^{(k)}$ is a nonzero constant on (a_i,b_i), so $f^{(k)}(a_i)\neq 0$ by continuity of the derivative. Thus $k<n$ and hence $f^{(n)}=0$ on (a_i,b_i). □

EXERCISE. *As the previous exercise shows the theorem is not true if we only assume that* $f\in C^{1000}$. *Where did we use in the proof the assumption* $f\in C^{\infty}(\mathbb{R})$*?*

Theorem 8.4. *If a complete metric space X has no isolated points, then every dense G_δ set is uncountable.*

Proof. Suppose that a dense G_δ set E is countable, $E = \{x_1, x_2, \dots\}$. Then there are open and dense sets V_n such that $E = \bigcap_{n=1}^\infty V_n$. Now the sets $W_n = V_n \setminus \{x_1, \dots, x_n\}$ are still dense (because the space has no isolated points) and hence $\bigcap_{n=1}^\infty W_n \neq \emptyset$ by the Baire theorem. On the other hand

$$\bigcap_{n=1}^\infty W_n = \bigcap_{n=1}^\infty V_n \setminus \{x_1, x_2, \dots\} = E \setminus E = \emptyset$$

which is a contradiction. $\square$

Theorem 8.5. *If $f : X \to Y$ is a continuous mapping between metric spaces, then the set of points where f is continuous is G_δ and the set of points where f is discontinuous if F_σ.*

Proof. We define oscillation of f at a point $x \in X$ by

$$\operatorname{osc} f(x) = \lim_{r \to 0^+} \operatorname{diam}\left(f(B(x, r)\right).$$

It is obvious that f is continuous at x if and only if $\operatorname{osc} f(x) = 0$. It is also easy to see that the sets

$$U_n = \{x \in X : \operatorname{osc} f(x) < 1/n\}$$

are open and hence

$$\{x : f \text{ is continuous at } x\} = \bigcap_{n=1}^\infty U_n$$

is G_δ. The second part of the theorem follows from the fact that the complement of a G_δ set is F_σ. $\square$

Riemann constructed a function $f : [0, 1] \to \mathbb{R}$ that is continuous exactly at irrational points. However as a consequence of the above two results we have

Corollary 8.6. *There is no function $f : [0, 1] \to \mathbb{R}$ that is continuous exactly at rational points.*

Since the set of irrational points if G_δ, Riemann's example is a consequence of the following more general result.

Exercise. *Let $E \subset \mathbb{R}$ be a given G_δ set. Prove that there is a function $f : \mathbb{R} \to \mathbb{R}$ such that the set of points where f is continuous equals E.*

The last application of the Baire theorem provided in this section is more important because the proof is based on the so called *Baire category method* which, in many instances, is used to prove the existence of functions (or other objects) that have strange properties.

Theorem 8.7 (Banach)**.** *The set of function in $C[0,1]$ that are not differentiable at* any *point of $[0,1]$ is a dense G_δ subset of $C[0,1]$. In particular it is not empty.*

The existence of a nowhere differentiable function has been proved for the first time by Weierstrass. His construction was explicit. The above result proves existence of such functions, but does not provide any explicit example. On the other hand it proves that nowhere differentiability is a typical property of continuous functions.

Proof of the theorem. Let $f\in C[0,1]$, $x\in[0,1]$ and $0<r\leq 1/2$. If both numbers $x+r, x-r$ belong to the interval $[0,1]$, then we define

$$D(f,x,r)=\min\left\{\frac{|f(x+r)-f(x)|}{r},\ \frac{|f(x-r)-f(x)|}{r}\right\}. \tag{8.5}$$

If only one of the numbers $x+r, x-r$ belongs to $[0,1]$ we define $D(f,x,r)$ to be value of that number on the right hand side of (8.5) that exists, for example if $x+r>1$, we set $D(f,x,r)=|f(x-r)-f(x)|/r$. Let

$$G_n=\bigcup_{a>n}\ \bigcup_{0<r<1/n}\left\{f\in C[0,1]:\ D(f,x,r)\geq a \text{ for all } x\in[0,1]\right\}.$$

It is easy to see that if $f\in\bigcap_{n=1}^\infty G_n$, then

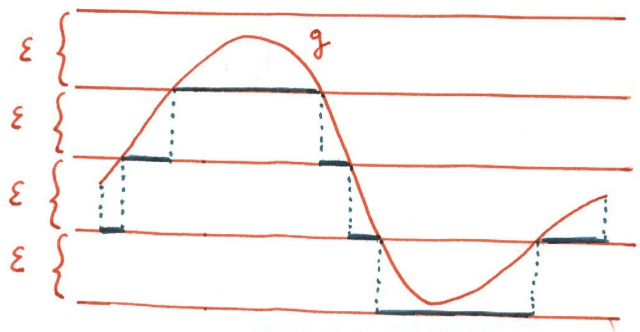
$$\limsup_{r\to 0}\left|\frac{f(x+r)-f(x)}{r}\right|=+\infty \quad \text{for all } x\in[0,1]$$

and hence f is not differentiable at any point of $[0,1]$. Thus it remains to prove that every set G_n is open and dense, because then the Baire theorem will imply that $\bigcap_{n=1}^\infty G_n$ is a dense G_δ subset of $C[0,1]$.[21] The fact that the sets G_n are open is easy and left to the reader. To prove density it suffices to prove that every $g\in C[0,1]$ can be uniformly approximated by functions $\tilde{g}\in C[0,1]$ such that

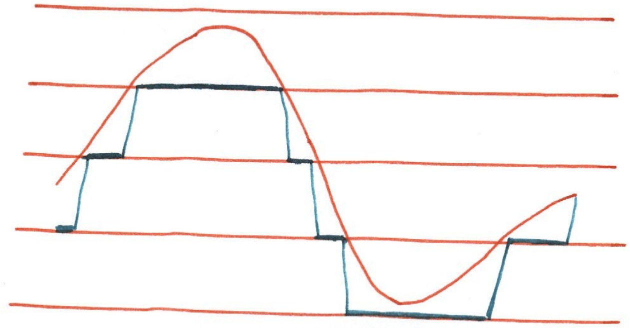
$$\exists_{0<r<1/n}\,\forall_{x\in[0,1]}\ D(\tilde{g},x,r)\geq n+1\,,$$

because this implies $\tilde{g}\in G_n$. The construction of the function $\tilde{g}$ approximating g is explained on a sequence of pictures. First we approximate g with ε accuracy by a simple function

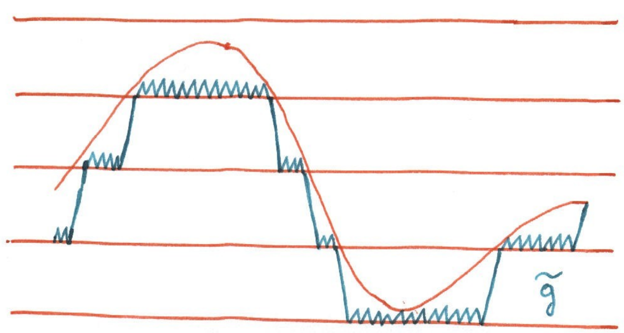
[21]Because $C[0,1]$ is complete.



Then we modify it to a piecewise linear function which has flat parts and parts with the slope larger than $n+1$.



Finally we add little teeth to the flat part, so the slope of each tooth is also larger than $n+1$.



Clearly $\sup_{x\in[0,1]} |g(x) - \tilde{g}(x)| \le \varepsilon$ and if r is very small $D(\tilde{g}, x, r) > n+1$ for all $x \in [0, 1]$. $\square$

The proof is based on the following general idea. Given a complete metric space X, we want to prove that there is an element $x \in X$ that has a certain property P. We find other, simpler to deal with, properties P_n, $n = 1, 2, \dots$ such that

(a) The sets $G_n = \{x \in X : x \text{ satisfies } P_n\}$ are open and dense;
(b) x has the property P if x has all the properties P_n.

Then $x \in \bigcap_{n=1}^\infty G_n$ has the property P and such an x exists, because the set $\bigcap_{n=1}^\infty G_n$ is nonempty by the Baire theorem. This is what is called the Baire category method.

Exercise. *We say that a function $f \in C^\infty(0,1)$ is analytic at $a \in (0,1)$ if there is $\varepsilon > 0$ such that $f(x) = \sum_{n=0}^\infty f^{(n)}(a)(x-a)^n/n!$ for $|x-a| < \varepsilon$. Use the Baire category method to prove that there is a function $f \in C^\infty(0,1)$ that is not analytic at any point.*

9. Banach-Steinhaus theorem

The following theorem as well as each of the four corollaries that follow are called Banach-Steinhaus theorem.

Theorem 9.1 (Banach-Steinhaus)**.** *Let X be a complete metric space and let $\{f_i\}_{i\in I}$ be a family of continuous real-valued functions on X. Then exactly one of the following two conditions is satisfied*

(a) *There is a nonempty open set $U \subset X$ and a constant $M > 0$ such that*
$$\sup_{i\in I} |f_i(x)| \leq M \quad \textit{for all } x \in U;$$
(b) *There is a dense G_δ set $E \subset X$ such that*
$$\sup_{i\in I} |f_i(x)| = \infty \quad \textit{for all } x \in E.$$

Proof. If $\varphi(x) = \sup_{i\in I} |f_i(x)|$, then the sets
$$V_n = \{x \in X : \varphi(x) > n\}$$
are open because of continuity of the functions f_i. Suppose that the condition (a) is not satisfied. Then for every open set $U \neq \emptyset$, $V_n \cap U \neq \emptyset$ and hence the sets V_n are dense. According to the Baire theorem $E = \bigcap_n V_n$ is a dense G_δ set and obviously points of E satisfy (b). $\square$

Corollary 9.2. *Let X be a complete metric space and let $\{f_i\}_{i\in I}$ be a family of continuous real-valued functions on X. If the functions in the family are pointwise bounded, i.e.*
$$\sup_{i\in I} |f_i(x)| < \infty \quad \textit{for every } x \in X,$$

then there is a nonempty open set $U \subset X$ and a constant $M > 0$ such taht

$$\sup_{i\in I} |f_i(x)| \le M \quad \text{for all } x \in U.$$

The above two results are very important in the case X is a Banach space.

Corollary 9.3. *Let X be a Banach space and Y a normed space. Let $\{L_i\}_{i\in I} \subset B(X,Y)$. Then either*

$$\sup_{i\in I} \|L_i\| < \infty$$

or there is a dense G_δ set $E \subset X$ such that

$$\sup_{i\in I} \|L_i x\| = \infty \quad \text{for all } x \in E.$$

Proof. The functions $f_i(x) = \|L_i x\|$, $i \in I$ are continuous and real-valued. If the second condition is not satisfied, then it follows from the Banach-Steinhaus theorem that there is an open ball $B(x_0, r_0) \subset X$ such that

$$\sup_{i\in I} |f_i(x)| = M < \infty \quad \text{for all } x \in B(x_0, r_0)\ .$$

For $x \neq 0$

$$y = x_0 + \frac{r_0}{2\|x\|}\, x \in B(x_0, r_0)$$

and hence for all $i \in I$ we have

$$\|L_i x\| = \frac{2\|x\|}{r_0} \Big\| L_i\Big(x_0 + \frac{r_0}{2\|x\|}\, x\Big) - L_i x_0 \Big\| \le \frac{4M}{r_0}\, \|x\|$$

which yields

$$\sup_{i\in I} \|L_i\| \le \frac{4M}{r_0} < \infty\,.$$

The claim is proved. □

Corollary 9.4. *Let X be a Banach space and Y a normed space. Let $\{L_i\}_{i\in I} \subset B(X,Y)$. If for every $x \in X$*

$$\sup_{i\in I} \|L_i x\| < \infty$$

then

$$\sup_{i\in I} \|L_i\| < \infty\,.$$

Corollary 9.5. *Let X be a Banach space and Y a normed space. If $\{L_n\}_{n=1}^\infty \subset B(X,Y)$ is a pointwise convergent sequence, i.e. for every $x \in X$ the limit*

$$Lx := \lim_{n\to\infty} L_n x$$

exists, then $L \in B(X,Y)$. Moreover

$$\sup_n \|L_n\| < \infty$$

and

$$\|L\| \leq \liminf_{n\to\infty} \|L_n\| .$$

Proof. Clearly $L : X \to Y$ is a linear mapping. The existence of the limit Lx implies that

$$\sup_n \|L_n x\| < \infty \quad \text{for every } x \in X$$

and hence

$$\sup_n \|L_n\| < \infty$$

by the Banach-Steinhaus theorem. We have

$$\|Lx\| = \lim_{n\to\infty} \|L_n x\| = \liminf_{n\to\infty} \|L_n x\| \leq \big(\liminf_{n\to\infty} \|L_n\|\big) \|x\|$$

and the result follows. □

Now we will present several applications of the Banach-Steinhaus theorem.

9.1. **Multilinear operators.** Let X and Y be normed spaces over $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. We say that

$$B : X \times Y \to \mathbb{K}$$

is a *two-linear functional* if it is a linear functional with respect to each variable, i.e.

$$B(x, \cdot) : Y \to \mathbb{K} \quad \text{is linear for every } x \in X,$$

$$B(\cdot, y) : X \to \mathbb{K} \quad \text{is linear for every } y \in Y.$$

Similarly as in the case of linear functionals one can prove

Lemma 9.6. *For a two-linear functional $B : X \times Y \to \mathbb{K}$ the following conditions are equivalent.*

(a) *The function B is continuous.*
(b) *The function B is continuous at $(0, 0)$.*
(c) *There is a constant $C > 0$ such that*

$$|B(x, y)| \leq C\|x\| \, \|y\| \quad \textit{for all } x \in X,\ y \in Y.$$

It is known that even for a function on $\mathbb{R}^2$ continuity with respect to each variable does not imply continuity. However we have

Theorem 9.7. *If X and Y are Banach spaces and $B : X \times Y \to \mathbb{K}$ is a two-linear functional which is continuous with respect to each variable, then B is continuous.*

Proof. It suffices to prove continuity at $(0,0)$, i.e. the implication

$$(x_n, y_n) \to (0,0) \quad \Rightarrow \quad B(x_n, y_n) \to 0\,.$$

Define a family of functionals $T_n \in Y^*$ by $T_n(y) = B(x_n, y)$. Since for every $y \in Y$, $T_n(y) \to B(0,y) = 0$, we have

$$\sup_n |T_n(y)| < \infty \quad \text{for every } y \in Y.$$

and thus

$$\sup_n \|T_n\| = M < \infty$$

by the Banach-Steihnaus theorem. Hence $|T_n(y)| \leq M\|y\|$ for every n and all $y \in Y$. In particular

$$|B(x_n, y_n)| = |T_n(y_n)| \leq M\|y_n\| \to 0$$

and the result follows. □

9.2. **Landau's theorem.**

Theorem 9.8 (Landau)**.** *If the series $\sum_{i=1}^\infty \eta_i\xi_i$ converges for every $(\xi_i) \in \ell^p$, $1 \leq p \leq \infty$, then $(\eta_i) \in \ell^q$, where $1/p + 1/q = 1$.*

Proof. We will prove the theorem in the case $1 < p \leq \infty$, but an obvious modification gives also the proof in the case $p = 1$. Define a bounded functional on ℓ^p by the formula

$$T_n x = \sum_{i=1}^n \eta_i\xi_i, \quad n = 1, 2, 3, \ldots,$$

where $x = (\xi_i) \in \ell^p$. The functional T_n is fiven by $(\eta_1, \ldots, \eta_n, 0, 0, \ldots) \in \ell^q$ and hence (see Theorem 2.12)

$$\|T_n\| = \Big(\sum_{i=1}^n |\eta_i|^q\Big)^{1/q}.$$

Thus $\{T_n\}_n$ is a sequence of bounded functionals of ℓ^p, such that $T_n x$ is convergent for every $x \in \ell^p$ (by the assumption in the theorem) and hence the Banach-Steinhaus theorem yields

$$\Big(\sum_{i=1}^\infty |\eta_i|^q\Big)^{1/q} = \sup_n \|T_n\| < \infty\,.$$

The proof is complete. □

9.3. **Matrix summability methods.** If a sequence (a_n) converges to a limit g, then also

$$c_n = \frac{a_1 + a_2 + \ldots + a_n}{n} \to g\,. \tag{9.1}$$

There are, however, divergent sequences (a_n) for which the sequence of arithmetic means (9.1) converges. Thus assigning to the sequence (a_n) the limit of the sequence of arithmetic means (c_n) allows us to extend the notion of the limit to a larger class of sequences, not necessarily convergent in the oridinary sense. This particular method is known as the *Cesaro summability method*.

We may, however, consider more general methods by taking instead of arithmetic means other linear combinations of the a_i's. This leads to the following

DEFINITION. We say that a sequence (a_n) is *summable* to a generalized limit g by the *matrix summability method* $A = (\xi_{ij})_{i,j=1}^{\infty}$ if

(1) the series $\sum_{i=1}^{\infty} \xi_{ni} a_i$ converges for every $n = 1, 2, 3, \ldots$;
(2) $\lim_{n\to\infty} \sum_{i=1}^{\infty} \xi_{ni} a_i = g\,.$

For example if

$$A_1 = \begin{bmatrix} 1 & & & & 0 \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & \ddots \\ 0 & & & & \end{bmatrix} \qquad A_2 = \begin{bmatrix} \frac{1}{1} & 0 & 0 & 0 & \ldots \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & \ldots \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & \ldots \\ & & \ldots & & \end{bmatrix}$$

then the sequence (a_n) is summable to g by the method A_1 if $a_n \to g$ and by the method A_2 if

$$\frac{a_1 + a_2 + \ldots + a_n}{n} \to g\,,$$

so we can recover both the classical notion of the convergence and the Cesaro summability method.

DEFINITION. We say that the matrix method A is *regular* if every convergent sequence if summable to the same limit by the method A.

Clearly methods A_1 and A_2 are regular.

Theorem 9.9 (Toeplitz). *The matrix summability method $A = (\xi_{ij})_{i,j=1}^{\infty}$ is regular if and only if the following conditions are satisfied:*

(a) $\sup_n \sum_{i=1}^{\infty} |\xi_{ni}| < \infty$,
(b) $\lim_{n\to\infty} \xi_{ni} = 0$ *for* $i = 1, 2, 3, \ldots$,
(c) $\lim_{n\to\infty} \sum_{i=1}^{\infty} \xi_{ni} = 1$.

Proof. Suppose that the matrix method A is regular. If $(a_n) = (0, 0, \ldots, 0, 1, 0, \ldots)$ with 1 on ith coordinate, i.e. $a_n = \delta_{ni}$, then $a_n \to 0$ and hence

$$\lim_{n\to\infty} \xi_{ni} = \lim_{n\to\infty} \sum_{j=1}^{\infty} \xi_{nj} a_j = 0$$

which is condition (b). If $(a_n) = (1, 1, 1, \ldots)$, then $a_n \to 1$ and hence

$$\lim_{n\to\infty} \sum_{i=1}^{\infty} \xi_{ni} = \lim_{n\to\infty} \sum_{i=1}^{\infty} \xi_{ni} a_i = 1$$

which is condition (c). The condition (a) is more difficult to prove and it will actually follow from the Banach-Steinhaus theorem.

On the space of convergent sequences we define functionals[22]

$$T_{nm}x = \sum_{i=1}^{m} \xi_{ni} a_i, \quad \text{where } x = (a_i) \in c.$$

We have

$$\|T_{nm}\| = \sum_{i=1}^{m} |\xi_{ni}| \,.$$

Since the series $\sum_{i=1}^{\infty} \xi_{ni} a_i$ converges for every $x = (a_i) \in c$, the sequence $(T_{nm}x)_m$ converges for every $x \in c$ and hence it follows from the Banach-Steinhaus theorem that for every n

$$\sup_m \|T_{nm}\| = \sum_{i=1}^{\infty} |\xi_{ni}| < \infty \,.$$

This, however, implies that

$$T_n x = \sum_{i=1}^{\infty} \xi_{ni} a_i$$

is a bounded functional on c with

$$\|T_n\| = \sum_{i=1}^{\infty} |\xi_{ni}| \,.$$

By the assumption, the sequence $T_n x$ converges for every $x \in c$ and the second application of the Banach-Steinhaus theorem gives

$$\sup_n \|T_n\| = \sup_n \sum_{i=1}^{\infty} |\xi_{ni}| < \infty$$

which is the condition (a) of the theorem.

[22] It is easy to see that every sequence $(\xi_i) \in \ell^1$ defines a functional on $T \in c^*$ by $Tx = \sum_{i=1}^{\infty} \xi_i a_i$, where $x = (a_i) \in c$ and $\|T\| = \sum_{i=1}^{\infty} |\xi_i|$. This follows from the argument used in the proof of Theorem 2.10.

It remains to prove that conditions (a), (b), (c) imply that the method A is regular.

For every $x = (a_n) \in c$ the series

$$T_n x = \sum_{i=1}^{\infty} \xi_{ni} a_i$$

converges and it actually defines a bounded functionals on c with

$$\sup_n \|T_n\| = \sup_n \sum_{i=1}^{\infty} |\xi_{ni}| < \infty\,.$$

Note that also

$$Tx = \lim_{n\to\infty} a_n$$

is a bounded functional on c. We have to prove that for every $x \in c$, $T_n x \to Tx$, as $n \to \infty$. Let A_0 be a subset of c consisting of the sequences $(1,0,0,\ldots)$, $(0,1,0,0,\ldots)$, $(0,0,1,0,0,\ldots)$,... and also of the sequence $(1,1,1,1,\ldots)$. The conditions (b) and (c) readily imply that $T_n x \to Tx$ for all $x \in A_0$. By the linearity we also have

$$T_n x \to Tx \quad \text{for } x \in X_0 = \operatorname{span} A_0. \tag{9.2}$$

It is easy to see that X_0 is a dense subset of c (why?), so we have convergence on a dense subset of c. Since

$$\sup_n \|T_n - T\| \le \|T\| + \sup_n \|T_n\| < \infty$$

it easily follows that

$$T_n x \to Tx \quad \text{for all } x \in c.$$

Indeed, given $x \in c$ and $\varepsilon > 0$ there is $x_0' \in X_0$ such that

$$\|x' - x\| \le \frac{\varepsilon}{2\sup_n \|T_n - T\|} \tag{9.3}$$

and hence there is n_0 such that for $n > n_0$

$$|(T_n - T)x| \le \|T_n - T\|\,\|x - x'\| + |T_n x' - Tx'| < \varepsilon$$

by (9.2) and (9.3). □

9.4. **Divergent Fourier series.** If $f \in C(S^1)$, then according to the Carleson theorem (Theorem 6.5) the sequence of partial sums of the Fourier series converges to f a.e. This is natural to expect that this sequence actually converges to f everywhere. Surprisingly, this is not always true. We will use the Banach-Steinhaus theorem to demonstrate existence of functions in $C(S^1)$ such that the Fourier series diverges on an uncountable and dense subset of S^1.

Theorem 9.10. *There is a dense G_δ set $E \subset C(S^1)$ such that for each $f \in E$ the set*

$$\{x \in [0,1] : \sup_n |s_n(f,x)| = +\infty\}$$

is a dense and uncountable G_δ subset of $[0,1]$.

Proof. In the first step we will prove existence of continuous functions with unbounded partial sums of the Fourier series at 0. For each integer n we define a bounded functional on $C(S^1)$

$$\Lambda_n f = s_n(f,0), \quad f \in C(S^1)\,.$$

By Proposition 6.2 we have

$$\Lambda_n f = \int_{-1/2}^{1/2} f(y) D_n(y)\, dy\,,$$

where

$$D_n(y) = \frac{\sin \pi(2n+1)y}{\sin \pi y}\,.$$

Clearly

$$\|\Lambda_n\| \le \int_{-1/2}^{1/2} |D_n(y)|\, dy\,. \tag{9.4}$$

Actually we have equality

$$\|\Lambda_n\| = \int_{-1/2}^{1/2} |D_n(y)|\, dy. \tag{9.5}$$

Indeed, let

$$g(y) = \begin{cases} 1 & \text{if } D_n(y) \ge 0, \\ -1 & \text{if } D_n(y) < 0. \end{cases}$$

It is easy to see that there is a sequence $f_j \in C(S^1)$ such that $-1 \le f_j \le 1$, $f_j \to g$ in $L^1(S^1)$ and $\|f_j\|_\infty = 1$. Hence

$$\begin{aligned} \|\Lambda_n\| &\ge \lim_{j\to\infty} \Lambda_n f_j = \lim_{j\to\infty} \int_{-1/2}^{1/2} f_j(y) D_n(y)\, dy \\ &= \int_{-1/2}^{1/2} g(y) D_n(y)\, dy = \int_{-1/2}^{1/2} |D_n(y)|\, dy \end{aligned}$$

which together with (9.4) proves (9.5). We claim that $\|\Lambda_n\| \to \infty$ as $n \to \infty$. Indeed, $|\sin \pi y| \leq |\pi y|$ and hence

$$
\begin{aligned}
\|\Lambda_n\| &= \int_{-1/2}^{1/2} |D_n(y)|\, dy \\
&\geq \frac{2}{\pi} \int_0^{1/2} |\sin \pi(2n+1)y| \, \frac{dy}{y} \\
&\overset{\substack{\pi(2n+1)y=t \\ dy/y=dt/t}}{=} \frac{2}{\pi} \int_0^{(n+1/2)\pi} |\sin t| \frac{dt}{t} \\
&> \frac{2}{\pi} \sum_{k=1}^{n} \frac{1}{k\pi} \underbrace{\int_{(k-1)\pi}^{k\pi} |\sin t|\, dt}_{2} \\
&= \frac{4}{\pi} \sum_{k=1}^{n} \frac{1}{k} \to \infty \quad \text{as } n \to \infty.
\end{aligned}
$$

We proved that

$$\sup_n \|\Lambda_n\| = \infty.$$

According to the Banach-Steinhaus theorem there is a dense G_δ set $E_0 \subset C(S^1)$ such that

$$\sup_n |\Lambda_n f| = \sup_n |s_n(f,0)| = \infty, \quad \text{for every } f \in E_0.$$

In the above reasoning we could replace 0 by any other point $x \in [0,1]$, so for each $x \in [0,1]$ there is a dense G_δ set $E_x \subset C(S^1)$ such that

$$\sup_n |s_n(f,x)| = \infty \quad \text{for all } f \in E_x.$$

Now let $\{x_1, x_2, \ldots\} \subset [0,1]$ be a dense subset. Then by the Baire theorem $E = \bigcap_i E_{x_i}$ is a dense G_δ set in $C(S^1)$ and for every $f \in E$

$$\sup_n |s_n(f,x_i)| = \infty \quad \text{for all } i = 1,2,3,\ldots$$

We still need the following lemma that we leave as an exercise.

Lemma 9.11. *If $f \in C(S^1)$, then the set $\{x \in [0,1] : \sup_n |s_n(f,x)| = \infty\}$ is G_δ.*

Accordingly $\sup_n |s_n(f,x)| = \infty$ for x in a dense G_δ set, which is uncountable by Theorem 8.4. □

10. Banach open mapping theorem

Another consequence of the Baire theorem is the following fundamental result.

Theorem 10.1 (Banach open mapping theorem)**.** *If X and Y are Banach spaces and $T \in B(X,Y)$ is a surjection, then T is an open mapping, i.e. $T(U) \subset Y$ is open whenever $U \subset X$ is open.*

Proof. Since every ball can be mapped by an affine transformation onto the unit ball, it suffices to prove that the image of the unit ball contains a neighborhood of 0, i.e.

$$B(0,\delta) \subset T(B(0,1)) \quad \text{for some } \delta > 0.$$

Surjectivity of T gives

$$Y = \bigcup_{k=1}^{\infty} T(B(0,k))$$

and by Baire theorem at least one of the sets $\overline{T(B(0,k))}$ has nonempty interior, i.e.

$$B(y_0,\eta) \subset \overline{T(B(0,k))}\,.$$

For y such that $\|y\| < \eta$ there exist sequences $(x_n'), (x_n'') \subset B(0,k)$ such that

$$Tx_n' \to y_0, \quad Tx_n'' \to y_0 + y\,.$$

This yields

$$T(x_n'' - x_n') \to y, \quad x_n'' - x_n' \in B(0,2k)$$

and hence

$$B(0,\eta) \subset \overline{T(B(0,2k))}\,.$$

Taking $2\delta = \eta/2k$ we obtain

$$B(0,2\delta r) \subset \overline{T(B(0,r))} \quad \text{for every } r > 0.$$

In particular for $y \in B(0,\delta)$ $(r = 1/2)$ there is $x_1 \in B(0,r) = B(0,1/2)$ such that $\|y - Tx_1\| < \delta/2$. Hence $y - Tx_1 \in B(0,\delta/2)$ $(r = 1/4)$ and now we find $x_2 \in B(0,r) = B(0,1/4)$ such that

$$\|(y - Tx_1) - Tx_2\| < \delta/4\,.$$

Hence

$$y - T(x_1 + x_2) \in B(0,\delta/4), \quad (r = 1/8)\,.$$

Next we find $x_3 \in B(0,1/8)$ such that

$$\|y - T(x_1 + x_2 + x_3)\| < \delta/8\,.$$

By induction, we construct a sequence $(x_n) \subset X$ such that

$$\|x_n\| < 2^{-n} \quad \text{and} \quad \|y - T(x_1 + \ldots + x_n)\| < 2^{-n}\delta\,.$$

Since X is a Banach space we conclude that

$$x = \sum_{n=1}^{\infty} x_n \in X, \quad \|x\| \le \sum_{n=1}^{\infty} \|x_n\| < 1$$

and

$$Tx = \lim_{n\to\infty} T(x_1 + \ldots + x_n) = y\,.$$

The proof is complete. $\square$

Corollary 10.2. *If X and Y are Banach spaces and $T \in B(X,Y)$ is an isomorphism of linear spaces X and Y, then T is an isomorphism of Banach spaces X and Y, i.e. $T^{-1} \in B(Y,X)$.*

Compare this result with an example that follows Proposition 2.6.

Corollary 10.3. *Let X be a Banach space with respect to any one of the two norms $\|\cdot\|_1$, $\|\cdot\|_2$. If there is a constant $C > 0$ such that*

$$\|x\|_1 \leq C\|x\|_2 \quad \textit{for all } x \in X$$

then the norms are equivalent and hence there is a constant $C' > 0$ such that

$$\|x\|_2 \leq C'\|x\|_1 \quad \textit{for all } x \in X.$$

Proof. The identity mapping $\mathrm{id} : (X, \|\cdot\|_2) \to (X, \|\cdot\|_1)$ is an isomorphism of Banach spaces by Corollary 10.2. $\square$

If $f : X \to Y$ is a function between Banach spaces, then its graph

$$G_f = \{(x,y) \in X \times Y : \, x \in X, \; y = f(x)\}$$

is a subset of the Banach space $X \oplus Y$.

Theorem 10.4 (Closed graph theorem)**.** *A linear mapping between Banach spaces $T : X \to Y$ is bounded if and only if G_T is a closed subset of $X \oplus Y$.*

Proof. The implication $\Rightarrow$ is a direct consequence of the definition of continuity, but the other implication $\Leftarrow$ is more difficult. G_T is a closed linear subspace of the Banach space $X \oplus Y$, so it is a Banach space with respect to the norm

$$\|(x, Tx)\| = \|x\| + \|Tx\| \, .$$

The projection on the first component

$$X \oplus Y \ni (x,y) \mapsto x \in X$$

is a bounded operator. Its restriction to G_T is a bounded operator as well. Since it is an isomorphism of linear spaces G_T and X, we conclude from the Banach open mapping theorem that it is an isomorphism of Banach spaces G_T and X. Hence the inverse mapping

$$X \ni x \mapsto (x, Tx) \in G_T$$

is bounded, i.e. $\|x\| + \|Tx\| \leq C\|x\|$ and thus $\|Tx\| \leq (C-1)\|x\|$, which proves boundedness of T. $\square$

According to the closed graph theorem in order to prove boundedness of a linear mapping between Banach spaces $T : X \to Y$ it suffices to prove the implication

$$x_n \to x, \; Tx_n \to y \quad \Rightarrow \quad y = Tx \, . \tag{10.1}$$

Now we will show two applications.

10.1. **Symmetric operators are bounded.**

Theorem 10.5 (Hellinger-Toeplitz)**.** *If a linear mapping of a Hilbert space $T : H \to H$ is symmetric, i.e.*

$$\langle Tx, y \rangle = \langle x, Ty \rangle$$

for all $x, y \in H$, then it is bounded.

Proof. We need to prove the implication (10.1). For every $z \in H$ we have

$$\begin{aligned} \langle y, z \rangle &= \lim_{n\to\infty} \langle Tx_n, z \rangle = \lim_{n\to\infty} \langle x_n, Tz \rangle \\ &= \langle x, Tz \rangle = \langle Tx, z \rangle . \end{aligned}$$

Hence $\langle Tx - y, z \rangle = 0$ for all $z \in H$ and taking $z = Tx - y$ we obtain $Tx - y = 0$. □

10.2. **Complemented subspaces.**

Theorem 10.6. *If M and N are closed subspaces of a Banach space X, such that*

$$M + N = X \quad \text{and} \quad M \cap N = \{0\} ,$$

then the space X is isomorphic to the direct sum $M \oplus N$. The isomorphism is given by

$$M \oplus N \ni (m, n) \mapsto m + n \in X .$$

Moreover the spaces X/M and X/N are isomorphic to N and M respectively under the quotient map.

Proof. M and N are Banach spaces as closed subspaces of a Banach space. Hence $M \oplus N$ is a Banach space. The mapping

$$M \oplus N \ni (m, n) \mapsto m + n \in X$$

is one-to-one (because $M \cap N = \{0\}$) and onto (because $M + N = X$). Hence it follows from the open mapping theorem that it is an isomorphism of Banach spaces. Similarly the mappings

$$M \to X/N, \qquad m \mapsto [m],$$

$$N \to X/M \qquad n \mapsto [n]$$

are one-to-one and onto, hence they are isomorphisms of Banach spaces. □

Proposition 10.7. *If $L : X \oplus Y \to Z$ is an isomorphism of Banach spaces, then $L(X)$ and $L(Y)$ are closed subspaces of Z and hence $Z = L(X) \oplus L(Y)$.*

Proof. The subspaces $L(X)$ and $L(Y)$ are closed since

$$L(X) = L^{-1}(X \times \{0\}), \quad L(Y) = L^{-1}(\{0\} \times Y)$$

and the result easily follows. □

DEFINITION. Let M be a closed subspace of a Banach space X. We say that M is *complemented* in X if there is a subspace N of X such that

$$M + N = X, \qquad M \cap N = \{0\}\,.$$

Proposition 10.8. *Every closed subspace M of a Hilbert space X is complemented.*

Proof. $H = M \oplus M^\perp$. □

DEFINITION. Let X be a Banach space. The mapping $P \in B(X) = B(X, X)$ is called a *projection* if $P^2 = P$, i.e. $P(P(x)) = P(x)$ for all $x \in X$. We denote the kernel (null space) and the range of the projection by

$$\mathcal{N}(P) = \{x \in X : \, Px = 0\}, \quad \mathcal{R}(P) = \{Px : \, x \in X\}\,.$$

Theorem 10.9.

(a) *If $P \in B(X)$ is a projection, then*[23]

$$X = \mathcal{R}(P) \oplus \mathcal{N}(P)\,.$$

(b) *If $X = M \oplus N$ is a direct sum of closed subspaces, then there is a projection $P \in B(X)$ such that*

$$M = \mathcal{R}(P) \quad \textit{and} \quad N = \mathcal{N}(P)\,.$$

Corollary 10.10. *A closed subspace of a Banach space is complemented if and only if it is the image of a projections.*

Proof of Theorem 10.9. (a) $\mathcal{N}(P)$ is closed as a preimage of 0 of a continuous mapping. Since

$$\mathcal{R}(P) = \mathcal{N}(I - P) = \{x \in X : \, x = Px\}\,,$$

also $\mathcal{R}(P)$ is closed. Now it suffices to show that

$$\mathcal{R}(P) \cap \mathcal{N}(P) = \{0\} \quad \text{and} \quad \mathcal{R}(P) + \mathcal{N}(P) = X\,.$$

If $x \in \mathcal{R}(P) \cap \mathcal{N}(P)$, then $x = Px$ and $Px = 0$, so $x = 0$. Moreover every element $x \in X$ can be represented as

$$x = \underbrace{Px}_{\in \mathcal{R}(P)} + \underbrace{(x - Px)}_{\in \mathcal{N}(P)}\,.$$

(b) If $P : M \oplus N \to M$, $P(x, y) = x$ is the projection onto the first component, then $M = \mathcal{R}(P)$ and $N = \mathcal{N}(P)$. The proof is complete. □

[23] i.e. the subspaces $\mathcal{N}(P)$ and $\mathcal{R}(P)$ are closed and $\mathcal{R}(P) + \mathcal{N}(P) = X$, $\mathcal{R}(P) \cap \mathcal{N}(P) = \{0\}$

The following theorem is a consequence of the Hahn-Banach theorem and we will prove it in the next section.

Theorem 10.11. *Let M be a closed subspace of a Banach space X.*

(a) *If $\dim M < \infty$, then M is complemented in X.*
(b) *If $\dim(X/M) < \infty$, then M is complemented in X.*

Actually it turns out that a property of being a complemented subspace is quite rare. Namely we have

Theorem 10.12 (Lindenstrauss-Tzafriri)**.** *If X is a real Banach space and every closed subspace of X is complemented, then X is isomorphic to a Hilbert space.*

This is a very difficult theorem and we will not prove it. We will prove, however

Theorem 10.13 (Phillips)**.** *The space c_0 is not complemented in ℓ^∞. Equivalently there is no bounded linear projection of ℓ^∞ onto c_0.*

Proof. We will need the following lemma.

Lemma 10.14. *There is an uncountable family $\{A_i\}_{i\in I}$ of subsets of $\mathbb{N}$ such that*

(a) *A_i is infinite for every $i \in I$;*
(b) *$A_i \cap A_j$ is finite for $i \neq j$.*

Proof. It suffices to prove the lemma with $\mathbb{N}$ replaced by $\mathbb{Q} \cap (0, 1)$. For each irrational number $i \in (0, 1) \setminus \mathbb{Q} := I$ let A_i be a sequence of rationals in $(0, 1)$ convergent to i. It is easy to see that the family $\{A_i\}_{i\in I}$ has the desired properties. □

By contradiction suppose that

$$\ell^\infty = c_0 \oplus X, \quad X \subset \ell^\infty,$$

so $X \simeq \ell^\infty/c_0$. Consider the family of functionals $\{e_n\}_{n\in\mathbb{N}}$ on X

$$\langle e_n, x\rangle = x_n, \quad x = (x_i) \in X \subset \ell^\infty.$$

The family $\{e_n\}_n$ in X^* is *total*, i.e. it has the property that

$$\langle e_n, x\rangle = 0 \text{ for all } n \quad \Rightarrow \quad x = 0.$$

It suffices to prove that there is no countable total family of functionals in $(\ell^\infty/c_0)^*$.

Let $\{A_i\}_{i\in I}$ be an uncountable family of subsets of $\mathbb{N}$ as in the lemma. We can identify ℓ^∞ with the space of bounded functions $f:\mathbb{N}\to\mathbb{C}$ and for each $i\in I$ we define

$$f_i=\chi_{A_i}\in\ell^\infty,\quad [f_i]\in\ell^\infty/c_0.$$

Note that for any finite collection of functions $f_{i_1},\ldots,f_{i_m}$ from the family, $\sum_{k=1}^m f_{i_k}>1$ on a finite subset $A\subset\mathbb{N}$ by the property (b) of the lemma. Since

$$\chi_A\sum_{k=1}^m f_{i_k}\in c_0$$

we conclude that

$$\Big\|\sum_{k=1}^m [f_{i_k}]\Big\|\leq\Big\|\sum_{k=1}^m f_{i_k}-\chi_A\sum_{k=1}^m f_{i_k}\Big\|_\infty=1\,.$$

By the same argument if $b_k\in\mathbb{C}$, $|b_k|\leq 1$, $k=1,2,\ldots,m$, then

$$\Big\|\sum_{k=1}^m b_k[f_{i_k}]\Big\|\leq 1\,.$$

We will prove now that for every $x^*\in(\ell^\infty/c_0)^*$ the set

$$\{[f_i]:\,\langle x^*,[f_i]\rangle\neq 0\}$$

is countable. To this end it suffices to prove that for each integer n the set

$$C(n)=\{[f_i]:\,\langle x^*,[f_i]\rangle\geq 1/n\}$$

is finite. Choose $[f_{i_1}],\ldots,[f_{i_m}]\in C(n)$ and let

$$b_k=\operatorname{sgn}\langle x^*,[f_{i_k}]\rangle=\overline{\langle x^*,[f_{i_k}]\rangle}/|\langle x^*,[f_{i_k}]\rangle|\,.$$

It follows from the above observation that for $x=\sum_{k=1}^m b_k[f_{i_k}]$ we have $\|x\|\leq 1$. Hence

$$\|x^*\|\geq|\langle x^*,x\rangle|\geq m/n,\quad m\leq n\|x^*\|,$$

so $C(n)$ is finite for every n and actually $\#C(n)\leq n\|x^*\|$. Accordingly, if $\{h_n\}_{n\in\mathbb{N}}$ is a countable subset of $(\ell^\infty/c_0)^*$, then

$$\{[f_i]:\,\langle h_n,[f_i]\rangle\neq 0\text{ for some }n=1,2,3,\ldots\}=\bigcup_{n=1}^\infty\{[f_i]:\,\langle h_n,[f_i]\rangle\neq 0\}$$

is countable and hence there is[24] f_i such that $\langle h_n,[f_i]\rangle=0$ for all n, so the family $\{h_n\}_n$ cannot be total. □

[24]Because the family $\{f_i\}_{i\in I}$ is uncountable.

11. Hahn-Banach Theorem

Let X be a linear space over $\mathbb{R}$. We *do not* require X to be a linear normed space. The Hahn-Banach theorem says that functionals defined on subspaces of X can be extended to functionals on X in a way that if they satisfy a certain inequality on the subspace, then the extension will satisfy the same inequality.

Definition. A function $p : X \to \mathbb{R}$ defined on a real linear space X is called a *Banach functional* if

$$p(x+y) \le p(x) + p(y), \quad p(tx) = tp(x)$$

for all $x, y \in X$ and $t \ge 0$.

Theorem 11.1 (Hahn-Banach)**.** *Let $p : X \to \mathbb{R}$ be a Banach functional on a linear space X over $\mathbb{R}$ and let M be a linear subspace of X. If $f : M \to \mathbb{R}$ is a linear functional satisfying*

$$f(x) \le p(x) \quad \text{for } x \in M\,,$$

then there is a linear functional $F : X \to \mathbb{R}$ being an extension of f, i.e.

$$f(x) = F(x) \quad \text{for } x \in M$$

and such that

$$-p(-x) \le F(x) \le p(x) \quad \text{for } x \in X.$$

Proof. In the first step we will prove that if $\tilde{f}$ is a functional defined on a proper subspace $\tilde{M}$ of X and it satisfies $\tilde{f}(x) \le p(x)$ on that subspace, then we can extend it to a bigger subspace $\tilde{\tilde{M}} \supset \tilde{M}$. Actually $\tilde{\tilde{M}}$ will be obtained from $\tilde{M}$ by adding one independent vector, so that $\dim(\tilde{\tilde{M}}/\tilde{M}) = 1$, and of course the extension will satisfy $\tilde{\tilde{f}}(x) \le p(x)$ for $x \in \tilde{\tilde{M}}$. In the second part we will use the Hausdorff maximality theorem to conclude the existence of the maximal extension. Then it will easily follow that the maximal extension is defined on X and has all required properties.

Suppose that $\tilde{f} : \tilde{M} \to \mathbb{R}$ satisfies

$$\tilde{f}(x) \le p(x), \quad \text{for } x \in \tilde{M}$$

and that $\tilde{M} \ne X$. Let $x_1 \in X \setminus \tilde{M}$ and define

$$\tilde{\tilde{M}} = \{x + tx_1 : x \in \tilde{M},\, t \in \mathbb{R}\}\,.$$

Then $\tilde{M}$ is a proper subspace of $\tilde{\tilde{M}}$. For $x, y \in \tilde{M}$ we have

$$\tilde{f}(x) + \tilde{f}(y) = \tilde{f}(x+y) \le p(x+y) \le p(x - x_1) + p(x_1 + y)\,.$$

Hence

$$\tilde{f}(x) - p(x - x_1) \le p(x_1 + y) - \tilde{f}(y)\,.$$

Let

$$\alpha = \sup_{x\in\tilde{M}} \tilde{f}(x) - p(x - x_1)\,.$$

Then

$$\tilde{f}(x) - \alpha \leq p(x - x_1) \quad \text{and} \quad \tilde{f}(y) + \alpha \leq p(x_1 + y)\,. \tag{11.1}$$

We extend $\tilde{f}$ to $\tilde{\tilde{M}}$ by the formula

$$\tilde{\tilde{f}}(x + tx_1) = \tilde{f}(x) + t\alpha\,.$$

Replacing x and y by $x/(-t)$, $t < 0$ and x/t, $t > 0$ in (11.1), after simple calculations we arrive at

$$\tilde{\tilde{f}}(x) \leq p(x) \quad \text{on } \tilde{\tilde{M}}.$$

Now we want to apply the Hausdorff maximality theorem. Consider the family of pairs $(\tilde{M}, \tilde{f})$, where $\tilde{M} \supset M$ is a subspace of X and $\tilde{f} : \tilde{M} \to \mathbb{R}$ is an extension of f satisfying

$$\tilde{f}(x) \leq p(x) \quad \text{for } x \in \tilde{M}.$$

The family is partially ordered by the relation

$$(\tilde{M}, \tilde{f}) \leq (\tilde{\tilde{M}}, \tilde{\tilde{f}})$$

if $\tilde{M} \subset \tilde{\tilde{M}}$ and $\tilde{\tilde{f}}$ is an extension of $\tilde{f}$. According to the Hausdorff maximality theorem there is a maximal element in the family. Because of the first part of the proof the maximal element must be a functional defined on all of X. Denote it by (X, F). Hence

$$F(x) = f(x) \text{ for } x \in M, \quad F(x) \leq p(x) \text{ for } x \in X.$$

Now it suffices to observe that

$$F(x) = -F(-x) \geq -p(-x)$$

by linearity of F. □

Theorem 11.2 (Hahn-Banach)**.** *Let M be a subspace of a linear space X over $\mathbb{K}$ (= $\mathbb{R}$ of $\mathbb{C}$), and let p be a seminorm*[25] *on X. Let f be a linear functional on M such that*

$$|f(x)| \leq p(x) \quad \textit{for } x \in M.$$

Then there is an extension $F : X \to \mathbb{K}$ of f such that

$$F(x) = f(x) \textit{ for } x \in M; \qquad |F(x)| \leq p(x) \textit{ for } x \in X.$$

[25] i.e. $p(x + y) \leq p(x) + p(y)$, $p(\alpha x) = |\alpha| p(x)$ for $x, y \in X$ and $\alpha \in \mathbb{K}$, see Section 4.2.

Proof. Observe that in Theorem 11.1 we required X to be a real space, but now it can also be a complex one. Actually when $\mathbb{K} = \mathbb{R}$, Theorem 11.2 immediately follows from Theorem 11.1, because any seminorm is Banach functional, so we are left with the case $\mathbb{K} = \mathbb{C}$.

Every linear space over $\mathbb{C}$ can be regarded as a linear space over $\mathbb{R}$. For example $\mathbb{C}^n$ can be regarded as $\mathbb{R}^{2n}$. Thus if X is a linear space over $\mathbb{C}$ we have two kinds of functionals: $\mathbb{C}$-linear and $\mathbb{R}$-linear. More precisely $\mathbb{C}$-linear functionals are

$$\Lambda : X \to \mathbb{C}, \quad \Lambda(ax + by) = a\Lambda x + b\Lambda y \ a, b \in \mathbb{C}$$

and $\mathbb{R}$-linear functionals are

$$\Lambda : X \to \mathbb{R}, \quad \Lambda(ax + by) = a\Lambda x + b\Lambda y \ a, b \in \mathbb{R}\,.$$

The functional $f : M \to \mathbb{C}$ given in the assumptions of the theorem is clearly $\mathbb{C}$-linear and we need find its $\mathbb{C}$-linear extension. The functional $u = \operatorname{re} f : M \to \mathbb{R}$ is $\mathbb{R}$-linear. Note that

$$f(x) = u(x) - iu(ix)\,.$$

Now according to Theorem 11.1, the functional u can be extended to an $\mathbb{R}$-linear functional

$$U : X \to \mathbb{R}, \quad U(x) \le p(x) \text{ for } x \in X,$$

and hence

$$F(x) = U(x) - iU(ix)$$

is a $\mathbb{C}$-linear extension of f. Since for every $x \in X$ there is $\alpha \in \mathbb{C}$ such that $|\alpha| = 1$ and $\alpha F x = |Fx|$ we conclude

$$|Fx| = F(\alpha x) = U(\alpha x) \le p(\alpha x) = p(x)\,.$$

where the second equality follows from the fact that $F(\alpha x) \in \mathbb{R}$ and $U = \operatorname{re} F$. □

Corollary 11.3. *If f is a bounded functional defined on a subspace M of a normed space X, then there is a functional on X, i.e. there is $F \in X^*$ such that*[26]

$$F(x) = f(x) \textit{ for } x \in X \quad \textit{and} \quad \|f\| = \|F\|\,.$$

Proof. $|f(x)| \le \|f\|\,\|x\|$ for $x \in M$. The function $p(x) = \|f\|\,\|x\|$ is a seminorm on X. Hence there is an extension F such that $|F(x)| \le \|f\|\,\|x\|$ for all $x \in X$. Thus $\|F\| \le \|f\|$ and the opposite inequality $\|F\| \ge \|f\|$ is obvious. □

Corollary 11.4. *If X is a normed space and $x_0 \in X$, $x_0 \ne 0$, then there is $x^* \in X^*$ such that*

$$\|x^*\| = 1 \quad \textit{and} \quad \langle x^*, x_0\rangle = \|x_0\|\,.$$

[26] Here $\|f\| = \sup_{\substack{\|x\| \le 1 \\ x \in M}} |f(x)|$ and $\|F\| = \sup_{\substack{\|x\| \le 1 \\ x \in X}} |F(x)|$.

Proof. On a subspace $M = \{\alpha x_0 : \alpha \in \mathbb{K}\}$ we define a functional $f(\alpha x_0) = \alpha\|x_0\|$. Clearly $\|f\| = 1$ and $f(x_0) = \|x_0\|$. Now it suffices to take $x^* = F$, where F in a norm preserving functional from Corollary 11.3. $\square$

The corollary implies that if $X \neq \{0\}$, then $X^* \neq \{0\}$.

Recall that the norm in X^* is defined by

$$\|x^*\| = \sup_{\substack{x\in X\\ \|x\|\leq 1}} \langle x^*, x\rangle\,.$$

As a corollary from the Hahn-Banach theorem we also have

Corollary 11.5. *If X is a normed space and $x \in X$, then*

$$\|x\| = \sup_{\substack{x^*\in X^*\\ \|x^*\|\leq 1}} \langle x^*, x\rangle\,.$$

Theorem 11.6. *Let X_0 be a linear subspace of a normed space X and $x_1 \in X \setminus X_0$. Suppose that the distance of x_1 to X_0 is positive, i.e.*

$$d = \inf\{\|x_1 - x_0\| : x_0 \in X_0\} > 0\,.$$

Then there is a functional $x^* \in X^*$ such that

$$\langle x^*, x_1\rangle = 1, \quad \|x^*\| = \frac{1}{d}, \quad \langle x^*, x_0\rangle = 0 \textit{ for } x_0 \in X_0.$$

Proof. Let $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$ and

$$X_1 = \{x_0 + \alpha x_1 : x_0 \in X_0,\ \alpha \in \mathbb{K}\}.$$

Define a functional on X_1 by

$$\langle x_1^*, x_0 + \alpha x_1\rangle = \alpha\,.$$

For $\alpha \neq 0$ we have

$$\|x_0 + \alpha x_1\| = |\alpha|\left\|x_1 - \Big(-\frac{x_0}{\alpha}\Big)\right\| \geq |\alpha| d = |\langle x_1^*, x_0 + \alpha x_1\rangle|\, d\,,$$

which yields

$$\|x_1^*\| \leq \frac{1}{d}\,. \tag{11.2}$$

Now let $x_0^n \in X_0$ be a sequence such that

$$\|x_0^n - x_1\| \to d\,.$$

We have

$$1 = |\langle x_1^*, x_0^n - x_1\rangle| \leq \|x_1^*\|\,\|x_0^n - x_1\| \to \|x_1^*\|\, d\,,$$

i.e. $\|x_1^*\| \geq 1/d$. This inequality together with (11.2) gives $\|x_1^*\| = 1/d$. Now let x^* be a norm preserving extension of x_1^* to X. Then $\|x^*\| = 1/d$. We have

$$\langle x^*, x_0\rangle = \langle x_1^*, x_0 + 0\cdot x_1\rangle = 0 \quad \text{for } x_0 \in X_0,$$
$$\langle x^*, x_1\rangle = \langle x_1^*, 0 + 1\cdot x_1\rangle = 1\,.$$

The proof is complete. $\square$

Theorem 11.7. *If the dual space X^* to a normed space X is separable, then X is separable too.*

Proof. Let $\{x_i^*\}_{i=1}^\infty \subset X^*$ be a countable and dense subset. For each i we can find $x_i \in X$ such that

$$\|x_i\| \leq 1 \quad \text{and} \quad \langle x_i^*, x_i\rangle \geq \frac{1}{2}\|x_i^*\|\,.$$

Let X_0 be a subspace of X that consists of finite linear combinations of the x_i's. It suffices to prove that X_0 is a dense subset of X (Why?).

Suppose X_0 is not dense in X. Then it follows from the previous theorem that there is a functional $0 \neq x^* \in X^*$ such that $\langle x^*, x\rangle = 0$ for all $x \in X_0$. Let $x_{i_j}^*$ be a sequence such that

$$\|x_{i_j}^* - x^*\| \to 0 \quad \text{as } j \to \infty.$$

We have

$$\begin{aligned} 0 \quad \leftarrow \quad & \|x_{i_j}^* - x^*\| = \sup_{\|x\|\leq 1} |\langle x_{i_j}^* - x^*, x\rangle| \geq \langle x_{i_j}^* - x^*, x_{i_j}\rangle \\ = \quad & \langle x_{i_j}^*, x_{i_j}\rangle \geq \frac{1}{2}\,\|x_{i_j}^*\| \to \frac{1}{2}\,\|x^*\| > 0\,. \end{aligned}$$

Contradiction. □

Example. Every sequence $s = (s_i) \in \ell^1$ defines a functional on ℓ^∞ by

$$\langle x^*, x\rangle = \sum_{i=1}^\infty s_i x_i, \quad \text{where } x = (x_i) \in \ell^\infty. \tag{11.3}$$

It is easy to see (see Exercise following Theorem 2.11) that $\|x^*\| = \|s\|_1$. That means ℓ^1 is isometrically isomorphic to a closed subspace of $(\ell^\infty)^*$. However $\ell^1 \neq (\ell^\infty)^*$ because ℓ^1 is separable and ℓ^∞ is not.

There is, however, a more direct proof of this fact.

$$\langle x^*, x\rangle = \lim_{i\to\infty} x_i, \quad \text{for } x = (x_i) \in c \tag{11.4}$$

is a bounded nonzero functional on $c \subset \ell^\infty$ and hence according to the Hahn-Banach theorem it can be extended to a bounded functional of ℓ^∞. It is clear that this functional cannot be of the form (11.3). Indeed, the value of the functional (11.4) does not change if we change value of any finite number of coordinates of $x = (x_i) \in c$ and this is not true for the functional (11.3). □

As an application of the Hahn-Banach theorem we will prove Theorem 10.11.

Theorem 11.8. *Let M be a closed subspace of a Banach space X.*

(a) *If $\dim M < \infty$, then M is complemented in X.*

(b) *If* $\dim(X/M) < \infty$, *then* M *is complemented in* X.

Proof. (a) Let $e_1, \ldots, e_n$ be a Hamel basis of M. Then every element $x \in M$ can be represented as

$$x = \alpha_1(x)e_1 + \ldots + \alpha_n(x)e_n\,.$$

The coefficients $\alpha_1(x), \ldots \alpha_n(x)$ are linear functionals on M. The space M is isomorphic as a Banach space to[27] $\mathbb{K}^n$ and hence every linear functional on M is continuous. Thus the functions α_i are continuous. By the Hahn-Banach theorem the functionals α_i can be extended to bounded functionals $\alpha_i^* \in X^*$. Let

$$N = \bigcap_{i=1}^n \ker \alpha_i^*\,.$$

It is easy to see that

$$M \cap N = \{0\} \quad \text{and} \quad M + N = X.$$

(b) Let $\pi : X \to X/M$, $\pi(x) = [x]$ be the standard quotient mapping. Let $e_1, \ldots, e_n$ be a basis in X/M. Let $x_i \in X$ be such that $\pi(x_i) = e_i$ and let $N = \operatorname{span}\{x_1, \ldots, x_n\}$. Now it easily follows that

$$M \cap N = \{0\}, \quad M + N = X\,.$$

The proof is complete. $\square$

We close this section with several applications of the Hahn-Banach theorem.

11.1. **Banach limits.** As we have seen in Section 9.3 matrix summability methods provide a way to extend the notion of limit to some sequences that are not necessarily convergent, but is it possible to extend the notion of limit to all bounded sequences? The following result provides a satisfactory answer.

Theorem 11.9 (Mazur)**.** *To each bounded sequence of real numbers* $x = (x_n) \in \ell^\infty$ *we can assign a generalized limit* $\operatorname{LIM}_{n\to\infty} x_n$ *(called* Banach limit*) so that*

(a) *For convergent sequences* $x = (x_n) \in c$

$$\operatorname*{LIM}_{n\to\infty} x_n = \lim_{n\to\infty} x_n\,;$$

(b) *For all* $x = (x_n), y = (y_n) \in \ell^\infty$ *and* $a, b \in \mathbb{R}$

$$\operatorname*{LIM}_{n\to\infty}(ax_n + by_n) = a \operatorname*{LIM}_{n\to\infty} x_n + b \operatorname*{LIM}_{n\to\infty} y_n\,;$$

(c) *For all* $x = (x_n) \in \ell^\infty$ *and all* $k \in \mathbb{N}$

$$\operatorname*{LIM}_{n\to\infty} x_{n+k} = \operatorname*{LIM}_{n\to\infty} x_n\,;$$

[27] $\mathbb{R}^n$ or $\mathbb{C}^n$.

(d) *For all* $x = (x_n) \in \ell^\infty$

$$\liminf_{n\to\infty} x_n \le \operatorname*{LIM}_{n\to\infty} x_n \le \limsup_{n\to\infty} x_n \,.$$

Proof. It is easy to see that

$$p : \ell^\infty \to \mathbb{R}, \qquad p(x) = \limsup_{n\to\infty} \frac{x_1 + x_2 + \ldots + x_n}{n}$$

is a Banach functional. Let $M \subset \ell^\infty$ be a linear subspace consisting of sequences for which the limit

$$\ell(x) = \lim_{n\to\infty} \frac{x_1 + x_2 + \ldots + x_n}{n} \in \mathbb{R}$$

exists. Clearly $c \subset M$ and $\ell : M \to \mathbb{R}$ is a linear functional such that

$$\ell(x) = p(x) \le p(x) \quad \text{for } x \in M.$$

According to the Hahn-Banach theorem ℓ can be extended to a linear functional on ℓ^∞ denoted by $\operatorname{LIM}_{n\to\infty} x_n$ such that

$$-p(-x) \le \operatorname*{LIM}_{n\to\infty} x_n \le p(x) \quad \text{for all } x \in \ell^\infty. \tag{11.5}$$

Properties (a) and (b) are obvious. It is easy to see that

$$p(x) \le \limsup_{n\to\infty} x_n, \qquad \liminf_{n\to\infty} x_n = -\limsup_{n\to\infty}(-x_n) \le -p(-x)\,.$$

The two inequalities combined with (11.5) yield (d). We are left with the proof of (c). We have

$$\operatorname*{LIM}_{n\to\infty} x_{n+1} - \operatorname*{LIM}_{n\to\infty} x_n = \operatorname*{LIM}_{n\to\infty} (x_{n+1} - x_n)\,.$$

Since

$$\begin{aligned} p((x_{n+1} - x_n)) &= \limsup_{n\to\infty} \frac{(x_2 - x_1) + (x_3 - x_2) + \ldots + (x_{n+1} - x_n)}{n} \\ &= \limsup_{n\to\infty} \frac{x_{n+1} - x_1}{n} = 0\,, \end{aligned}$$

and similarly $p(x_n - x_{n+1}) = 0$, inequality (11.5) yields

$$\operatorname*{LIM}_{n\to\infty} x_{n+1} = \operatorname*{LIM}_{n\to\infty} x_n$$

and (c) follows by induction. □

This result can be generalized to the class of bounded real-valued functions on $[0,\infty)$. Let $B\,[0,\infty)$ be the class of all bounded functions $f : [0,\infty) \to \mathbb{R}$ (no measurability condition). $B\,[0,\infty)$ is a Banach space with respect to the supremum norm[28] $\|\cdot\|_\infty$.

Theorem 11.10 (Banach)**.** *In the space* $B\,[0,\infty)$ *there is a functional* $\operatorname{LIM}_{t\to\infty} x(t)$ *such that*

[28]In the notation used in Section 5.1, $B\,[0,\infty) = \ell^\infty([0,\infty))$.

(a) *If the limit* $\lim_{t\to\infty} x(t)$ *exists, then*

$$\lim_{t\to\infty} x(t) = \operatorname*{LIM}_{t\to\infty} x(t)\,;$$

(b)

$$\operatorname*{LIM}_{t\to\infty}(ax(t)+by(t)) = a\operatorname*{LIM}_{t\to\infty} x(t) + b\operatorname*{LIM}_{t\to\infty} y(t)\,;$$

(c)

$$\operatorname*{LIM}_{t\to\infty} x(t+\tau) = \operatorname*{LIM}_{t\to\infty} x(t) \quad \textit{for every } \tau > 0\,;$$

(d)

$$\liminf_{t\to\infty} x(t) \le \operatorname*{LIM}_{t\to\infty} x(t) \le \limsup_{t\to\infty} x(t)\,.$$

Note that Mazur's theorem follows from that of Banach. Indeed, if $x = (x_n) \in \ell^\infty$, it suffices to apply the Banach theorem to the function

$$x(t) = \sum_{n=1}^{\infty} x_n \chi_{[n-1,n)}$$

and set

$$\operatorname*{LIM}_{n\to\infty} x_n := \operatorname*{LIM}_{t\to\infty} x(t)\,.$$

Proof of Theorem 11.10. The subspace $M \subset B\,[0,\infty)$ consisting of functions for which the limit

$$\lim_{t\to\infty} x(t) \tag{11.6}$$

exists is a linear subspace and the limit (11.6) is a linear functional on M. We want to extend it to an appropriate functional on $B\,[0,\infty)$ and the main difficulty is a construction of a suitable Banach functional.

For $x \in B\,[0,\infty)$ and $t_1,\ldots,t_n \ge 0$ we define

$$\beta(x;t_1,\ldots,t_n) = \limsup_{t\to\infty} \frac{1}{n}\sum_{i=1}^{n} x(t+t_i)$$

and then

$$p(x) = \inf\left\{\beta(x;t_1,\ldots,t_n) :\, t_i \ge 0,\, n \in \mathbb{N}\right\}.$$

We will prove that $p(x)$ is a Banach functional. Clearly $p(tx) = tp(x)$ for $t > 0$ and we only need to prove subadditivity $p(x+y) \le p(x)+p(y)$.

Let $t_1,\ldots,t_n \ge 0$ and $s_1,\ldots,s_m \ge 0$. Then $t_i + s_j$ is a collection of nm numbers. Denote these numbers by $u_1,\ldots,u_{nm}$. It it easy to see that the subadditivity of p follows from the lemma.

Lemma 11.11.

$$\beta(x+y;u_1,\ldots,u_{nm}) \le \beta(x;t_1,\ldots,t_n) + \beta(y;s_1,\ldots,s_m)\,.$$

Proof. We have

$$\frac{1}{nm}\sum_{k=1}^{nm}(x+y)(t+u_k) = \frac{1}{m}\sum_{j=1}^{m}\frac{1}{n}\sum_{i=1}^{n}x(t+t_i+s_j)+\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m}\sum_{j=1}^{m}y(t+t_i+s_j)\,.$$

Applying $\limsup_{t\to\infty}$ to both sides yields

$$\begin{aligned}\beta(x+y;u_1,\ldots,u_{nm}) &\le \frac{1}{m}\sum_{j=1}^{m}\underbrace{\limsup_{t\to\infty}\frac{1}{n}\sum_{i=1}^{n}x(t+t_i+s_j)}_{\beta(x;t_1,\ldots,t_n)\text{ because }t+s_j\to\infty} \\ &+ \frac{1}{n}\sum_{i=1}^{n}\underbrace{\limsup_{t\to\infty}\frac{1}{m}\sum_{j=1}^{m}y(t+t_i+s_j)}_{\beta(y;s_1,\ldots,s_m)\text{ because }t+t_i\to\infty} \\ &= \beta(x;t_1,\ldots,t_n)+\beta(y;s_1,\ldots,s_m)\,.\end{aligned}$$

This completes the proof of the lemma and hence that of the fact that $p(x)$ is a Banach functional. $\square$

If $x \in M \subset B(0,\infty)$, then

$$\lim_{t\to\infty} x(t) = p(x) \le p(x)$$

and hence according to the Hahn-Banach theorem the functional $x \mapsto \lim_{t\to\infty} x(t)$ extends from M to a linear functional on $B\,[0,\infty)$ that we denote by $\mathrm{LIM}_{\,t\to\infty}\, x(t)$ such that

$$-p(-x) \le \mathop{\mathrm{LIM}}_{t\to\infty} x(t) \le p(x) \quad \text{for all } x \in B\,[0,\infty).$$

It is easy to see that $p(x) \le \limsup_{t\to\infty} x(t)$ and hence[29]

$$\liminf_{t\to\infty} x(t) \le \mathop{\mathrm{LIM}}_{t\to\infty} x(t) \le \limsup_{t\to\infty} x(t)\,.$$

It remains to show that

$$\mathop{\mathrm{LIM}}_{t\to\infty} x_\tau(t) = \mathop{\mathrm{LIM}}_{t\to\infty} x(t)$$

where $x_\tau(t) = x(t+\tau)$. Let $y = x_\tau - x$. Then

$$p(y) \le \beta(y;0,\tau,2\tau,\ldots,(n-1)\tau) = \limsup_{t\to\infty}\frac{1}{n}(x(t+n\tau)-x(t)) \le \frac{2}{n}\|x\|_\infty\,.$$

Since this inequality holds for any $n \in \mathbb{N}$, we have $p(y) \le 0$. Similarly $p(-y) \le 0$ and hence

$$0 \le -p(-y) \le \mathop{\mathrm{LIM}}_{t\to\infty}(x_\tau - x) \le p(y) \le 0\,,$$

so $\mathrm{LIM}_{\,t\to\infty}(x_\tau - x) = 0$, $\mathrm{LIM}_{\,t\to\infty}\, x_\tau(t) = \mathrm{LIM}_{\,t\to\infty}\, x(t)$. The proof is complete. $\square$

[29]Because $\liminf_{t\to\infty} x(t) = -\limsup_{t\to\infty}(-x(t)) \le -p(-x)$.

11.2. **Finitely additive measures.** In this section we will prove the following surprising result.

Theorem 11.12 (Banach)**.** *There is a* finitely *additive measure* $\mu : 2^{\mathbb{R}} \to [0,\infty]$*, i.e.*
$$\mu(A\cup B) = \mu(A) + \mu(B) \quad \textit{for } A, B \subset \mathbb{R},\ A\cap B = \emptyset$$
defined on all *subsets of* $\mathbb{R}$ *and such that*

(a) μ *is invariant with respect to isometries, i.e.*
$$\mu(A+t) = \mu(A), \quad \mu(-A) = \mu(A)$$
for all $A\subset\mathbb{R}$*,* $t\in\mathbb{R}$*.*

(b)
$$\mu(A) = \mathcal{L}^1(A)$$
for all Lebesgue measurable sets $A\subset\mathbb{R}$*.*

Banach proved that the above result holds also in $\mathbb{R}^2$, i.e. there is a finitely additive measure
$$\mu : 2^{\mathbb{R}^2} \to [0,\infty]$$
defined on *all* subsets of $\mathbb{R}^2$, invariant under isometries of $\mathbb{R}^2$ and equal to the Lebesgue measure on the class of Lebesgue measurable sets. According to the Vitali example such measures cannot be, however, countably additive.

Surprisingly, the Banach theorem does not hold in $\mathbb{R}^n$, $n\geq 3$. This is related to an algebraic fact that the group of isometries of $\mathbb{R}^n$, $n\geq 3$ contains a free supgroup of rank 2. Namely Banach and Tarski proved[30] that the unit ball in $\mathbb{R}^3$ can be decomposed into a finite number of disjoint sets (later it was shown that it suffices to take 5 sets)
$$B^3(0,1) = A_1\cup A_2\cup A_3\cup A_4\cup A_5$$
is a way that there are isometries $\tau_1,\dots,\tau_5$ of $\mathbb{R}^3$ such that
$$B^3(0,1) = \tau_1(A_1)\cup\tau_2(A_2), \quad B^3(0,1) = \tau_3(A_3)\cup\tau_4(A_4)\cup\tau_5(A_5)$$
and a similar decomposition is possible for $B^n(0,1)$ for any $n\geq 3$.

Clearly the sets A_i cannot be Lebesgue measurable, because we would obtain a contradiction by comparing volumes. This also shows that there is no finitely additive measure in $\mathbb{R}^n$, $n\geq 3$ invariant under isometries and equal to the Lebesgue measure on the class of Lebesgue measurable sets.

The Banach theorem will follow from a somewhat stronger result. As in the case of Fourier series we can identify bounded functions on $\mathbb{R}$ with period 1 with bounded functions on S^1 via the exponential mapping $t\mapsto e^{2\pi i t}$. Thus the integral of a function on S^1 corresponds to the integral $\int_0^1 x(t)\, dt$ of a function $x:\mathbb{R}\to\mathbb{R}$ with period 1.

[30]See S. Wagon, *The Banach-Tarski paradox*, Cambridge Univ. Press 1999.

The space of all (not necessarily measurable) real-valued functions on $\mathbb{R}$ with period 1 will be denoted by $B(S^1)$. Since every bounded function on $\mathbb{R}$ with period 1 is uniquely determined by a bounded function on $[0,1)$ we can further identify $B(S^1)$ with bounded real-valued functions on $[0,1)$.

Theorem 11.13 (Banach)**.** *In the space $B(S^1)$ there is a linear functional denoted by $\int x(t)\,dt$ such that*

(a)
$$\int \big(ax(t)+by(t)\big)\,dt = a\int x(t)\,dt + b\int y(t)\,dt\,;$$

(b)
$$\int x(t)\,dt \geq 0 \quad \textit{if } x \geq 0$$

(c)
$$\int x(t+\tau)\,dt = \int x(t)\,dt \quad \textit{for all } \tau\in\mathbb{R};$$

(d)
$$\int x(1-t)\,dt = \int x(t)\,dt\,;$$

(e) *For any Lebesgue measurable function $x\in B(S^1)$*
$$\int x(t)\,dt = \int_0^1 x(t)\,dt\,,$$
where on the right hand side we have the integral with respect to the Lebesgue measure.

Before we prove this theorem we show how to conclude Theorem 11.12.

Proof of Theorem 11.12. Theorem 11.13 defines a generalized integral of a bounded function on $[0,1)$. For $A\subset\mathbb{R}$ and $k\in\mathbb{Z}$ let
$$x_k = \chi_{(A-k)\cap[0,1)}\,.$$
Thus $\int x_k(t)\,dt$ corresponds to a generalized measure of the set $A\cap[k,k+1)$ and it is natural to define
$$\mu(A) = \sum_{k=-\infty}^{\infty}\int x_k(t)\,dt\,.$$
It is easy to check now that μ satisfies the claim of Theorem 11.12. $\square$

Proof of Theorem 11.13. For $x\in B(S^1)$ and $t_1,\dots,t_n\geq 0$ we define
$$\beta(x;t_1,\dots,t_n) = \sup\Big\{\frac{1}{n}\sum_{i=1}^{n} x(t+t_i)\,:\, t\in\mathbb{R}\Big\}$$
and then
$$p(x) = \inf\big\{\beta(x;t_1,\dots,t_n)\,:\, t_i\geq 0,\, n\in\mathbb{N}.\big\}$$

As in the proof of Theorem 11.10 one can show that $p(x)$ is a Banach functional such that

$$p(x_\tau - x) \leq 0, \quad p(x - x_\tau) \leq 0\,. \tag{11.7}$$

For a Lebesgue measurable function $x \in B(S^1)$ we define a functional

$$\langle x_1^*, x\rangle = \int_0^1 x(t)\, dt\,.$$

It is easy to see that

$$\langle x_1^*, x\rangle \leq p(x)$$

and hence x_1^* can be extended to a functional x^* on $B(S^1)$ such that

$$-p(-x) \leq \langle x^*, x\rangle \leq p(x) \quad \text{for all } x \in B(S^1).$$

Finally we define

$$\int x(t)\, dt = \frac{1}{2}\langle x^*, x + x_-\rangle\,,$$

where $x_-(t) = x(1-t)$. The properties (a), (d) and (e) are obvious. Property (b) follows from the observation that if $x \geq 0$, then $p(-x) \leq 0$ and hence

$$0 \leq -p(-x) \leq \langle x^*, x\rangle, \quad 0 \leq -p(-x_-) \leq \langle x^*, x_-\rangle\,.$$

Finally, property (c) follows from the inequality (11.7). □

11.3. **Runge's theorem.** Let $\Omega = B^2(0,2) \setminus \overline{B^2}(0,1)$ be an annulus. The function $f(z) = z^{-1}$ is holomorphic in Ω, but it cannot be uniformly approximated by polynomials on on compact subsets of Ω. Indeed, if γ is a positively oriented circle inside Ω centered at 0, then

$$\int_\gamma z^{-1}\, dz = 2\pi i$$

and the corresponding integral for any complex polynomial equals 0. Note that the domain Ω is not simply connected. Hence there is no complex version of the Weierstrass approximation theorem. However, we have

Theorem 11.14 (Runge)**.** *If $\Omega \subset \mathbb{C}$ is simply connected, then every holomorphic function on Ω can be uniformly approximated on compact subsets of Ω by complex polynomials.*

Proof. Since Ω is simply connected, every compact subset of Ω is contained in a simply connected compact set $K \subset \Omega$, so the complement $\mathbb{C} \setminus K$ is connected. Let γ be a positively oriented Jordan curve in Ω such that K is in the interior of γ. If $f \in H(\Omega)$, then

$$f(z) = \frac{1}{2\pi i}\int_\gamma \frac{f(\xi)}{\xi - z}\, d\xi, \quad z \in K\,.$$

The integral on the right hand side can be uniformly approximated by Riemann sums

$$\frac{1}{2\pi i}\int_\gamma \frac{f(\xi)}{\xi - z}\,d\xi \approx \frac{1}{2\pi i}\sum_{k=1}^{m}\frac{f(\xi_k)}{\xi_k - z}\,\Delta\xi_k$$

which is a linear combination of functions

$$z \mapsto \frac{1}{\xi_k - z} \in C(K)\,.$$

Thus it remains to prove that every function

$$z \mapsto \frac{1}{\xi_0 - z} \in C(K), \quad \xi_0 \notin K \tag{11.8}$$

can be uniformly approximated on K by complex polynomials. Let $M \subset C(K)$ be the closure of the subspace of complex polynomials. Suppose that the function (11.8) does not belong to M. According to Theorem 11.6 there is a functional $x^* \in C(K)^*$ which vanishes on M and is nonzero on the function (11.8), i.e.

$$\langle x^*, p(z)\rangle = 0, \quad p \text{ any complex polynomial},$$

$$\langle x^*, \frac{1}{\xi_0 - z}\rangle \neq 0\,. \tag{11.9}$$

The function

$$\xi \mapsto \langle x^*, \frac{1}{\xi - z}\rangle \tag{11.10}$$

is complex differentiable in $\mathbb{C}\setminus K$, so it is holomorphic in $\mathbb{C}\setminus K$. If $|\xi| > R = \max_{z\in K}|z|$, then

$$\frac{1}{\xi - z} = \frac{1}{\xi}\,\frac{1}{1 - z/\xi} = \sum_{n=0}^{\infty}\frac{z^n}{\xi^{n+1}}$$

and hence the function

$$z \mapsto \frac{1}{\xi - z}$$

can be uniformly approximated in $C(K)$ by complex polynomials, so

$$\langle x^*, \frac{1}{\xi - z}\rangle = 0 \quad \text{if } |\xi| > R.$$

Since the function (11.10) is homomorphic in $\mathbb{C}\setminus K$, $\mathbb{C}\setminus K$ is connected and it vanishes for ξ sufficiently large

$$\langle x^*, \frac{1}{\xi - z}\rangle = 0 \quad \text{for all } \xi \in \mathbb{C}\setminus K$$

which contradicts (11.9). $\square$

11.4. **Separation of convex sets.** Let X be a linear space over $\mathbb{K}$.[31]

DEFINITION. We say that a convex set $W \subset X$ is *absorbing* if for each $x \in X$ there is $\varepsilon > 0$ such that the whole interval $\{tx : t \in [0, \varepsilon]\} \subset W$ is contained in W.[32]

We say that W is *absorbing at* $a \in W$ if $W - a$ is absorbing. Equivalently if for every $x \in X$ there is $\varepsilon > 0$ such that $\{a + tx : t \in [0, \varepsilon]\} \subset W$.

We say that W is *balanced* if $\alpha x \in W$ for all $|\alpha| \leq 1$ whenever $x \in W$.

Clearly $0 \in W$ if W is absorbing or balanced. If $\mathbb{K} = \mathbb{R}$, then W is balanced if it is central symmetric, i.e. symmetric with respect to 0 and if $\mathbb{K} = \mathbb{C}$, then W is balanced if the set contains the whole disc $\{\alpha x : |\alpha| \leq 1\}$ in the complex plane generated by x, span $\{x\}$.

The above definitions are are motivated by the following obvious result:

Proposition 11.15. *If p is a seminorm on X, then the unit ball $W = \{x : p(x) < 1\}$ is convex, balanced and absorbing at any point of W.*

In Theorem 11.17 we will show that the conditions from the proposition characterize the unit ball for a seminorm.

DEFINITION. For each convex and absorbing set $W \subset X$ the *Minkowski functional* is

$$\mu_W(x) = \inf \left\{ s > 0 : \frac{x}{s} \in W \right\}.$$

Theorem 11.16. *Let $W \subset X$ be convex and absorbing. The Minkowski functional has the following properties*

(a) $\mu_W(x) \geq 0$;
(b) $\mu_W(x + y) \leq \mu_W(x) + \mu_W(y)$;
(c) $\mu_W(tx) = t\mu_W(x)$ *for all* $t \geq 0$;
(d) $\mu_W(x) = 0$ *if and only if* $\{tx : t \geq 0\} \subset W$.
(e) *If* $x \in W$, *then* $\mu_W(x) \leq 1$;
(f) *If* $\mu_W(x) < 1$, *then* $x \in W$
(g) *If* $\mu_W(x) > 1$, *then* $x \notin W$.

Proof. All the properties but (b) are obvious. To prove (b) let $x, y \in X$ and $s_1, s_2 > 0$ be such that $x/s_1, y/s_2 \in W$. It follows from the convexity of W that

$$\frac{x + y}{s_1 + s_2} = \frac{s_1}{s_1 + s_2} \, \frac{x}{s_1} + \frac{s_2}{s_1 + s_2} \, \frac{y}{s_2} \in W$$

Hence $\mu_W(x + y) \leq s_1 + s_2$ and the claim follows upon taking the infimum. □

[31] As always $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$.

[32] Equivalently a convex set $W \subset X$ is absorbing if $X = \bigcup_{n=1}^{\infty} nW$.

Note that the properties (b) and (c) show that the Minkowski functional is a Banach functional.

The following result provides a geometric characterization of sets that are unit balls for norms and seminorms.

Theorem 11.17. *If $W \subset X$ is convex, balanced and absorbing at any point of W, then there is a unique seminorm p such that $W = \{x \in W : p(x) < 1\}$. Moreover p is a norm if and only if for each $x \neq 0$ there is $t > 0$ such that $tx \notin W$.*

Proof. Let $p = \mu_W$. First we will prove that p is a seminorm. The inequality $p(x+y) \leq p(x) + p(y)$ follows from the previous result. Let now $x \in X$, $\alpha \in \mathbb{K}$. If $\alpha = 0$, then $p(\alpha x) = |\alpha| p(x)$, so we can assume $\alpha \neq 0$. If $|\alpha| = 1$, then for $s > 0$, $x/s \in W$ if and only if $\alpha x/s \in W$, because W is balanced and hence

$$p(\alpha x) = p(x) \quad \text{for } |\alpha| = 1\,. \tag{11.11}$$

If $\alpha \neq 0$ is arbitrary, then

$$p(\alpha x) = p\Big(\frac{\alpha}{|\alpha|}\,|\alpha| x\Big) = p(|\alpha| x) = |\alpha| p(x)$$

by (11.11) and Theorem 11.16(c). This proves that p is a seminorm. According to Theorem 11.16(e,f)

$$\{x : p(x) < 1\} \subset W \subset \{x : p(x) \leq 1\}\,. \tag{11.12}$$

Let $x \in W$. Since W is absorbing at x there is $t > 0$ such that $x + tx = y \in W$ and thus $x = (1+t)^{-1} y$. Since $y \in W$, $p(y) \leq 1$, so

$$p(x) = (1+t)^{-1} p(y) \leq (1+t)^{-1} < 1\,.$$

This inequality together with (11.12) proves that

$$W = \{x : p(x) < 1\}\,.$$

Uniqueness is easy, because if q is another seminorm such that $W = \{x : q(x) < 1\}$, then it is easy to show that

$$q(x) = \inf\Big\{s > 0 : \frac{x}{s} \in W\Big\} = p(x)\,.$$

Now it follows from condition (d) of Theorem 11.16 that $p(x)$ is a norm, i.e. $p(x) \neq 0$ for $x \neq 0$ is and only if for each $x \neq 0$ there is $t > 0$ such that $tx \notin W$. □

Now we will use the fact that the Minkowski functional is a Banach functional and we will prove results about separation of convex sets known also as geometric Hahn-Banach theorems.

Observe that the Hahn-Banach theorem in a form involving a Banach functional applies only to real linear spaces. However, if X is a complex linear space, the real part of a functional is $\mathbb{R}$-linear and we can still apply the Hahn-Banach theorem (cf. Theorem 11.2).

Theorem 11.18. *Let W_1, W_2 be disjoint convex subsets of a normed space X and assume that W_1 is open. Then there is a functional $x^* \in X^*$ and $c \in \mathbb{R}$ such that*

$$\operatorname{re}\langle x^*, x\rangle < c \leq \operatorname{re}\langle x^*, y\rangle$$

for all $x \in W_1$ and $y \in W_2$.

Proof. It suffices to prove the theorem in the case of real normed spaces. Indeed, if X is a complex normed space, then it can still be regarded as a real normed space. If we can find a bounded $\mathbb{R}$-linear functional $\tilde{x}^*$ such that

$$\langle \tilde{x}^*, x\rangle < c \leq \langle \tilde{x}^*, y\rangle$$

for all $x \in W_1$ and $y \in W_2$, then

$$\langle x^*, x\rangle = \langle \tilde{x}^*, x\rangle - i\langle \tilde{x}^*, ix\rangle$$

is a bounded $\mathbb{C}$-linear functional that satisfies $\operatorname{re}\langle x^*, x\rangle = \langle \tilde{x}^*, x\rangle$.

Henceforth we will assume that X is a real normed space. Fix $x_1 \in W_1$ and $x_2 \in W_2$. Then the set

$$W = W_1 - W_2 + \underbrace{x_2 - x_1}_{x_0}$$

is open, convex and $0 \in W$. Since W is open it is absorbing and hence we can define the Minkowski functional μ_W for W.

Define a functional x_1^* on the one dimensional space $\{\alpha x_0 : \alpha \in \mathbb{R}\}$ by

$$\langle x_1^*, \alpha x_0\rangle = \alpha\,.$$

Since the sets W_1 and W_2 are disjoint, $x_0 = x_2 - x_1 \notin W$. Thus $\mu_W(x_0) \geq 1$ and hence it easily follows that

$$\langle x_1^*, \alpha x_0\rangle \leq \mu_W(\alpha x_0)$$

for all $\alpha \in \mathbb{R}$. Now the fact that μ_W is a Banach functional allows us to apply the Hahn-Banach theorem according to which there is a functional x^* being an extension of x_1^* such that

$$\langle x^*, x\rangle \leq \mu_W(x) \quad \text{for all } x \in X.$$

The functional x^* is bounded because W is open. Indeed, if $B(0,\varepsilon) \subset W$, then for $\|x\| < \varepsilon$, $\langle x^*, x\rangle \leq \mu_W(x) \leq 1$, so $\|x^*\| \leq \varepsilon^{-1}$.

For any $x \in W_1$, $y \in W_2$, $x - y + x_0 \in W$ and thus

$$\mu_W(x - y + x_0) < 1\,,$$

because W is open. Using $\langle x^*, x_0\rangle = 1$ we obtain

$$\langle x^*, x - y + x_0\rangle = \langle x^*, x\rangle - \langle x^*, y\rangle + 1 \leq \mu_W(x - y + x_0) < 1\,,$$

so

$$\langle x^*, x\rangle < \langle x^*, y\rangle \quad \text{for all } x \in W_1,\ y \in W_2. \tag{11.13}$$

The sets $x^*(W_1)$ and $x^*(W_2)$ are intervals in $\mathbb{R}$ as convex sets. Moreover $x^*(W_1)$ is open, since W_1 is open.[33] Therefore (11.13) yields the existence of $c \in \mathbb{R}$ such that $x^*(W_1) \subset (-\infty, c)$, $x^*(W_2) \subset [c, \infty)$. □

Theorem 11.19. *Let W_1, W_2 be disjoint convex subsets of a normed space X. Assume that W_1 is compact and W_2 is closed. Then there is $x^* \in X^*$, $c \in \mathbb{R}$ and $\varepsilon > 0$ such that*

$$\mathrm{re}\,\langle x^*, x\rangle \le c - \varepsilon < c \le \mathrm{re}\,\langle x^*, y\rangle$$

for all $x \in W_1$ and $y \in W_2$.

Proof. The distance between W_1 and W_2 is positive

$$\mathrm{dist}\,(W_1, W_2) = \inf\{d(x, y) : x \in W_1,\ y \in W_2\} > 0\,,$$

because W_1 is compact and W_2 is closed. Hence there is $\delta > 0$ such that

$$W_1^\delta = W_1 + B(0, \delta) = \{x + y : x \in W_1,\ \|y\| < \delta\}$$

is disjoint with W_2. Clearly W_1^δ is open and convex so we can apply previous result to the pair of convex sets W_1^δ, W_2. Therefore

$$\mathrm{re}\,\langle x^*, x\rangle < c \le \mathrm{re}\,\langle x^*, y\rangle$$

for all $x \in W_1^\delta$, $y \in W_2$ and some $c \in \mathbb{R}$. Since W_1 is compact, the set

$$\{\mathrm{re}\,\langle x^*, x\rangle : x \in W_1\}$$

is a compact subset of $(-\infty, c)$, so it is contained in $(-\infty, c - \varepsilon]$ for some $\varepsilon > 0$. □

Corollary 11.20. *Let W be a closed convex subset of a real normed space X. Then W is the intersection of all closed half-spaces $\{x : \langle x^*, x\rangle \ge c\}$, $x^* \in X^*$, $c \in \mathbb{R}$ that contain W.*

Proof. For $x \notin W$ we set $W_1 = \{x\}$, $W_2 = W$ and apply Theorem 11.19. □

11.5. **Convex hull.**

Definition. Let X be a vector space and $E \subset X$. The *convex hull* $co(E)$ of E is the intersection of all convex sets that contain E. In other words $co(E)$ is the smallest convex set that contains E.

If X is a normed space and $E \subset X$, then it is easy to see that the closure $\overline{co}(E)$ is convex. This set is called the *closed convex hull* of E and it is obvious that $\overline{co}(E)$ is the smallest closed convex set that contains E.

[33] $x^*(B(x, r))$ is an open interval centered at $\langle x^*, x\rangle$.

Theorem 11.21 (Mazur)**.** *If X is a Banach space and $K \subset X$ is compact, then $\overline{co}(K)$ is compact.*

Proof. A metric space is compact if and only if it is complete and totally bounded.[34] Clearly $\overline{co}(K)$ is complete as a closed subset of a complete metric space, and it remains to prove that it is totally bounded. Since in any metric space the closure of a totally bounded set is totally bounded it suffices to prove that $co(K)$ is totally bounded. Fix $\varepsilon > 0$. Since K is compact, there is a finite covering of K by balls of radius $\varepsilon/2$.

$$\begin{aligned} K &\subset B(x_1, \varepsilon/2) \cup \ldots \cup B(x_n, \varepsilon/2) = \{x_1, \ldots, x_n\} + B(0, \varepsilon/2) \\ &\subset co(\{x_1, \ldots, x_n\}) + B(0, \varepsilon/2)\,. \end{aligned}$$

The set on the right hand side is convex, so

$$co(K) \subset co(\{x_1, \ldots, x_n\}) + B(0, \varepsilon/2)\,.$$

Note that the set $co(\{x_1, \ldots, x_n\})$ is totally bounded as a bounded set in a finitely dimensional space span $\{x_1, \ldots, x_n\}$ and hence it has a finite covering by balls of radius $\varepsilon/2$

$$co(\{x_1, \ldots, x_n\}) \subset \{y_1, \ldots, y_m\} + B(0, \varepsilon/2)$$

and hence

$$co(K) \subset \{y_1, \ldots, y_m\} + B(0, \varepsilon/2) + B(0, \varepsilon/2) = B(y_1, \varepsilon) \cup \ldots \cup B(y_m, \varepsilon)$$

which proves that $co(K)$ is totally bounded. □

12. Banach space valued integration

The existence of vector valued integrals is described in the next theorem.

Theorem 12.1. *Let μ be a probability Borel measure on a compact metric space E and let $f : E \to X$ be a continuous function with values in a Banach space X. Then there is a unique element $y \in X$ such that*

$$\langle x^*, y \rangle = \int_E \langle x^*, f \rangle \, d\mu \quad \textit{for every } x^* \in X^*. \tag{12.1}$$

Moreover $y \in \overline{co}(f(E))$.

Observe that $z \mapsto \langle x^*, f(z) \rangle$ is a continuous scalar valued function on E, so the integrals on the right hand side of (12.1) are well defined.

The vector y from the above theorem is denoted by

$$y = \int_E f \, d\mu$$

[34] A metric space is totally bounded if for every $\varepsilon > 0$ there is a finite covering of the space by balls of radius ε.

and hence we can rewrite (12.1) as

$$\left\langle x^*, \int_E f\, d\mu \right\rangle = \int_E \langle x^*, f\rangle\, d\mu\,. \tag{12.2}$$

Moreover

$$\int_E f\, d\mu \in \overline{co}(f(E))\,. \tag{12.3}$$

If μ is a finite Borel measure, then up to a constant factor it is a probability measure and hence we can define the integrals of vector valued functions for any finite measures. However, if μ is not a probability measure, then, in general, the property (12.3) will be lost.

Proof. Uniqueness of y is obvious, because if $\langle x^*, y_1\rangle = \langle x^*, y_2\rangle$ for all $x^* \in X^*$, then $y_1 = y_2$, so we are left with the proof of the existence.

It follows from Mazur's theorem that the set $K = \overline{co}(f(E))$ is compact. We have to find $y \in K$ such that (12.1) is satisfied. For each finite subset $F \subset X^*$ let

$$K_F = \left\{ y \in K : \langle x^*, y\rangle = \int_E \langle x^*, f\rangle\, d\mu \text{ for all } x^* \in F \right\}.$$

K_F is a closed subset of K, so it is compact. We have to prove that the intersection of sets K_F over all finite subsets $F \subset X^*$ is nonempty. To this end it suffices to prove that each of the sets K_F in nonempty, because the intersection of a finite number of sets K_F is also a set of the same type and the rest follows from the well known lemma.

Lemma 12.2. *If $\{K_i\}_{i\in I}$ is a collection of compact subsets of a metric space Z such that the intersection of every finite subcollection of $\{K_i\}_{i\in I}$ is nonempty, then $\bigcap_{i\in I} K_i$ is nonempty.*

Proof. Suppose $\bigcap_{i\in I} K_i = \emptyset$. Let $G_i = Z \setminus K_i$. Fix $j \in I$. Then $K_j \cap \bigcap_{i\in I} K_i = \emptyset$, so $K_j \subset \bigcup_{i\in I} G_i$. This is an open covering of a compact set K_j, so there is a finite subcovering $K_j \subset G_{i_1} \cup \ldots \cup G_{i_n}$, i.e. $K_j \cap K_{i_1} \cap \ldots \cap K_{i_n} = \emptyset$ which is a contradiction with our assumptions. □

Let $F = \{x_1^*, \ldots, x_n^*\} \subset X^*$. We have to prove that $K_F \neq \emptyset$, i.e. that there is $y \in K = \overline{co}(f(E))$ such that

$$\langle x_i^*, y\rangle = \int_E \langle x_i^*, f\rangle\, d\mu \quad \text{for } i = 1, 2, \ldots, n. \tag{12.4}$$

Let

$$L : X \to \mathbb{K}^n, \quad L(x) = \big(\langle x_1^*, x\rangle, \ldots, \langle x_n^*, x\rangle\big)$$

and let $W = L(f(E))$. Since L is linear the set $L(K)$ is compact, convex and it contains W, so $\overline{co}(W) \subset L(K)$. It suffices to prove that

$$z = (z_1, \ldots, z_n) := \left(\int_E \langle x_1^*, f\rangle\, d\mu, \ldots, \int_E \langle x_n^*, f\rangle\, d\mu \right) \in \overline{co}(W)\,.$$

Indeed, since $\overline{co}(W) \subset L(K)$ it will imply that $z = L(y)$ for some $y \in K$ which is (12.4).

Suppose $z \notin \overline{co}(W)$. The sets $\{z\}$ and $\overline{co}(W)$ are convex, compact and disjoint, so according to Theorem 11.19 there is a functional in $(\mathbb{K}^n)^*$ that separates the two sets, i.e. there is $(c_1, \ldots, c_n) \in \mathbb{K}^n$ such that

$$\operatorname{re} \sum_{i=1}^{n} c_i t_i < \operatorname{re} \sum_{i=1}^{n} c_i z_i \quad \text{for all } t = (t_1, \ldots, t_n) \in \overline{co}(W).$$

If $t = L(f(s))$, $s \in E$, we have

$$\operatorname{re} \sum_{i=1}^{n} c_i \langle x_i^*, f(s) \rangle < \operatorname{re} \sum_{i=1}^{n} c_i z_i \,.$$

Since μ is a probability measure, integration of the inequality over E gives

$$\operatorname{re} \sum_{i=1}^{n} c_i \underbrace{\int_E \langle x_i^*, f \rangle \, d\mu}_{z_i} < \operatorname{re} \sum_{i=1}^{n} c_i z_i$$

which is an obvious contradiction. The proof of the theorem is complete. □

Theorem 12.3. *Under the assumptions of Theorem 12.1*

$$\Big\| \int_E f \, d\mu \Big\| \leq \int_E \|f\| \, d\mu \,.$$

Proof. Let $y = \int_E f \, d\mu$. According to the Hahn-Banach theorem there is $x^* \in X^*$ such that $\|x^*\| = 1$ and $\langle x^*, y \rangle = \|y\|$. Hence

$$\Big\| \int_E f \, d\mu \Big\| = \langle x^*, y \rangle = \int_E \langle x^*, f \rangle \, d\mu \leq \int_E \|f\| \, d\mu \,.$$

The proof is complete. □

12.1. **Banach space valued holomorphic functions.**

Definition. Let X be a complex banach space and $\Omega \subset \mathbb{C}$ an open set. We say that $f : \Omega \to X$ is *(strongly) holomorphic* if the limit

$$f'(z) = \lim_{w \to z} \frac{f(w) - f(z)}{w - z}$$

exists in the topology of X for every $z \in \Omega$.

Holomorphic functions are continuous and hence we can integrate them along curves in Ω. If γ is a positively oriented Jordan curve in Ω such that the interior of the curve Δ_γ is contained in Ω, then for every $z \in \Delta_\gamma$ we have the Cauchy formula

$$f(z) = \frac{1}{2\pi i} \int_\gamma \frac{f(\xi)}{\xi - z} \, d\xi \,. \tag{12.5}$$

Indeed, if $x^* \in X^*$, then $\Omega \ni z \mapsto \langle x^*, f(z)\rangle \in \mathbb{C}$ is a holomorphic function and hence

$$\langle x^*, f(z)\rangle = \frac{1}{2\pi i}\int_\gamma \frac{\langle x^*, f(\xi)\rangle}{\xi - z}\, d\xi = \left\langle x^*, \frac{1}{2\pi i}\int_\gamma \frac{f(\xi)}{\xi - z}\, d\xi\right\rangle$$

by (12.2) and hence the Cauchy formula (12.5) follows.

By the same argument we prove that

$$\int_\gamma f(z)\, dz = 0\,. \tag{12.6}$$

This easily implies that a substantial part of the theory of homomorphic functions can be generalized to the vectorial case. In particular one can easily show that holomorphic functions in a domain $\Omega \subset \mathbb{C}$ can be represented as a power series in any disc contained in Ω and a similar result hold for Laurent expansions. That implies that holomorphic functions are infinitely differentiable. the following result is an easy exercise.

Proposition 12.4. *Let X be a complex Banach space, Y be a closed subspace and $\Omega \subset \mathbb{C}$ be an open connected set. If $f : \Omega \to X$ is holomorphic and on a certain open subset of Ω it takes values into Y, then it takes values into Y on all of Ω.*

13. Reflexive spaces

If X is a Banach space, then $X^{**} = (X^*)^*$ is the dual to the dual space of X, called *the second dual space.* Observe that every element $x \in X$ defines a bounded functional on X^* by the formula

$$x^* \mapsto \langle x^*, x\rangle\,.$$

Denote this functional by $\kappa(x) \in X^{**}$, i.e.

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

Observe that

$$\|\kappa(x)\| = \sup_{\substack{x^* \in X^* \\ \|x^*\| \le 1}} \langle x^*, x\rangle = \|x\|$$

by Corollary 11.5. Therefore the *canonical embedding*

$$\kappa : X \to X^{**}$$

is an isometrical isomorphism between X and a closed subspace of X^{**}.

Definition. We say that a Banach space X is reflexive if $\kappa(X) = X^{**}$.

Warning. James constructed a *nonreflexive* space X such that X is isometrically isomorphic to X^{**}.

Theorem 13.1. *Let μ be a σ-finite measure and $1 < p, q < \infty$, $p^{-1} + q^{-1} = 1$. Then the space $L^p(\mu)$ is reflexive.*

Proof.[35] The space $(L^p(\mu))^*$ is isometrically isomorphic to $L^q(\mu)$ by Theorem 2.13. If $f \in L^p(\mu)$, then $\kappa(f)$ defines a functional on $(L^p)^* = L^q$ by

$$\langle \kappa(f), g\rangle = \int_X gf\, d\mu \quad g \in L^q(\mu)\,.$$

Since every functional on $(L^p)^* = L^q$ is of that form[36] we obtain that $\kappa(L^p(\mu)) = (L^p(\mu))^{**}$ and hence $L^p(\mu)$ is reflexive. □

The canonical embedding $\kappa : \ell^1 \to (\ell^1)^{**} = (\ell^\infty)^*$ gives an isometric isomorphism between ℓ^1 and a closed subspace of $(\ell^\infty)^*$, but as we know κ is not surjective, so ℓ^1 is not reflexive.

Theorem 13.2. *Any Hilbert space is reflexive.*

Proof. According to the Riesz representation theorem (Theorem 5.5) for every $x^* \in H^*$ there is a unique element $T(x^*) \in H$ such that

$$\langle x^*, x\rangle = \boldsymbol{\langle} x, T(x^*)\boldsymbol{\rangle} \quad \text{for } x \in H\,,$$

where on the left hand side we have evaluation of the functional and inner product on the right hand side.[37] Moreover $\|x^*\|_{H^*} = \|T(x^*)\|_H$.

If $x \in H$, then $\kappa(x)$ is a functional on H^* defined by

$$\langle \kappa(x), x^*\rangle = \langle x^*, x\rangle = \boldsymbol{\langle} x, T(x^*)\boldsymbol{\rangle} \tag{13.1}$$

and we want to show that every functional on H^* is of that form. The mapping $T : H^* \to H$ is bounded one-to-one and onto, however, if H is a complex Hilbert space, T is not quite linear, because

$$T(x_1^* + x_2^*) = T(x_1^*) + T(x_2^*) \quad \text{but} \quad T(\alpha x^*) = \overline{\alpha} T(x^*)\,.$$

The space H^* is a Hilbert space with the inner product

$$\boldsymbol{\langle} x_1^*, x_2^*\boldsymbol{\rangle} := \boldsymbol{\langle} T(x_2^*), T(x_1^*)\boldsymbol{\rangle}\,.$$

Now if $x^{**} \in H^{**}$, then it follows from the Riesz representation theorem that there is $z^* \in H^*$ such that

$$\langle x^{**}, x^*\rangle = \boldsymbol{\langle} x^*, z^*\boldsymbol{\rangle} = \boldsymbol{\langle} T(z^*), T(x^*)\boldsymbol{\rangle}\,,$$

and hence

$$\langle x^{**}, x^*\rangle = \langle \kappa(T(z^*)), x^*\rangle$$

by (13.1). The proof is complete. □

[35] We can briefly summarize the proof as follows: $(L^p)^* = L^q$, $(L^p)^{**} = (L^q)^* = L^p$, so the second dual to L^p is L^p itself and hence L^p is reflexive.

[36] Again by Theorem 2.13.

[37] We use thick notation for the inner product to distinguish it from the notation for the functional.

Theorem 13.3. *A closed subspace of a reflexive space is reflexive.*

Proof. Let Y be a closed subspace of a reflexive space X. We shall prove that Y is reflexive.

Every functional $x^* \in X^*$ defines a functional on Y as a restriction to Y,

$$\sigma : X^* \to Y^*, \quad \langle \sigma(x^*), y\rangle = \langle x^*, y\rangle \, .$$

Clearly

$$\|\sigma(x^*)\| \leq \|x^*\|, \quad \text{so} \quad \sigma(x^*) \in Y^*$$

Fix $y_0^{**} \in Y^{**}$. We have to prove that there is $x_0 \in Y$ such that

$$\langle y_0^{**}, y^*\rangle = \langle y^*, x_0\rangle \quad \text{for } y^* \in Y^*. \tag{13.2}$$

Let $\tau(y_0^{**}) \in X^{**}$ be defined by[38]

$$\langle \tau(y_0^{**}), x^*\rangle = \langle y_0^{**}, \sigma(x^*)\rangle \, .$$

Since

$$|\langle \tau(y_0^{**}), x^*\rangle| \leq \|y_0^{**}\| \, \|x^*\|$$

we have $\tau(y_0^{**}) \in X^{**}$. The canonical embedding $\kappa : X \to X^{**}$ is an isomorphism[39] and hence it is invertible $\kappa^{-1} : X^{**} \to X$. Thus

$$x_0 = \kappa^{-1}(\tau(y_0^{**})) \in X \, ,$$

but we will prove that $x_0 \in Y$. Note that

$$\langle \tau(y_0^{**}), x^*\rangle = \langle \kappa(x_0), x^*\rangle = \langle x^*, x_0\rangle \, .$$

Suppose $x_0 \notin Y$. According to Theorem 11.6 there is $x^* \in X^*$ such that

$$\langle x^*, x_0\rangle \neq 0, \quad \langle x^*, y\rangle = 0 \text{ for all } y \in Y$$

and hence $\sigma(x^*) = 0$, so

$$0 = \langle y_0^{**}, \sigma(x^*)\rangle = \langle \tau(y_0^{**}), x^*\rangle = \langle x^*, x_0\rangle \neq 0$$

which is a contradiction. We proved that

$$x_0 = \kappa^{-1}(\tau(y_0^{**})) \in Y \, .$$

For $y^* \in Y^*$ let $x^* \in X^*$ be an extension of y^*, so $y^* = \sigma(x^*)$. We have

$$\langle y_0^{**}, y^*\rangle = \langle y_0^{**}, \sigma(x^*)\rangle = \langle \tau(y_0^{**}), x^*\rangle = \langle x^*, x_0\rangle \overset{x_0 \in Y}{=} \langle y^*, x_0\rangle$$

which proves (13.2). $\square$

Theorem 13.4. *A Banach space isomorphic to a reflexive space is reflexive.*

---

[38] Roughly speaking $Y \subset X$, $X^* \subset Y^*$, $Y^{**} \subset X^{**}$, so every element in Y^{**} can be regarded as an element in X^{**}.

[39] Because X is reflexive.

Proof. Let $T : X \to Y$ be an isomorphism of Banach spaces and let X be reflexive. Let $y_0^{**} \in Y^{**}$. We have to prove that there is $y_0 \in Y$ such that

$$\langle y_0^{**}, y^* \rangle = \langle y^*, y_0 \rangle \quad \text{for } y^* \in Y^* . \tag{13.3}$$

The mapping $T^* : Y^* \to X^*$ defined by $\langle T^* y^*, x\rangle = \langle y^*, Tx\rangle$ is an isomorphism and hence it is invertible $(T^*)^{-1} : X^* \to Y^*$. Let $x_0^{**} \in X^{**}$ be defined by

$$\langle x_0^{**}, x^* \rangle = \langle y_0^{**}, (T^*)^{-1} x^* \rangle .$$

Since X is reflexive, there is $x_0 \in X$ such that

$$\langle x_0^{**}, x^* \rangle = \langle x^*, x_0 \rangle \quad \text{for } x^* \in X^*$$

and hence

$$\begin{aligned} \langle y_0^{**}, y^* \rangle &= \langle y_0^{**}, (T^*)^{-1}(T^* y^*) \rangle = \langle x_0^{**}, T^* y^* \rangle \\ &= \langle T^* y^*, x_0 \rangle = \langle y^*, T x_0 \rangle . \end{aligned}$$

Thus (13.3) holds with $y_0 = T x_0 \in Y$. □

Corollary 13.5. *Finitely dimensional Banach spaces are reflexive.*

Theorem 13.6. *X is reflexive if and only if X^* is reflexive.*

Proof. Suppose that X is reflexive. In order to prove reflexivity of X^* we have to show that for every

$$x_0^{***} \in X^{***} = (X^*)^{**} = (X^{**})^*$$

there is $x_0^* \in X^*$ such that

$$\langle x_0^{***}, x^{**} \rangle = \langle x^{**}, x_0^* \rangle \quad \text{for } x^{**} \in X^{**}.$$

Since X is reflexive, $\kappa : X \to X^{**}$ is an isomorphism and hence if is invertible. Note that the composition

$$X \xrightarrow{\kappa} X^{**} \xrightarrow{x_0^{***}} \mathbb{K}$$

defines $x_0^* = x_0^{***} \circ \kappa \in X^*$ and hence

$$\begin{aligned} \langle x_0^{***}, x^{**} \rangle &= \langle x_0^{***}, \kappa(\kappa^{-1}(x^{**})) \rangle = \langle x_0^{***} \circ \kappa, \kappa^{-1}(x^{**}) \rangle \\ &= \langle x_0^*, \kappa^{-1}(x^{**}) \rangle = \langle \kappa(\kappa^{-1}(x^{**})), x_0^* \rangle \\ &= \langle x^{**}, x_0^* \rangle , \end{aligned}$$

where in the second to last inequality we employed the definition of κ according to which $\langle \kappa(x), x^* \rangle = \langle x^*, x \rangle$. This proves reflexivity of X^*.

Now if X^* is reflexive, then by what we already proved X^{**} is reflexive. The canonical embedding $\kappa : X \to X^{**}$ gives an isomorphism between X and a closed subspace of the reflexive space X^{**}. Hence reflexivity of X follows from Theorem 13.3 and Theorem 13.4. □

Theorem 13.7. *If X is reflexive and $M \subset X$ is a closed subspace, then the quotient space X/M is reflexive.*

Proof. The space X^* is reflexive by Theorem 13.6. The space $(X/M)^*$ is isomorphic to the annihilator $M^\perp$ by Theorem 4.7. Since $M^\perp$ is a closed subspace of X^* it is reflexive by Theorem 13.3 and thus Theorem 13.4 implies reflexivity of $(X/M)^*$. This and Theorem 13.6 yields reflexivity of X/M. □

Corollary 13.8. *If $T : X \to Y$ is a bounded linear surjection of a reflexive space X onto a Banach space Y, then Y is reflexive.*

Proof. The space Y is isomorphic to $X/\ker T$ and hence it is reflexive as a consequence of Theorem 13.7 and Theorem 13.4. □

14. Weak convergence

Definition. Let X be a normed space. We say that a sequence $\{x_n\}_{n=1}^\infty \subset X$ *converges weakly* to $x_0 \in X$ if for every functional $x^* \in X^*$

$$\langle x^*, x_n\rangle \to \langle x^*, x_0\rangle \quad \text{as } n \to \infty.$$

We denote weak convergence by $x_n \rightharpoonup x_0$.

Clearly if $x_n \to x_0$ in norm, then $x_n \rightharpoonup x_0$.

Theorem 14.1. *Let X be a normed space. If $x_n \rightharpoonup x_0$ in X, then*

$$\sup_n \|x_n\| < \infty$$

and

$$\|x_0\| \le \liminf_{n\to\infty} \|x_n\| . \tag{14.1}$$

Proof. The canonical embedding defines $\kappa(x_n) \in X^{**}$ by

$$\langle \kappa(x_n), x^*\rangle = \langle x^*, \kappa(x_n)\rangle .$$

Observe that X^* is a Banach space, even if X is only a normed space. Since

$$\langle \kappa(x_0), x^*\rangle = \langle x^*, x_0\rangle = \lim_{n\to\infty}\langle x^*, x_n\rangle = \lim_{n\to\infty}\langle \kappa(x_n), x^*\rangle ,$$

the sequence $\kappa(x_n)$ of functionals on X^* is pointwise convergent to $\kappa(x_0)$. Therefore the Banach-Steinhaus theorem (Corollary 9.5) yields

$$\sup_n \|x_n\| = \sup_n \|\kappa(x_n)\| < \infty$$

and

$$\|x_0\| = \|\kappa(x_0)\| \le \liminf_{n\to\infty} \|\kappa(x_n)\| = \liminf_{n\to\infty} \|x_n\| .$$

The proof is complete. □

Note that we can prove (14.1) as a direct consequence of the Hahn-Banach theorem. Indeed, let $x^* \in X^*$ be such that $\|x^*\| = 1$, $\langle x^*, x_0\rangle = \|x_0\|$. Then

$$\|x_0\| = \langle x^*, x_0\rangle = \lim_{n\to\infty}\langle x^*, x_n\rangle \le \liminf_{n\to\infty} \|x_n\| .$$

In the previous proof instead of the Hahn-Banach theorem we employed the Banach-Steinhaus theorem, but the Hahn-Banach theorem was still there – it was needed for the proof of the equality $\|x\| = \|\kappa(x)\|$.

Weak convergence is, in general, much weaker than the convergence in norm. For example if

$$x_n = (x_{ni}) = (0, 0, \ldots, 0, 1, 0, \ldots) \in \ell^p, \quad 1 < p < \infty,$$

i.e. $x_{ni} = \delta_{ni}$, then $\|x_n - x_m\|_p = 2^{1/p}$ and hence no subsequence of x_n can be convergent in the norm. However, x_n is weakly convergence to 0. Indeed, according to Theorem 2.12 for every $x^* \in (\ell^p)^*$ there is $s = (s_i) \in \ell^q$ such that

$$\langle x^*, x_n \rangle = \sum_{i=1}^{\infty} s_i x_{ni} = s_n$$

and clearly $\langle x^*, x_n \rangle = s_n \to 0$ as $n \to \infty$. Note also that the same sequence is not weakly convergent in ℓ^1. Indeed, if we take $s = (1, -1, 1, -1, 1 \ldots) \in \ell^\infty = (\ell^1)^*$, then $\langle x^*, x_n \rangle = \pm 1$ is not convergent. Actually in ℓ^1 weak convergence is equivalent with the convergence in norm.

Theorem 14.2 (Schur)**.** *In the space ℓ^1 a sequence is weakly convergent if and only if it is convergent in norm.*

Proof. Suppose that $x_n \rightharpoonup x_0$ weakly in ℓ^1. Then $y_n = x_n - x_0 \rightharpoonup 0$ and we have to prove that $\|y_n\|_1 \to 0$. By contradiction suppose that $\|y_n\|_1$ does not converge to 0. Then there is $\varepsilon > 0$ and a subsequence (still denoted by y_n) such that $\|y_n\|_1 \geq \varepsilon$. Denote $y_n = (y_{nk})$. Note that for every integer p

$$\sum_{k=1}^{p} |y_{nk}| \to 0 \quad \text{as } n \to \infty \tag{14.2}$$

because of weak convergence to 0. Hence for each p we can find n so large that that the sum at (14.2) is less than $\varepsilon/4$. On the other hand the sum of the series is greater or equal to ε, so we can find $q > p$ such that $\sum_{k=p+1}^{q} |y_{nk}| > 3\varepsilon/4$ and $\sum_{k=q}^{\infty} |y_{nk}| < \varepsilon/4$. Hence using induction we can find two sequences of integers $\{n_i\}$ and $\{p_i\}$ such that

$$\sum_{k=1}^{p_i} |y_{n_i k}| < \frac{\varepsilon}{4}\,, \qquad \sum_{k=p_i+1}^{p_{i+1}} |y_{n_i k}| > \frac{3\varepsilon}{4}\,, \qquad \sum_{k=p_{i+1}}^{\infty} |y_{n_i k}| < \frac{\varepsilon}{4}\,.$$

Now we need to define a functional on which we will evaluate the sequence. For $k \leq p_1$ we set $s_k = 0$ and for $p_i < k \leq p_{i+1}$, $i = 1, 2, \ldots$ we define

$$s_k = \operatorname{sgn}(y_{n_i k}) = \begin{cases} \frac{\overline{y_{n_i k}}}{|y_{n_i k}|} & \text{if } y_{n_i k} \neq 0, \\ 0 & \text{if } y_{n_i k} = 0. \end{cases}$$

Then $x^* = (s_k) \in \ell^\infty = (\ell^1)^*$ is a bounded functional. We have

$$\langle x^*, y_{n_i} \rangle \geq \sum_{k=p_i+1}^{p_{i+1}} |y_{n_i k}| - \frac{\varepsilon}{4} - \frac{\varepsilon}{4} > \frac{\varepsilon}{4}$$

so $\langle x^*, y_{n_i} \rangle$ does not converge to 0. Contradiction. □

Weak convergence in the space of continuous functions is descried in the following result.

Theorem 14.3. *Let X be a compact metric space. Then a sequence $f_n \in C(X)$ converges weakly in $C(X)$ to $f \in C(X)$ if and only if there is $M > 0$ such that*

(a)

$$|f_n(x)| \leq M \quad \textit{for all } x \in X \textit{ and } n = 1, 2, 3 \ldots$$

and the sequence converges pointwise

(b)

$$f_n(x) \to f(x) \quad \textit{for all } x \in X.$$

Proof. Suppose $f_n \rightharpoonup f$. Then condition (a) follows from Theorem 14.1. Since for every $x \in X$ the value at the point x, i.e. $g \mapsto g(x)$ is a bounded functional on $C(X)$, the pointwise convergence (b) follows from the definition of weak convergence.

Now suppose that both conditions (a) and (b) are satisfied. According to the Riesz representation theorem for every functional $\Phi \in C(X)^*$ there is a signed Borel measure μ of finite total variation such that

$$\langle \Phi, g \rangle = \int_X g \, d\mu$$

and hence $\langle \Phi, f_n \rangle \to \langle \Phi, f \rangle$ by the dominated convergence theorem. □

While weak convergence is weaker than the convergence in norm, the following result shows that a sequence of convex combinations of a weakly convergent sequence converges in norm.

Theorem 14.4 (Mazur's lemma)**.** *Let X be a normed space and let $x_n \rightharpoonup x_0$ weakly in X. Then $u_n \to x_0$ in norm for some sequence u_n of the form*

$$u_n = \sum_{k=n}^{N(n)} a_k^n x_k$$

where $a_k^n \geq 0$, $\sum_{k=n}^{N(n)} a_k^n = 1$.

Proof. Let W_n be the closure of the set of all convex combinations of elements of the sequence $\{x_n, x_{n+1}, x_{n+2} \ldots\}$. It remains to prove that $x_0 \in W_n$, because it will imply existence of coefficients $a_k^n \geq 0$, $\sum_{k=n}^{N(n)} a_k^n = 1$ such that

$$\Big\| \sum_{k=n}^{N(n)} a_k^n x_k - x_0 \Big\| < \frac{1}{n}\,.$$

Suppose $x_0 \notin W_n$. The set W_n is convex and closed and the set $W' = \{x_0\}$ is convex and compact. According to Theorem 11.19 there is $x^* \in X^*$, $c \in \mathbb{R}$ and $\varepsilon > 0$ such that

$$\mathrm{re}\,\langle x^*, x_0\rangle \leq c - \varepsilon < c \leq \mathrm{re}\,\langle x^*, x\rangle$$

for all $x \in W_n$. In particular

$$\mathrm{re}\,\langle x^*, x_0\rangle \leq c - \varepsilon < c \leq \mathrm{re}\,\langle x^*, x_i\rangle \quad \text{for } i = n, n+1, \ldots$$

which contradicts weak convergence $x_i \rightharpoonup x_0$. □

Definition. Let X be a normed space. A sequence of functionals $\{x_n^*\} \subset X^*$ *converges weakly-$*$* to $x_0^* \in X^*$ if for every $x \in X$

$$\langle x_n^*, x\rangle \to \langle x_0^*, x\rangle \quad \text{as } n \to \infty.$$

We denote weak-$*$ convergence by $x_n^* \overset{*}{\rightharpoonup} x_0^*$.

Theorem 14.5. *If $x_n^* \overset{*}{\rightharpoonup} x_0^*$, then*

$$\sup_n \|x_n^*\| < \infty \quad \textit{and} \quad \|x_0^*\| \leq \liminf_{n\to\infty} \|x_n^*\|\,.$$

Proof. It is an immediate consequence of Corollary 9.5. □

The following result is a special case of the Banach-Alaoglu theorem.

Theorem 14.6 (Banach-Alaoglu, separable case)**.** *Let X be a separable normed space. Then every bounded sequence in X^* has a weakly-$*$ convergent subsequence.*

Proof. Let $\{x_n^*\}_n \subset X^*$ be a bounded sequence and let $\{x_1, x_2, \ldots\} \subset X$ be a countable and dense set. For each i, $\langle x_n^*, x_i\rangle$ is a bounded sequence of scalars so it has a convergent subsequence. Using the diagonal method we find a subsequence $\{x_{n_k}^*\}_k$ such that for each $i = 1, 2, 3, \ldots$ the sequence $\langle x_{n_k}^*, x_i\rangle$ is convergent. Thus $\{x_{n_k}^*\}_k$ is a bounded sequence of functionals that converges on a dense subset of X and it easily follows[40] that $\langle x_{n_k}^*, x\rangle$ converges for every $x \in X$. □

[40] Compare with an argument that follows (9.2).

Corollary 14.7. *Let X be a locally compact metric space and let $\{\mu_n\}_n$ be a sequence of finite positive Borel measures on X such that*

$$\sup_n \mu_n(X) < \infty \, .$$

Then there is a subsequence μ_{n_k} and a finite Borel measure μ on X such that

$$\mu(X) \leq \liminf_{k\to\infty} \mu_{n_k}(X)$$

and for every $f \in C_0(X)$

$$\int_X f \, d\mu_{n_k} \to \int_X f \, d\mu \, .$$

Proof. The measures μ_n define bounded functionals in $C_0(X)^*$

$$\langle \Phi_n, f\rangle = \int_X f \, d\mu$$

with

$$\sup_n \|\Phi_n\| = \sup_n \mu_n(X) < \infty \, ,$$

see Theorem 2.15. Since the space $C(X)$ is separable[41], there is a weakly-$*$ convergent subsequence $\Phi_{n_k} \xrightharpoonup{*} \Phi$. Clearly μ is represented by a signed Borel measure of finite total variation and

$$\int_X f \, d\mu_{n_k} \to \int_X f \, d\mu \quad \text{for all } f \in C_0(X). \tag{14.3}$$

Since the measures μ_n are positive it easily follows from (14.3) that μ is positive and Theorem 14.5 yields

$$\mu(X) = \|\Phi\| \leq \liminf_{k\to\infty} \|\Phi_{n_k}\| = \liminf_{k\to\infty} \mu_{n_k}(X) \, .$$

The proof is complete. □

In the situation described in the above result we say that the measures μ_{n_k} converge weakly-$*$ to the measure μ, $\mu_{n_k} \xrightharpoonup{*} \mu$.[42]

If X is a reflexive space, then X is isometric to the dual space to X^*, so the weak convergence is equivalent with the weak-$*$ convergence. Note also that for reflexive spaces, X is separable if and only if X^* is separable (Theorem 11.7) and hence Theorem 14.6 shows that a bounded sequence in a separable reflexive space has a weakly convergent subsequence. However, this result is also true without assuming separability.

Theorem 14.8. *A bounded sequence in a reflexive space has a weakly convergent subsequence.*

---

[41] Why?

[42] Sometimes in the literature this convergence is called weak convergence of measures $\mu_{n_k} \rightharpoonup \mu$, but we prefer to call it weak-$*$ convergence to be consistent with the language of functional analysis.

Proof. Let $\{x_n\}_n \subset X$ be a bounded sequence in a reflexive space X. Let

$$X_0 = \overline{\text{span}\,\{x_1, x_2, \ldots\}}\,.$$

Then X_0 is a closed separable subspace of X, so X_0 is reflexive. Since the dual space to X_0^* is separable as isometric to X_0, we conclude that X_0^* is separable (Theorem 11.7). Let $\{x_i^*\}_i \subset X_0^*$ be a countable and dense subset. By the diagonal argument there is a subsequence $\{x_{n_k}\}_k$ such that for every $i = 1, 2, 3, \ldots$

$$\langle x_i^*, x_{n_k}\rangle \quad \text{is convergent as } k \to \infty.$$

The density of $\{x_i^*\}_i$ in X_0^* easily implies that[43]

$$\langle x^*, x_{n_k}\rangle \quad \text{is convergent for every } x^* \in X_0^*.$$

Observe that every functional x^* defines a functional in X_0^* as a restriction to X_0, so

$$\langle x^*, x_{n_k}\rangle \quad \text{is convergent for every } x^* \in X^*.$$

If $\kappa : X \to X^{**}$ is the canonical embedding, then the limit

$$\lim_{k\to\infty} \langle \kappa(x_{n_k}), x^*\rangle = \lim_{k\to\infty} \langle x^*, x_{n_k}\rangle := \langle x^{**}, x^*\rangle$$

defines an element $x^{**} \in X^{**}$. Indeed, linearity of x^{**} is obvious and boundedness follows from the fact that

$$\sup_k \|\kappa(x_{n_k})\| = \sup_k \|x_{n_k}\| < \infty\,.$$

Now it is clear that $x_{n_k} \rightharpoonup x = \kappa^{-1}(x^{**})$. Indeed, for $x^* \in X^*$

$$\langle x^*, x_{n_k}\rangle \to \langle x^{**}, x^*\rangle = \langle x^*, x\rangle\,.$$

The proof is complete. □

Corollary 14.9. *Let mu be a σ-finite measure and $1 < p < \infty$. Then every bounded sequence $\{f_n\}_n \subset L^p(\mu)$ has a weakly convergence subsequence, i.e. there is a subsequence $\{f_{n_k}\}_k$ and $f \in L^p(\mu)$ such that for every $g \in L^q(\mu)$, $p^{-1} + q^{-1} = 1$*

$$\int_X f_{n_k} g\, d\mu \to \int_X fg\, d\mu \quad \textit{as } k \to \infty.$$

The following result generalizes Theorem 5.1 to any reflexive space, see also an example that follows Theorem 5.1.

Theorem 14.10. *Every nonempty, convex and closed set E in a reflexive space X contains an element of smallest norm.*

Proof. Let $\{x_n\} \subset E$ be a sequence such that $\|x_n\| \to \inf_{x\in E} \|x\|$. Hence the sequence $\{x_n\}$ is bounded, so it has a weakly convergent subsequence

[43] Verify the Cauchy condition.

$x_n \rightharpoonup x$. By Mazur's lemma a sequence of convex combinations of x_{n_k} converge to x in norm. Since convex combinations belong to E we conclude that $x \in E$. Now Theorem 14.1 yields

$$\|x\| \leq \liminf_{k\to\infty} \|x_{n_k}\| = \inf_{x\in E} \|x\|$$

and the claim follows. □

In the case of reflexive spaces the Riesz lemma (Theorem 3.6) has the following stronger version.[44]

Theorem 14.11. *Let $X_0 \neq X$ be a closed linear subspace of a reflexive space X. Then there is $y \in X$ such that*

$$\|y\| = 1 \quad \textit{and} \quad \|y - x\| \geq 1 \textit{ for all } x \in X_0.$$

Proof. Fix $y_0 \in X \setminus X_0$ and let $M = X_0 - y_0$. Since M is convex and closed, Theorem 14.10 yields the existence of $z_0 \in M$ such that

$$\|z_0\| = \inf_{z\in M} \|z\|\,.$$

Hence $z_0 = x_0 - y_0$, $x_0 \in X_0$ satisfies

$$\|x_0 - y_0\| = \inf_{x\in X_0} \|x - y_0\|$$

and the proof of the Riesz lemma shows that the vector $y = (y_0 - x_0)/\|y_0 - x_0\|$ satisfies the claim. □

14.1. **Direct methods in the calculus of variations.** In this section we will show an application of weak convergence to an abstract approach to existence of minimizers of variational problems.

Let $I : X \to \mathbb{R}$ be a function defined on a normed space X. Henceforth I will be called *functional* even if it is not linear. Actually in all interesting cases it will not be linear.

The problem is to find reasonable conditions that will guarantee existence of $\tilde{x} \in X$ such that

$$I(\tilde{x}) = \inf_{x\in X} I(x) \tag{14.4}$$

An element $\tilde{x} \in X$ satisfying (14.4) is called *minimizer* of I and a problem of finding a minimizer is called a *variational problem.*

Definition. We say that a functional I is *sequentially weakly lower semicontinuous* (swlsc) if for every sequence $x_n \rightharpoonup x$ weakly convergent in X,

$$I(x) \leq \liminf_{n\to\infty} I(x_n)\,.$$

[44] Compare with Corollary 3.8 and an example that follows it.

We say that I is *coercive* if

$$\|x_n\| \implies I(x_n) \to \infty\,.$$

Theorem 14.12. *If X is a reflexive Banach space and $I : X \to \mathbb{R}$ is swlsc and coercive, then there is $\tilde{x} \in X$ such that*

$$I(\tilde{x}) = \inf_{x\in X} I(x)\,.$$

Proof. Let $x_n \in X$ be a sequence such that

$$I(x_n) \to \inf_{x\in X} I(x)\,.$$

Such a sequence is called a *minimizing sequence*. Coercivity of I implies that x_n is bounded in X. Since the space is reflexive, x_n has a weakly convergent subsequence

$$x_{n_k} \rightharpoonup \tilde{x} \quad \text{in } X$$

and the sequential weak lower semicontinuity yields

$$I(\tilde{x}) \le \liminf_{k\to\infty} I(x_{n_k}) = \inf_{x\in X} I(x)\,.$$

The proof is complete. $\square$

In general the swlsc condition is very difficult to check, because it does not follow from continuity of I.

Definition. A functional $I : X \to \mathbb{R}$ is called *convex* if

$$I(tx + (1-t)y) \le tI(x) + (1-t)I(y) \quad \text{for } x, y \in X,\ t \in [0,1]$$

and *strictly convex* if

$$I(tx + (1-t)y) < tI(x) + (1-t)I(y) \quad \text{for } x, y \in X,\ x \ne y,\ t \in (0,1).$$

We say that I is *lower semicontinuous* if

$$x_n \to x \implies \liminf_{n\to\infty} I(x_n)\,.$$

The lower semicontinuity is much weaker than swlsc, because we require convergence of x_n to x in norm. In particular if I is continuous, then it is lower semicontinuous.

Theorem 14.13. *If X is a normed space and $I : X \to \mathbb{R}$ is convex and lower semicontinuous, then I is swlsc.*

Proof. We have to prove that

$$x_n \rightharpoonup x \implies I(x) \le \liminf_{n\to\infty} I(x_n)\,.$$

We can assume that $I(x_n)$ has a limit.[45] Denote the limit by g,

$$\lim_{n\to\infty} I(x_n) = g\,.$$

[45] Otherwise we choose a subsequence x_{n_k} such that $I(x_{n_k}) \to \liminf_{n\to\infty} I(x_n)$.

It follows from Mazur's lemma that for some sequence

$$u_n = \sum_{k=n}^{N(n)} a_k^n x_k, \quad a_k^n \geq 0, \quad \sum_{k=n}^{N(n)} a_k^n = 1$$

of convex combinations of x_k, $u_n \to x$ in norm. Now lower semicontinuity and coercivity of I yields

$$\begin{aligned} I(x) &\leq \liminf_{n\to\infty} I(u_n) = \liminf_{n\to\infty} I\Big(\sum_{k=n}^{N(n)} a_k^n x_k\Big) \\ &\leq \liminf_{n\to\infty} \sum_{k=n}^{N(n)} a_k^n I(x_k) = g\,. \end{aligned}$$

The proof is complete. □

As a corollary we obtain a result of fundamental importance in the convex calculus of variations.

Theorem 14.14 (Mazur-Schauder)**.** *If $I : X \to \mathbb{R}$ is a convex, lower semicontinuous and coercive functional defined on a reflexive Banach space X, then I attains minimum in X, i.e. there is $\tilde{x} \in X$ such that*

$$I(\tilde{x}) = \inf_{x\in X} I(x)\,.$$

If in addition I is strictly convex, then the minimizer $\tilde{x}$ is unique.

15. Weak topology

The following result which is interesting on its own will be useful later.

Proposition 15.1. *Let s be be the space of* all *(real or complex) sequences $x = (x_k)_{k=1}^\infty$. The space s with the metric*

$$d(x,y) = \sum_{k=1}^\infty 2^{-k} \frac{|x_k - y_k|}{1 + |x_k - y_k|}$$

is a complete metric space. Moreover for each sequence of positive numbers, $r_1, r_2, \ldots > 0$ the set

$$K = \{x \in s : |x_k| \leq r_k \textit{ for } k = 1, 2, \ldots\}$$

is compact in (s,d).

Proof. The fact that d is a metric directly follows from an elementary inequality

$$\frac{a+b}{1+a+b} \leq \frac{1}{1+a} + \frac{b}{1+b} \quad a, b \geq 0.$$

Observe that d is the metric of convergence on each coordinate, i.e.

$$x_n = (x_k^n)_{k=1}^\infty \to x = (x_k)_{k=1}^\infty$$

if and only if $x_k^n \to x_k$ for each $k = 1, 2, 3 \ldots$ If $x_n = (x_k^n)_{k=1}^\infty$ is a Cauchy sequence, them for each k, x_k^n is a Cauchy sequence of numbers, so it is convergent. That means x_n converges on each of its coordinates and thus it converges in the metric d. this proves that (s, d) is complete.

If $x_k \in K$, for $n = 1, 2, 3 \ldots$, then for each n, the sequence x_k^n is bounded, so it has a convergent subsequence. Using the diagonal argument we can find a subsequence x_{n_i} that converges on each of its coordinates, so it converges in the metric d. Thus K is compact. □

As an application we have

Theorem 15.2. *If X is a separable Banach space, then there is a metric ρ on X^* such that for every bounded set $E \subset X^*$ and $\{x_n^*\} \subset E$, $x^* \in E$,*

$$x_n^* \stackrel{*}{\rightharpoonup} x^* \quad \text{if and only if} \quad \rho(x_n^*, x^*) \to 0.$$

That means for each bounded set $E \subset X^*$, the metric ρ_E being the restriction of ρ to E is such that the convergence in ρ_E is equivalent to the weak-$*$ convergence in E, i.e. the weak-$*$ convergence in *bounded* subsets of X^* is metrizable. This is true in bounded sets only as the weak-$*$ convergence in X^*, $\dim X = \infty$ is *not* metrizable (see Theorem 15.4).

The following is a version of the separable Banach-Alaoglu theorem (Theorem 14.6).

Corollary 15.3. *If X is a separable Banach space, then every closed ball*

$$\overline{B} = \{x^* \in X^* : \|x^*\| \le r\}$$

is compact in the metric $\rho_{\overline{B}}$.

Proof. If $x_n^* \in \overline{B}$, then we can find a weakly-$*$ convergent subsequence $x_{n_i}^* \stackrel{*}{\rightharpoonup} x^*$ by Theorem 14.6. It follows then from Theorem 14.5 that $x^* \in \overline{B}$ and hence $\rho_{\overline{B}}(x_{n_i}^*, x^*) \to 0$. □

Proof of Theorem 15.2. Let $\{x_1, x_2, \ldots\} \subset B(0, 1) \subset X$ be a dense subset of the unit ball in X. Consider the mapping

$$\Phi : X^* \to s, \quad \Phi(x^*) = (\langle x^*, x_1\rangle, \langle x^*, x_2\rangle, \ldots).$$

The mapping is one-to-one. Indeed, if $\langle x^*, x_i\rangle = \langle y^*, x_i\rangle$ for all i, then $\langle x^* - y^*, x\rangle = 0$ for all $x \in X$, $\|x\| \le 1$ by the density argument and hence $x^* = y^*$. Thus

$$\rho(x^*, y^*) = d(\Phi(x^*), \Phi(y^*))$$

is a metric in X^*. It remains to prove that for every bounded set $E \subset X^*$, $\{x_n^*\} \subset E$ and $x^* \in E$,

$$x_n^* \stackrel{*}{\rightharpoonup} x^* \quad \text{if and only if} \quad d(\Phi(x_n^*), \Phi(x^*)) \to 0.$$

If $x_n^* \xrightarrow{*} x^*$, then $\langle x_n^*, x_i\rangle \to \langle x^*, x_i\rangle$ for every i, so every coordinate of $\Phi(x_n^*)$ converges to the corresponding coordinate of $\Phi(x^*)$ and hence $d(\Phi(x_n^*), \Phi(x^*)) \to 0$.

Conversely, if $d(\Phi(x_n^*), \Phi(x^*)) \to 0$, then coordinates of $\Phi(x_n^*)$ converge to coordinates of $\Phi(x^*)$, i.e.

$$\langle x_n^*, x_i\rangle \to \langle x^*, x_i\rangle \tag{15.1}$$

and since the sequence x_n^* is bounded

$$\langle x_n^*, x\rangle \to \langle x^*, x\rangle \tag{15.2}$$

for every x with $\|x\| \leq 1$ and hence for every $x \in X$, i.e. $x_n^* \xrightarrow{*} x^*$. □

Note that the metric ρ is defined by an explicit formula

$$\rho(x^*, y^*) = \sum_{k=1}^{\infty} 2^{-k} \frac{|\langle x^* - y^*, x_k\rangle|}{1 + |\langle x^* - y^*, x_k\rangle|}\,,$$

where $\{x_1, x_2, \ldots\} \subset B(0,1)$ is a dense subset.

The assumption that E is bounded was employed only once in the proof of the implication from (15.1) to (15.2) and the boundedness was a crucial assumption here as the following result shows.

Theorem 15.4. *Let X be a separable Banach space, $\dim X = \infty$. Then there is no metric d in X^* such that $x_n^* \xrightarrow{*} x^*$ if and only if $d(x_n^*, x^*) \to 0$.*

Proof. Let $\{x_1, x_2, \ldots\} \subset X$ be a dense subset and let

$$X_n = \operatorname{span}\{x_1, x_2, \ldots, x_n\}$$

By the Hahn-Banach theorem there is a functional $x_n^* \in X^*$ such that $\|x_n^*\| = 1$ and $\langle x_n^*, x\rangle = 0$ for $x \in X_n$. Note that $x_n^* \xrightarrow{*} 0$. Indeed, if $x \in X$, then for every $\varepsilon > 0$ there is m such that

$$\|x - x_m\| < \varepsilon\,.$$

Hence for $n > m$ we have

$$|\langle x_n^*, x\rangle| = |\langle x_n^*, x - x_m\rangle| < \varepsilon\,.$$

Similarly for every $k > 0$, $kx_n^* \xrightarrow{*} 0$ as $n \to \infty$. Suppose now that there is a metric d as in the statement on the theorem. Then $d(kx_n^*, 0) \to 0$ as $n \to \infty$. Hence for each k we can find $n(k)$ such that $d(kx_{n(k)}^*, 0) < 1/k$, so $kx_{n(k)}^* \xrightarrow{*} 0$ as $k \to \infty$. However, $\|kx_{n(k)}^*\| = k$ and we arrive to a contradiction with the fact that the weakly-$*$ convergent sequence is bounded, Theorem 14.5. □

15.1. **Topological spaces.** In addition to the norm topology Banach spaces are equipped with so called weak topology and dual spaces with weak-$*$ topology. These topologies do not come from metric, so we have to introduce topological spaces.

Definition. A *topological space* is a set X with a family $\mathcal{T}$ of subsets of X called *open sets* that satisfy the following properties

(a) $\emptyset, X \in \mathcal{T}$;
(b) $\mathcal{T}$ is closed under finite intersections, i.e. if $U, V \in T$, then $U \cap V \in \mathcal{T}$;
(c) $\mathcal{T}$ is closed under arbitrary unions, i.e. if $\{U_i\}_{i\in I} \subset \mathcal{T}$, then $\bigcup_{i\in I} U_i \in \mathcal{T}$.

The family $\mathcal{T}$ is called *topology* in X.

Let $X_0 \subset X$ be a subset of a topological space. Then
$$\mathcal{T}_0 = \{U \cap X_0 : U \in \mathcal{T}\}$$
is the *induced topology* in X_0. Obviously $(X_0, \mathcal{T}_0)$ is a topological space.

We say that $E \subset X$ is *closed* is $X \setminus E$ is open, i.e. $X \setminus E \in \mathcal{T}$.

If $A \subset X$, the *closure* of A denoted by $\overline{A}$ is the intersection of all closed sets that contain A, i.e. it is the smallest closed set that contain A. Clearly, if $x \in \overline{A}$ then for every open set U such that $x \in U$ we have $U \cap A \neq \emptyset$.

A topological space $(X, \mathcal{T})$ is called *Hausdorff* if for every $x, y \in X$, $x \neq y$ there are opens sets $U, V \in \mathcal{T}$ such that $x \in U$, $y \in V$, $U \cap V = \emptyset$.

Example. If (X, d) is a metric space, then the family of all open sets is a Hausdorff topology in X.

Definition. Let $(X, \mathcal{T})$, $(Y, \mathcal{F})$ be two topological spaces. We say that a mapping $f : X \to Y$ is *continuous* if preimages of open sets are open, i.e. $f^{-1}(U) \in \mathcal{T}$ whenever $U \in \mathcal{F}$.

We say that a topology $\mathcal{T}_1$ is *weaker* than a topology $\mathcal{T}_2$ if $\mathcal{T}_1 \subset \mathcal{T}_2$. In this situation $\mathcal{T}_2$ is called a stronger topology. Clearly if a mapping is continuous with respect to $\mathcal{T}_1$, then it is also continuous with respect to the stronger topology $\mathcal{T}_2$.

A family $\mathcal{B} \subset \mathcal{T}$ is called a *base* if every $U \in \mathcal{T}$ is a union of elements of $\mathcal{B}$.

Example. If (X, d) is a metric space, then the family of all open balls is a base for the topology generated by the metric.

Definition. We say that a topological space is *compact* if every open covering has a finite subcovering. A subset of a topological space is compact if it is compact with respect to the induced topology.

Proposition 15.5.

(a) *Let E be a closed subset of a compact space. Then E is compact with respect to the induced topology.*
(b) *Let E be a compact subset of a Hausdorff topological space. Then E is closed.*[46]
(c) *A continuous image of a compact set is compact.*
(d) *Let $f : X \to Y$ be a continuous one-to-one mapping of a compact space into a Hausdorff space. Then $f^{-1} : f(X) \to X$ is continuous.*

Proof. We leave the proofs of (a)-(c) as an exercise and we will only prove property (d). We need to show that for every open set $U \subset X$, $(f^{-1})^{-1}(U) = f(U)$ is an open subset of $f(X)$. $X \setminus U$ is closed, so it is compact by (a). Hence $f(X \setminus U)$ is compact by (c). Since $f(X)$ is Hausdorff, $f(X \setminus U)$ is closed and hence $f(U) = f(X) \setminus f(X \setminus U)$ is open. $\square$

DEFINITION. Let $\mathcal{K}$ be a family of functions from a set Y into a topological space $(X, \mathcal{T})$. The $\mathcal{K}$-*weak* topology in Y is the weakest topology on Y for which all the functions in the family $\mathcal{K}$ are continuous.

The $\mathcal{K}$-weak topology is constructed as follows. Observe that the sets

$$\mathcal{B} = \{f_1^{-1}(U_1) \cap \ldots \cap f_n^{-1}(U_n) : f_i \in \mathcal{K},\ U_i \in \mathcal{T}, i = 1, 2, \ldots, n\}$$

must be open and that the family of all possible unions of the sets from $\mathcal{B}$ is a topology, so it is the $\mathcal{K}$-weak topology and $\mathcal{B}$ is a base.

Let $(X_1, \mathcal{T}_1), \ldots, (X_n, \mathcal{T}_n)$ be topological spaces. In the Cartesian product

$$\prod_{i=1}^n X_i = X_1 \times \ldots \times X_n$$

we consider open rectangles

$$U_1 \times \ldots \times U_n, \quad U_i \in \mathcal{T}_i,\ i = 1, 2, \ldots, n.$$

Then the family of all unions of open rectangles defines a topology in $\prod_{i=1}^n X_i$ which is called the *product topology* and denoted by $\prod_{i=1}^n \mathcal{T}_i = \mathcal{T}_1 \times \ldots \times \mathcal{T}_n$. Clearly open rectangles form a base for this topology.

Note that each projection

$$\pi_i : X_1 \times \ldots \times X_n \to X_i, \quad \pi_i(x_1, \ldots, x_n) = x_i$$

is continuous with respect to the product topology and that the product topology is the weakest one which makes all the projections continuous.

[46]The claim is not true if the space is not Hausdorff. Indeed, $\{\emptyset, X\}$ is a topology in X and every subset is compact, but only $\emptyset$ and X are closed.

If $(X_1,d_1),\ldots,(X_n,d_n)$ are metric spaces, then it is easy to see that the topology induced by the metric d in the product $\prod_{i=1}^n X_i$

$$d((x_1,\ldots,x_n),(y_1,\ldots,y_n))=\sum_{i=1}^n d_i(x_i,y_i)$$

coincides with the product topology.

One can also define a product topology in a product of an arbitrary (possibly uncountable) number of topological spaces.

Definition. Let $(X_i,\mathcal{T}_i)_{i\in I}$ be an arbitrary family of topological spaces. The Cartesian product

$$\prod_{i\in I} X_i=\{(x_i)_{i\in I}:\, x_i\in X_i\}$$

is equipped with the *product topology* which is the weakest topology for which each of the projection

$$\pi_j:\prod_{i\in I} X_i\to X_j,\quad \pi_j((x_i)_{i\in I})=x_j$$

is continuous.

Let $i_1,\ldots,i_n\in I$ and $U_{i_1}\in\mathcal{T}_{i_1},\ldots,U_{i_n}\in\mathcal{T}_{i_n}$ be chosen arbitrarily. It is easy to see that the sets

$$\{(x_i)_{i\in I}:\, x_{i_1}\in U_{i_1},\ldots,x_{i_n}\in U_{i_n}\}=\pi_{i_1}^{-1}(U_{i_1})\cap\ldots\cap\pi_{i_n}^{-1}(U_{i_n})$$

form a base for the product topology.

It is easy to see that if the spaces $\{X_i\}_{i\in I}$ are Hausdorff, then $\prod_{i\in I} X_i$ is Hausdorff.

Theorem 15.6 (Tychonov)**.** *Let $\{X_i\}_{i\in I}$ be an arbitrary collection of compact topological spaces. Then the product $\prod_{i\in I} X_i$ is compact.*

We will not prove it.

15.2. **Weak topology in Banach spaces.**

Definition. Let X be a Banach space. The *weak topology* in X is the weakest topology with respect to which all functionals $x^*\in X^*$ are continuous.

Let X be a normed space. The *weak-$*$ topology* is the weakest topology in X^* with respect to which all functions of the form $x^*\mapsto\langle x^*,x\rangle$ for $x\in X$ are continuous.

Exercise. *Prove that weak and weak-$*$ topologies are Hausdorff.*

Theorem 15.7 (Banach-Alaoglu)**.** *Let X be a normed space. The closed unit ball in X^*, i.e.*

$$\overline{B} = \{x^* \in X^* : \|x^*\| \leq 1\}$$

is compact in the weak-$*$ topology.

Proof. For each $x \in X$ let

$$B_x = \{\lambda \in \mathbb{C} : |\lambda| \leq \|x\|\} .$$

Each set B_x is compact, so is the product

$$K = \prod_{x \in X} B_x .$$

Elements of K can be identified with functions $f : X \to \mathbb{C}$ such that $|f(x)| \leq \|x\|$ for all $x \in X$. Observe that the unit ball $\overline{B}$ in X^* is a subset of K. Actually, $X^* \cap K = \overline{B}$. It remains to prove that the topology in $\overline{B}$ induced from K coincides with the weak-$*$ topology and that $\overline{B}$ is a closed subset of K.

The sets

$$\mathcal{U} = \bigcap_{i=1}^{n} \{f : X \to \mathbb{C} : |f(x)| \leq \|x\|, \ f(x_i) \in U_i\}$$

form a base of the topology in K and the sets

$$\mathcal{V} = \bigcap_{i=1}^{n} \{x^* \in X^* : \langle x^*, x_i\rangle \in U_i\}$$

form a base of the weak-$*$ topology in X^*. Since

$$\mathcal{U} \cap \overline{B} = \mathcal{V} \cap \overline{B}$$

it easily implies that the weak-$*$ topology and the topology induced from K coincide on $\overline{B}$.

It remains to prove that $\overline{B}$ is a closed subset of K. Let f_0 be in the closure of $\overline{B}$ in the topology of K. We have to prove that $f_0 \in \overline{B}$.

Fix $x, y \in X$, $\alpha, \beta \in \mathbb{C}$. The set

$$\begin{aligned} &V_\varepsilon = \{f \in K : \\ &|f(x) - f_0(x)| < \varepsilon, \, |f(y) - f_0(y)| < \varepsilon, \, |f(\alpha x + \beta y) - f_0(\alpha x + \beta y)| < \varepsilon\} \end{aligned}$$

is open and contains f_0. Hence $V_\varepsilon \cap \overline{B} \neq \emptyset$ by the definition of the closure. Thus there is $x^* \in V_\varepsilon \cap \overline{B}$, and hence it satisfies

$$\begin{gathered} |\langle x^*, x\rangle - f_0(x)| < \varepsilon, \quad |\langle x^*, y\rangle - f_0(y)|, \varepsilon \\ |\underbrace{\alpha\langle x^*, x\rangle + \beta\langle x^*, y\rangle}_{\langle x^*, \alpha x + \beta y\rangle} - f_0(\alpha x + \beta y)| < \varepsilon . \end{gathered}$$

The above inequalities yield

$$|f_0(\alpha x+\beta y)-\alpha f_0(x)-\beta f_0(y)|<(1+|\alpha|+|\beta|)\,\varepsilon\,.$$

Since $\varepsilon>0$ can be arbitrarily small we have

$$f_0(\alpha x+\beta y)=\alpha f_0(x)+\beta f_0(y)\,.$$

Moreover $|f_0(x)|\le\|x\|$ as an element of K which implies $f_0\in\overline{B}$. $\square$

The following result shows that in the case of separable Banach spaces, Corollary 15.3 is equivalent to the Banach-Alaoglu theorem.

Theorem 15.8. *Let X be a separable Banach space. Then the weak-$*$ topology in the unit ball $\overline{B}$ in X^* coincides with the topology induced by the metric $\rho_{\overline{B}}$.*

Proof. Let

$$\Phi:X^*\to s,\quad \Phi(x^*)=(\langle x^*,x_1\rangle,\langle x^*,x_2\rangle,\ldots)$$

be a mapping defined in the proof of Theorem 15.2. The mapping is one-to-one. It remains to prove that Φ is continuous. Indeed, since s is Hausdorff as a metric space and $\overline{B}$ is compact in the weak-$*$ topology, it will follow from Proposition 15.5(d) that

$$\Phi|_{\overline{B}}:\overline{B}\to s$$

is a homeomorphism onto the image and hence the weak-$*$ topology in $\overline{B}$ will coincide with the topology generated by the metric $\rho(x^*,y^*)=d(\Phi(x^*),\Phi(y^*))$.

Fix $r>0$ and let $w=(w_i)\in s$. Let N be such that $\sum_{k=N+1}^\infty 2^{-k}<r/2$. Then the set

$$A(w,r,N)=\{t\in s:\,|t_i-w_i|<r/2,\ i=1,2,\ldots,N\}$$

is open, contained in $B(w,r)$ and $w\in A(w,r,N)$. Hence the sets $A(w,r,N)$ form a base in s. Thus it remains to prove that $\Phi^{-1}(A(w,r,N))$ is open in the weak-$*$ topology. By the definition of the weak-$*$ topology the functions $x^*\mapsto\langle x^*,x_i\rangle$ are continuous, so the sets

$$\{x^*\in X^*:\,|\langle x^*,x_i\rangle-w_i|<r/2\}$$

are open as preimages of $B(w_i,r/2)\subset\mathbb{C}$ and hence

$$\Phi^{-1}(A(w,r,N))=\bigcap_{i=1}^N\{x^*\in X^*:\,|\langle x^*,x_i\rangle-w_i|<r/2\}$$

is open. $\square$

16. Compact operators

Definition. Let $A \in B(X,Y)$ be a bounded operator between Banach spaces. We say that A is *compact* if it maps bounded sets onto relatively compact sets, i.e. if for every bounded set $E \subset X$, $\overline{A(E)} \subset Y$ is compact.

The class of compact operators will be denoted by $K(X,Y)$ with $K(X) = K(X,X)$.

Equivalently $A \in B(X,Y)$ is compact if for every bounded sequence $\{x_n\}_n \subset X$, $\{Ax_n\}_n \subset Y$ has a convergent subsequence.

It is also easy to see that A is compact if and only if

$$\overline{A(B(0,1))} \subset Y \quad \text{is compact.}$$

For a bounded mapping $A \in B(X,Y)$ we define[47]

$$\mathcal{R}(A) = A(X), \qquad \mathcal{N}(A) = \ker A\,.$$

Proposition 16.1. *If* $\dim \mathcal{R}(A) < \infty$*, then* A *is compact.*

Since the closed unit ball in an infinitely dimensional space is not compact (Corollary 3.7) we have

Proposition 16.2. *If* $\dim X = \infty$*, then the identity mapping* $\mathrm{id} : X \to X$ *is not compact.*

Theorem 16.3. *If* $A_n \in B(X,Y)$ *is a sequence of compact operators between Banach spaces and* $A_n \to A$ *in norm, then* A *is compact.*

Proof. Let $\{x_n\}_n \subset X$ be a bounded sequence, say $\|x_n\| \leq M$ for all n. We need to show that $Ax_n \in Y$ has a convergent subsequence. Since each sequence $\{A_i x_n\}_n$ has a convergent subsequence, by diagonal argument we find a subsequence $\{x_{n_k}\}_k$ such that for each $i = 1, 2, 3 \ldots$

$$A_i x_{n_k} \quad \text{is convergent as } k \to \infty.$$

Given $\varepsilon > 0$, let i be such that

$$\|A - A_i\| < \varepsilon/3M\,.$$

The Cauchy condition gives existence of k_0 such that for $k, l \geq k_0$

$$\|A_i x_{n_k} - A_i x_{n_l}\| < \varepsilon/3$$

and hence

$$\begin{aligned} \|Ax_{n_k} - Ax_{n_l}\| &\leq \|A_i x_{n_k} - A_i x_{n_l}\| + \|Ax_{n_k} - A_i x_{n_k}\| + \|Ax_{n_k} - A_i x_{n_l}\| \\ &\leq \frac{\varepsilon}{3} + \|A - A_i\|M + \|A - A_i\|M < \varepsilon\,. \end{aligned}$$

Thus the sequence $\{Ax_{n_k}\}_k$ satisfies the Cauchy condition and hence it is convergent, because Y is a Banach space. $\square$

[47] $\mathcal{R}$ stands for range and $\mathcal{N}$ for null space.

Theorem 16.4. *Let $A \in B(X,Y)$ be a compact operator between Banach spaces. If $x_n \rightharpoonup x$ weakly in X, then $Ax_n \to Ax$ in norm.*

Proof. First we will prove that $Ax_n \rightharpoonup Ax$ weakly in Y. If $y^* \in Y^*$, then $x^* = y^* \circ A$, i.e. $\langle x^*, x\rangle = \langle y^*, Ax\rangle$ is a bounded functional in X^* and hence weak convergence $x_n \rightharpoonup x$ yields

$$\langle y^*, Ax_n\rangle = \langle x^*, x_n\rangle \to \langle x^*, x\rangle = \langle y^*, Ax\rangle\, .$$

Suppose that Ax_n does not converge to Ax in norm. Then there is a subsequence Ax_{n_k} and $\varepsilon > 0$ such that

$$\|Ax_{n_k} - Ax\| \geq \varepsilon\, . \tag{16.1}$$

The sequence x_n is bounded (as weakly convergent) and since A is compact, there is a convergent subsequence $Ax_{n_{k_l}} \to y$. Since $Ax_{n_{k_l}} \rightharpoonup Ax$ weakly we easily conclude that $y = Ax$, so $Ax_{n_{k_l}} \to Ax$ in norm which contradicts (16.1). □

Theorem 16.5. *Let X, Y, Z, V be Banach spaces. If $A \in B(X,Y)$ is compact and $B \in B(Y,Z)$, $C \in B(V,X)$ are bounded operators, then the operators $BA \in B(X,Z)$, $AC \in B(V,Y)$ are compact.*

Proof. Easy exercise. □

It follows from Proposition 16.1 and Theorem 16.3 that if an operator $A \in B(X,Y)$ can be approximated in norm by operators with finitely dimensional range, then A is compact. In the case of operators into Hilbert spaces this property characterizes compact operators.

Theorem 16.6. *Let $A \in B(X,H)$ be a bounded operator between a Banach space and a Hilbert space. Then A is compact if and only if there is a sequence of operators $A_n \in B(X,H)$ such that $\dim \mathcal{R}(A_n) < \infty$ and $A_n \to A$ in norm.*

Proof. We only need to prove the implication from left to right. We can assume that $\dim \mathcal{R}(A) = \infty$ as otherwise the claim is obvious. Since A is compact, $\mathcal{R}(A)$ is a union of countably many relatively compact sets, so $\overline{\mathcal{R}(A)}$ is separable. Let $\{\varphi_i\}_{i=1}^\infty$ be an orthonormal basis in $\overline{\mathcal{R}(A)}$ and let

$$P_n : \overline{\mathcal{R}(A)} \to \operatorname{span}\{\varphi_1, \varphi_2, \ldots, \varphi_n\}$$

be the orthogonal projection. It remains to prove that $A_n = P_nA \to A$ in norm. If not, there is $\varepsilon > 0$ and a subsequence (still denoted by A_n) such that

$$\|A_n - A\| \geq \varepsilon\, .$$

Hence there is a sequence $x_n \in X$, $\|x_n\| = 1$ such that

$$\|(A_n - A)x_n\| \geq \frac{\varepsilon}{2}\, . \tag{16.2}$$

Since A is compact, $A_{n_k}x_{n_k} \to y \in \overline{\mathcal{R}(A)}$ for some subsequence. We have

$$(A - A_{n_k})x_{n_k} = (I - P_{n_k})Ax_{n_k} = (I - P_{n_k})y + (I - P_{n_k})(A_{n_k} - y) \to 0$$

as $k \to \infty$ which contradicts (16.2). $\square$

Recall that for a bounded operator $A \in B(X, Y)$ we define the adjoint operator $A^* \in B(Y^*, X^*)$ by

$$\langle A^* y^*, x\rangle = \langle y^*, Ax\rangle \, .$$

Theorem 16.7. *Let $\Omega \subset \mathbb{R}^n$ be open and $K \in L^2(\Omega \times \Omega)$. Then the integral operator*

$$Kf = \int_\Omega K(x,y) f(y)\, dy \, , \quad x \in \Omega$$

defines a compact operator $K : L^2(\Omega) \to L^2(\Omega)$.

Proof. The operator $K : L^2(\Omega) \to L^2(\Omega)$ is well defined, because

$$\begin{aligned} \|Kf\|_2^2 &= \int_\Omega \Big| \int_\Omega K(x,y) f(y)\, dy \Big|^2 dx \\ &\leq \int_\Omega \Big(\int_\Omega |K(x,y)|^2 \, dy \int_\Omega |f(y)|^2 \, dy \Big) \, dx \\ &= \|K\|^2_{L^2(\Omega\times\Omega)} \, \|f\|_2^2 \, , \end{aligned}$$

i.e.

$$\|Kf\|_2 \leq \|K\|_2 \, \|f\|_2 \, . \tag{16.3}$$

Let $\{\varphi_i\}_{i=1}^\infty$ be an orthonormal basis in $L^2(\Omega)$. According to Theorem 5.15 the functions $\{\varphi_i(x)\varphi_j(y)\}_{i,j=1}^\infty$ form an orthonormal basis in $L^2(\Omega \times \Omega)$. We can write

$$K(x,y) = \sum_{i,j=1}^\infty a_{ij} \varphi_i(x) \varphi_j(y)$$

where the series converges to K in the norm of $L^2(\Omega \times \Omega)$, i.e. the functions

$$K_m(x,y) = \sum_{i,j=1}^m a_{ij} \varphi_i(x) \varphi_j(y)$$

converge to K in $L^2(\Omega \times \Omega)$. Now it follows from inequality (16.3) that the operators $K_m \in B(L^2(\Omega), L^2(\Omega))$ converge to $K \in B(L^2(\Omega), L^2(\Omega))$ in norm. Since the range of each operator K_m is finitely dimensional

$$K_m f(x) = \sum_{i,j=1}^m a_{ij} \varphi_i(x) \int_\Omega f(y) \varphi_j(y) \, dy \in \operatorname{span}\{\varphi_1, \ldots, \varphi_m\}$$

the claim follows from Theorem 16.6. $\square$

Theorem 16.8 (Schauder)**.** *An operator $A \in B(X, Y)$ between Banach spaces is compact if and only if $A^* \in B(Y^*, X^*)$ is compact.*

Proof. $\Rightarrow$. Let $A \in B(X,Y)$ be compact. In order to prove compactness of A^* we need to show that for every bounded sequence $\{y_n^*\}_n \subset Y^*$, say $\sup_n \|y_n^*\| \leq M$, $\{A^*y_n^*\}_n \subset X^*$ has a convergent subsequence.

Let $S = \{x \in H : \|x\| \leq 1\}$. The set $K = \overline{A(S)} \subset Y$ is compact. The family of functions

$$f_n : K \to \mathbb{K}, \quad f_n(y) = \langle y_n^*, y\rangle$$

is bounded and equicontinuous. Boundedness is easy and equicontinuity follows from the estimate

$$|f_n(y_1) - f_n(y_2)| \leq M\|y_1 - y_2\|\,.$$

According to the Arzela-Ascoli theorem, there is a uniformly convergent subsequence $\{f_{n_k}\}$. Hence also the functions

$$g_k : S \to \mathbb{K}, \quad g_k(x) = f_n(Ax) = \langle A^*y_{n_k}^*, x\rangle$$

converge uniformly on S.[48] The limit

$$\langle x^*, x\rangle := \lim_{k\to\infty} \langle A^*y_{n_k}^*, x\rangle$$

exists for every $x \in X$ and it defines a bounded functional $x^* \in X^*$. Moreover

$$\|A^*y_{n_k}^* - x^*\| = \sup_{x\in S} |g_k(x) - \langle x^*, x\rangle| \to 0 \quad \text{as } k \to \infty.$$

This completes the proof of compactness of A^*.

$\Leftarrow$. Suppose now that $A^* \in B(Y^*, X^*)$ is compact. Let $\{x_n\}_n \subset X$ be a bounded sequence. We need to prove that $\{Ax_n\}_n \subset Y$ has a convergent subsequence. By the first part of the proof $A^{**} \in B(X^{**}, Y^{**})$ is compact and hence $A^{**}(\kappa_X(x_n))$ has a convergent subsequence $A^{**}(\kappa_X(x_{n_k}))$, where $\kappa_X : X \to X^{**}$ is the canonical embedding. Since

$$\kappa_Y(Ax_{n_k}) = A^{**}(\kappa_X(x_{n_k}))$$

where $\kappa_Y : Y \to Y^{**}$ is the canonical embedding and hence an isometric embedding, the sequence $\{Ax_{n_k}\}_k$ satisfies the Cauchy condition, so it is convergent. □

16.1. **Fredholm operators.** Fredholm operators play an important role in the theory of integral and elliptic equations. We will describe later an application to the Fredholm theory of integral equations. Another application appears in the famous Atiyah-Singer index theorem which deals with the index of a special Fredholm operator defined as an elliptic operator on manifolds. The Atiyah-Singer theorem plays a fundamental role in algebraic topology and in contemporary physics.

[48]Although the family g_k is bounded and equicontinuous, we could not apply the Arzela-Ascoli theorem directly to g_k, because S is not compact.

DEFINITION. A bounded operator $A \in B(X,Y)$ between Banach spaces is called *Fredholm operator* if

$$\dim \mathcal{N}(A) < \infty, \quad \dim(Y/\mathcal{R}(A)) < \infty.$$

Note that according to Theorem 4.8 $\mathcal{R}(A)$ is a closed subspace on Y.

The space of Fredholm operators will be denoted by $\mathrm{Fred}\,(X,Y)$. The quantity

$$\operatorname{ind} A = \dim \mathcal{N}(A) - \dim(Y/\mathcal{R}(A))$$

is called the *Fredholm index* of A.

Theorem 16.9. $\mathrm{Fred}\,(X,Y)$ *is an open subset of* $B(X,Y)$*. The Fredholm index* $\operatorname{ind} : \mathrm{Fred}\,(X,Y) \to \mathbb{Z}$ *is continuous and hence constant on connected components of* $\mathrm{Fred}\,(X,Y)$*.*

Proof. Let $A \in \mathrm{Fred}\,(X,Y)$. By Theorem 11.8 the subspaces $\mathcal{N}(A) \subset X$ and $\mathcal{R}(A) \subset Y$ are complemented, so there are closed subspaces $X_1 \subset X$ and $Y_1 \subset Y$ such that

$$X = \mathcal{N}(A) \oplus X_1, \quad Y = \mathcal{R}(A) \oplus Y_1.$$

For any operator $L \in B(X,Y)$ we define

$$\tilde{L} : X_1 \oplus Y_1 \to Y, \quad \tilde{L}(x_1, y_1) = Lx_1 + y_1\,.$$

It is easy to see that

$$\tilde{A} : X_1 \oplus Y_1 \to Y$$

is a bijection and hence it is an isomorphism. If $B \in B(X,Y)$, then $\|\tilde{A} - \tilde{B}\| \leq \|A - B\|$ and since isomorphisms form an open subset in the space of all bounded transformations (Theorem 2.7) it follows that if $\|A - B\|$ is sufficiently small, then also $\tilde{B}$ is an isomorphism. In particular $\mathcal{N}(B) \cap X_1 = \{0\}$, so $\dim \mathcal{N}(B) \leq \dim N(A) < \infty$, because the quotient map $X \to X/X_1 \simeq \mathcal{N}(A)$ restricted to $\mathcal{N}(B)$ is one-to-one. Moreover for any $y \in Y$ there is $x_1 \in X_1 \subset X$ and $y_1 \in Y$ such that

$$Bx_1 + y_1 = \tilde{B}(x_1, y_1) = y$$

and hence

$$\mathcal{R}(B) + Y_1 = Y$$

which gives

$$\dim(Y/\mathcal{R}(B)) \leq \dim Y_1 = \dim(Y/\mathcal{R}(A)) < \infty.$$

This proves that B is a Fredholm operator and it remains to show that $\operatorname{ind} B = \operatorname{ind} A$.

Since $\mathcal{N}(B) \oplus X_1$ is a closed subspace of X of finite codimension, there is a finitely dimensional subspace $Z \subset X$ such that

$$X = \mathcal{N}(B) \oplus Z \oplus X_1\,, \tag{16.4}$$

so

$$\dim \mathcal{N}(B) \oplus Z = \dim(X/X_1) = \dim \mathcal{N}(A)\,,$$

$$\dim \mathcal{N}(B) = \dim \mathcal{N}(A) - \dim Z. \tag{16.5}$$

The mapping

$$\tilde{B} : X_1 \oplus Y_1 \to Y$$

is an isomorphism, so $Y = B(X_1) \oplus Y_1$ (see Proposition 10.7) and hence $\dim Y_1 = \dim(Y/B(X_1))$. On the other hand (16.4) shows that B restricted to $Z \oplus X_1$ gives an isomorphism onto $\mathcal{R}(B)$, so

$$\mathcal{R}(B) = B(Z) \oplus B(X_1)$$

and in particular $\dim B(Z) = \dim Z$. Hence

$$\begin{aligned} \dim(Y/\mathcal{R}(B)) &= \dim(Y/B(X_1)) - \dim B(Z) = \dim Y_1 - \dim Z \\ &= \dim(Y/\mathcal{R}(A)) - \dim Z \end{aligned}$$

which together with (16.5) immediately implies that $\operatorname{ind} A = \operatorname{ind} B$. □

The above result shows in particular that if $[0,1] \ni t \to A(t)$ is a continuous one parameter family of Fredholm operators, then $\operatorname{ind} A(0) = \operatorname{ind} A(1)$.

Theorem 16.10. *Let $K \in B(X)$ be a compact operator. Then $I + K \in$ Fred (X) and* ind $(I + K) = 0$.

Proof. First we will prove that $\dim \mathcal{N}(I+K) < \infty$. To this end it suffices to prove that the closed unit ball in $\mathcal{N}(I+K)$ is compact, see Theorem 3.7. Let $x_n \in \mathcal{N}(I+K)$, $\|x_n\| \leq 1$. Since K is compact, after passing to a subsequence we may assume that $Kx_n \to y$ converges. Since $x_n + Kx_n = 0$, $x_n \to -y$.

Now we will prove that $\mathcal{R}(I+K)$ is closed. The space $\mathcal{N}(I+K)$ is finitely dimensional, so it is complemented

$$X = \mathcal{N}(I+K) \oplus V \tag{16.6}$$

for some closed subspace $V \subset X$. The mapping $I + K$ restricted to V is a bijection onto $\mathcal{R}(I+K)$ and to prove that $\mathcal{R}(I+K)$ is closed it suffices to show that the inverse mapping is continuous

$$\big((I+K)|_V\big)^{-1} : \mathcal{R}(I+K) \to V\,.$$

We need to prove continuity at 0 only. Suppose the mapping is not continuous at 0. Then there is a sequence $x_n \in V$ such that $(I+K)x_n \to 0$, but $x_n \not\to 0$. Without loss of generality we may assume that $\|x_n\| \geq \varepsilon$. Then the sequence $y_n = x_n/\|x_n\|$ satisfies $\|y_n\| = 1$, $y_n + Ky_n \to 0$. By compactness of K, after passing to a subsequence we have $Ky_n \to y$, $y_n \to -y$. Hence $\|y\| = 1$, $y \in V$, $y \in \mathcal{N}(I+K)$ which is a contradiction with (16.6).

In order to prove that $I+K$ is a Fredholm operator we still need to show that the space $\mathcal{R}(I+K)$ has a finite codimension. Suppose the codimension is infinite. Then we can find a sequence of closed subspaces

$$\mathcal{R}(I+K) = H_0 \subset H_1 \subset H_2 \subset \dots$$

such that $\dim(H_{n+1}/H_n) = 1$. Applying the Riesz lemma (Theorem 3.6), we can find $x_n \in H_n$, $\|x_n\| = 1$ such that

$$\|x_n - y\| \geq \frac{1}{2} \quad \text{for all } y \in H_{n-1}.$$

For $k < n$ we have

$$\|Kx_n - Kx_k\| = \|\underbrace{(x_n + Kx_n)}_{\in H_0} - \underbrace{(x_k + Kx_k)}_{\in H_0} + \underbrace{x_k}_{\in H_k} - x_n\| \geq \frac{1}{2}\,.$$

This shows that the sequence $\{Kx_n\}$ cannot have a convergent subsequence, which contradicts compactness of K. Hence $\mathcal{R}(I+K)$ has finite codimension. We proves that $I+K \in \operatorname{Fred}(X)$. Now $[0,1] \ni t \mapsto I + tK$ is a continuous family of Fredholm operators and from the previous result we have $\operatorname{ind}(I+K) = \operatorname{ind} I = 0$. □

Definition. Let $A, B \in B(X,Y)$. We say that A is *congruent to B modulo compact operators* if $A - B$ is compact and we write

$$A \equiv B \mod K(X,Y)\,.$$

It is easy to see that this is an equivalence relation. Moreover if

$$A \equiv B \mod K(X,Y), \quad A_1 \equiv B_1 \mod K(Y,Z)\,,$$

then

$$A_1 A \equiv B_1 B \mod K(X,Z)\,.$$

Indeed

$$A_1 A - B_1 B = A_1(A-B) + (A_1 - B_1)B$$

is compact by Theorem 16.5.

Definition. We say that $A \in B(X,Y)$ is *invertible modulo compact operators* if there is $A_1 \in B(Y,X)$ such that

$$AA_1 \equiv I_Y \mod K(Y), \quad A_1 A \equiv I_X \mod K(X)\,,$$

i.e. $A_1 A = I_X + K_1$ and $AA_1 = I_Y + K_2$ for some $K_1 \in K(X)$, $K_2 \in K(Y)$. We call A_1 *inverse of A modulo compact operators.*

Theorem 16.11. *Let $A \in B(X,Y)$ be a bounded operator between Banach spaces. Then A is Fredholm if and only if it is invertible modulo compact operators.*

Proof. Let $A \in \operatorname{Fred}(X,Y)$. We can write

$$X = \mathcal{N}(A) \oplus X_1, \quad Y = \mathcal{R}(A) \oplus Y_1.$$

Note that A restricted to X_1 is an isomorphism from X_1 onto $\mathcal{R}(A)$, so it is invertible. Consider the following composition of operators

$$Y = \mathcal{R}(A) \oplus Y_1 \xrightarrow{\pi} \mathcal{R}(A) \xrightarrow{(A|_{X_1})^{-1}} X_1 \xrightarrow{\iota} X\,,$$

where π is the projection onto the first component and ι is the inclusion. Denote this composition by

$$A_1 = \iota \circ (A|_{X_1})^{-1} \circ \pi : Y \to X\,.$$

It is easy to check that $I_Y - AA_1$ is the projection onto Y_1 and $I_X - A_1 A$ is the projection onto $\mathcal{N}(A)$. Both mappings are compact as mappings with finitely dimensional range.[49]

Conversely, suppose that $A \in B(X,Y)$ is invertible modulo compact operators, i.e. there is $A_1 \in B(Y,X)$, $K_1 \in K(X)$, $K_2 \in K(Y)$ such that

$$A_1 A = I_X + K_1, \quad AA_1 = I_Y + K_2\,.$$

Hence $\mathcal{N}(A) \subset \mathcal{N}(A_1 A) = \mathcal{N}(I_X + K_1)$ shows that $\dim \mathcal{N}(A) < \infty$, because $I_X + K_1 \in \mathrm{Fred}\,(X)$ and $\mathcal{R}(I_Y + K_2) = \mathcal{R}(AA_1) \subset \mathcal{R}(A)$ shows that $\mathcal{R}(A)$ has finite codimension, so $A \in \mathrm{Fred}\,(X,Y)$. □

Corollary 16.12. *The composition of Fredholm operators is Fredholm. If A is Fredholm and K is compact, then $A+K$ is Fredholm and* $\mathrm{ind}(A+K) = \mathrm{ind}\,A$.

Proof. Let $A \in \mathrm{Fred}\,(X,Y)$, $B \in \mathrm{Fred}\,(Y,Z)$ and A_1, B_1 be the inverse operators modulo compact ones, i.e.

$$A_1 A = I_X + K_1, \quad AA_1 = I_Y + K_2,$$
$$B_1 B = I_Y + K_3, \quad BB_1 = I_Z + K_4\,.$$

Then

$$A_1 B_1 B A = I_X + \underbrace{K_1 + A_1 K_3 A}_{\text{compact}}, \quad BAA_1B_1 = I_Z + \underbrace{K_4 + BK_2B_1}_{\text{compact}},$$

so $A_1 B_1$ is the inverse of BA modulo compact operators and hence $BA \in \mathrm{Fred}\,(X,Z)$.

A similar argument can be used to show that $A+K$ is Fredholm. Finally, $[0,1] \ni t \mapsto A + tK$ is a continuous family of Fredholm operators, so $\mathrm{ind}\,A = \mathrm{ind}\,(A+K)$. □

Theorem 16.13. *If $A \in \mathrm{Fred}\,(X,Y)$, $B \in \mathrm{Fred}\,(Y,Z)$, then*

$$\mathrm{ind}\,(BA) = \mathrm{ind}\,B + \mathrm{ind}\,A\,.$$

We will not prove it.[50]

[49] Hence we can always select an inverse of A modulo compact operators with finitely dimensional range.

[50] For a proof, see S. Lang, *Real and Functional Analysis.* Third edition, p. 423.

16.2. **Spectrum of compact operators.** In this section we will assume that X is a complex Banach space.

Definition. For a bounded operator $T \in B(X)$ the *spectrum* $\sigma(T)$ is defined as

$$\sigma(T) = \{\lambda \in \mathbb{C} : T - \lambda I \text{ is not invertible}\}.$$

If $\lambda \in \sigma(T)$, then $T - \lambda I$ is not surjective or it is not ont-to-one. In the latter case λ is called *eigenvalue of* T and there is $0 \neq x \in X$ such that

$$Tx = \lambda x.$$

Each such vector is called *eigenvector* of T.

Since $T - \lambda I = \lambda(\lambda^{-1}T - I)$ is invertible for $|\lambda| > \|T\|$, by Theorem 2.7, the spectrum $\sigma(T)$ is a bounded set satisfying $|\lambda| \leq \|T\|$ for $\lambda \in \sigma(T)$.

Theorem 16.14. *If $T \in B(X)$ is a bounded operator, then eigenvectors corresponding to distinct eigenvalues are linearly independent.*

Proof. We argue by induction. Suppose that any collection of $(n-1)$ eigenvectors corresponding to distinct eigenvalues is linearly independent. Suppose that $Tx_i = \lambda_i x_i$, $x_i \neq 0$, $i = 1, 2, \ldots, n$ and $\lambda_i \neq \lambda_j$ for $i \neq j$. We need to prove that the vectors $x_1, \ldots, x_n$ are linearly independent. Let $c_1, \ldots, c_n$ be scalars such that

$$c_1 x_1 + \ldots + c_n x_n = 0. \tag{16.7}$$

Applying the operator T we have

$$c_1 \lambda_1 x_1 + \ldots + c_n \lambda_n x_n = 0.$$

At least one of the eigenvalues is nonzero, say $\lambda_1 \neq 0$. Dividing the second equation by λ_1 and subtracting from the first one gives

$$c_2 \underbrace{(1 - \lambda_2/\lambda_1)}_{\neq 0} x_1 + \ldots + c_n \underbrace{(1 - \lambda_n/\lambda_1)}_{\neq 0} x_n = 0.$$

Since by the assumption the vectors $x_2, \ldots, x_n$ are linearly independent, $c_2 = \ldots = c_n = 0$ and hence (16.7) gives $c_1 = 0$ which proves linear independence of $x_1, \ldots, x_n$. $\square$

Theorem 16.15. *If $A \in B(X)$ is compact, then every nonzero element in the spectrum $0 \neq \lambda \in \sigma(A)$ is an eigenvalue. If $\dim X = \infty$, then $0 \in \sigma(A)$.*

Proof. If $0 \neq \lambda \in \sigma(A)$, then $A - \lambda I$ is a noninvertible Fredholm operator. Since $\operatorname{ind}(A - \lambda I) = 0$, $\mathcal{N}(A - \lambda I) \neq 0$ and hence there is $0 \neq x \in X$ such that $Ax = \lambda x$. If $\dim X = \infty$, then A is not surjective by Corollary 3.7 and hence $A = A - 0 \cdot I$ is not invertible, so $0 \in \sigma(A)$. $\square$

Theorem 16.16. *If $A \in B(X)$ is compact, then eigenvalues from a finite or a countable set. If there are infinitely many eigenvalues, then we can order them as $\lambda_1, \lambda_2, \lambda_3, \ldots$ in such a way that*

$$|\lambda_1| \geq |\lambda_2| \geq |\lambda_3| \geq \ldots \qquad \lim_{n\to\infty} \lambda_n = 0.$$

Proof.[51] It suffices to prove that for any $r > 0$ the number of eigenvalues satisfying $|\lambda| \geq r$ is finite. If this is not true, then there are distinct eigenvalues λ_i, $|\lambda_i| \geq r$, $i = 1, 2, 3, \ldots$ and corresponding eigenvectors x_i, $\|x_i\| = 1$. Let

$$H_n = \operatorname{span}\{x_1, \ldots, x_n\}.$$

By the Riesz lemma (Theorem 3.6) we can find $w_n \in H_n$ such that

$$\|w_n\| = 1, \quad \|w_n - y\| \geq \frac{1}{2} \text{ for any } y \in H_{n-1}.$$

We have $w_n = a_n x_n + y_{n-1}$, $y_{n-1} \in H_{n-1}$ Thus for $k < n$, $Aw_k \in H_k \subset H_{n-1}$ and hence

$$\begin{aligned} \|Aw_n - Aw_k\| &= \|a_n\lambda_n x_n + Ay_{n-1} - Aw_k\| \\ &= |\lambda_n| \| \underbrace{a_n x_n + y_{n-1}}_{w_n} - (\underbrace{y_{n-1} - \lambda_n^{-1}(Ay_{n-1} - Aw_k)}_{\in H_{n-1}})\| \geq \frac{r}{2}. \end{aligned}$$

Therefore the sequence $\{Aw_n\}$ has no convergent subsequence which contradicts compactness of A. $\square$

16.3. **The Fredholm-Riesz-Schauder theory.** Our aim is to apply the above theory to the following problem. Given a compact operator $A \in B(X)$ in a Banach space X and a parameter $0 \neq \lambda \in \mathbb{C}$ we want to solve the equation

$$Ax - \lambda x = y \quad \text{where } y \in X \text{ is given and } x \in X \text{ is unknown.} \tag{16.8}$$

Theorem 16.17. *If $\lambda \neq 0$ is not an eigenvalue of A, then for every $y \in X$, (16.8) has a unique solution.*

Proof. By Theorem 16.15, $\lambda \notin \sigma(A)$, so $A - \lambda I$ is invertible and hence (16.8) has the unique solution

$$x = (A - \lambda I)^{-1} y.$$

The proof is complete. $\square$

Note that by Theorem 16.16 we have at most countably many eigenvalues that form a discrete set whenever we are away from 0. Thus for most of the values of λ, (16.8) can be uniquely solved for any $y \in X$. The interesting case is, however, when λ is an eigenvalue of A.

[51]The proof is very similar to the proof that if K is compact, then $\mathcal{R}(I + K)$ has finite codimension, see the proof of Theorem 16.10.

Along with the equation (16.8) we consider the adjoint equation

$$A^*x^* - \lambda x^* = y^* . \tag{16.9}$$

The first part of the following definition already appears in Section 4.2.

Definition. Let X be a normed space. For linear subspaces $M \subset X$, $N \subset X^*$ we define *annihilators*

$$M^\perp = \{x^* \in X^* : \langle x^*, x\rangle = 0 \text{ for all } x \in M\} ,$$

$$^\perp N = \{x \in X : \langle x^*, x\rangle = 0 \text{ for all } x^* \in N\} .$$

Clearly $M^\perp$ and $^\perp N$ are closed subspaces of X^* and X respectively.

The following result provides a complete description of solvability of the equations (16.8) and (16.9).

Theorem 16.18 (Riesz-Schauder)**.** *If X is a Banach space, $A \in B(X)$ is compact and $\lambda \neq 0$, then*

$$\dim \mathcal{N}(A - \lambda I) = \dim \mathcal{N}(A^* - \lambda I) < \infty . \tag{16.10}$$

Moreover

$$\mathcal{R}(A - \lambda I) = {}^\perp\mathcal{N}(A^* - \lambda I), \quad \mathcal{R}(A^* - \lambda I) = \mathcal{N}(A - \lambda I)^\perp . \tag{16.11}$$

Before we prove the theorem we will see how it applies to the equations (16.8) and (16.9). The following description is called the *Fredholm alternative.*

It follows from (16.10) that $0 \neq \lambda \in \mathbb{C}$ is not an eigenvalue of A if an only if it is not an eigenvalue of A^* and in this case both equations

$$Ax - \lambda x = y, \quad A^*x^* - \lambda x^* = y^*$$

have unique solutions for all $y \in X$ and $y^* \in X^*$.

Also $0 \neq \lambda \in \mathbb{C}$ is an eigenvalue of A if and only if it is an eigenvalue of A^* and the dimensions corresponding of eigenspaces for A and A^* are finite and equal. Moreover the equation

$$Ax - \lambda x = y$$

has a solution if and only if $y \in {}^\perp\mathcal{N}(A^* - \lambda I)$, i.e. if

$$\langle y^*, y\rangle = 0 \quad \text{whenever } A^*y^* = \lambda y^*.$$

Note that since $\dim(A^* - \lambda I) < \infty$, this is a finite number of conditions to check. Similarly the equation

$$A^*x^* - \lambda x^* = y^*$$

has a solution if and only if $y^* \in \mathcal{N}(A - \lambda I)^\perp$, i.e.

$$\langle y^*, y\rangle = 0 \quad \text{whenever } Ay = \lambda y.$$

The above theory was first considered by Fredholm in the setting of integral equations. Let $\Omega \subset \mathbb{R}^n$ be an open set and let $K \in L^2(\Omega \times \Omega)$. As we know the operator $K : L^2(\Omega) \to L^2(\Omega)$,

$$Kf(x) = \int_\Omega K(x,y)f(y)\,dy$$

is compact, see Theorem 16.7. We consider the *Fredholm integral equation*

$$f(x) - \mu \int_\Omega K(x,y)f(y)\,dy = g(x)\,, \tag{16.12}$$

where $0 \neq \mu \in \mathbb{C}$ and $g \in L^2(\Omega)$ are given.

Theorem 16.19 (Fredholm)**.**

(a) *The equation (16.12) has a solution for every $g \in L^2(\Omega)$ if and only if the only solution to*

$$f(x) - \mu \int_\Omega K(x,y)f(y)\,dy = 0 \tag{16.13}$$

is $f = 0$.

(b) *The equation (16.13) has nonzero solutions, if and only if there are nonzero solutions to*

$$f(y) - \overline{\mu} \int_\Omega \overline{K(x,y)}f(x)\,dx = 0 \tag{16.14}$$

The spaces of solutions to (16.13) and (16.14) are finitely dimensional and have the same dimension.

(c) *If the equation (16.13) has nonzero solutions, then (16.12) has solutions if and only if*

$$\int_\Omega g(x)\overline{f(x)}\,dx$$

for every $f \in L^2(\Omega)$ that solves (16.14).

(d) *The set of numbers μ for which (16.13) has nonzero solutions is at most countable. If this set is infinite, then the numbers form an infinite sequence μ_n, $|\mu_n| \to \infty$.*

Proof. The equation (16.12) can be written as

$$Kf - \lambda f = -\mu^{-1}g, \quad \lambda = \mu^{-1}\,.$$

It has solutions for all $g \in L^2(\Omega)$ if and only if $\mathcal{R}(K - \lambda I) = L^2(\Omega)$ which us equivalent to $\mathcal{N}(K - \lambda I) = \{0\}$, so (a) follows.

The equation (16.13) with $\mu \neq 0$ has nonzero solutions if and only if $\lambda = \mu^{-1}$ is an eigenvalue of K. Thus (d) follows from Theorem 16.16.

The dual space to $L^2(\Omega)$ is isometrically isomorphic to $L^2(\Omega)$ and the isomorphism is given by[52]

$$L^2(\Omega) \ni f \mapsto \Lambda_f \in (L^2(\Omega))^*, \quad \langle \Lambda_f, g\rangle = \int_\Omega g(x)f(x)\,dx\,.$$

We have

$$\begin{aligned}\langle K^*\Lambda_f, g\rangle &= \langle \Lambda_f, Kg\rangle = \int_\Omega \Big(\int_\Omega K(x,y)g(y)\,dy\Big)f(x)\,dx\\ &= \int_\Omega g(y)\Big(\int_\Omega K(x,y)f(x)\,dx\Big)\,dy\,,\end{aligned}$$

so

$$K^*\Lambda_f = \Lambda_{\tilde{K}f}, \quad \tilde{K}f(y) = \int_\Omega K(x,y)f(x)\,dx\,.$$

Hence the adjoint operator to $K - \lambda I$ is $\tilde{K} - \lambda I$ and the equation adjoint to (16.13) reads as

$$f(y) - \mu\int_\Omega K(x,y)f(x)\,dx = 0\,. \tag{16.15}$$

Now (16.12) has a solution if and only if $g \in {}^\perp\mathcal{N}(\tilde{K} - \lambda I) = {}^\perp\mathcal{N}(K^* - \lambda I)$, i.e.

$$\langle \Lambda_f, g\rangle = 0 \quad \text{whenever } K^*\Lambda_f = \lambda\Lambda_f\,,$$

i.e.

$$\int_\Omega g(x)f(x)\,dx = 0 \quad \text{whenever } f \text{ solves (16.15)}.$$

Since f solves (16.15) if and only if $\overline{f}$ solves (16.14), the part (c) follows. □

We prepare for the proof of Theorem 16.18 now.

Theorem 16.20. *Let X and Y be normed spaces and $T \in B(X,Y)$. Then*

$$\mathcal{N}(T^*) = \mathcal{R}(T)^\perp \quad \text{and} \quad \mathcal{N}(T) = {}^\perp\mathcal{R}(T^*)\,.$$

Proof. It follows from an obvious sequence of identifications. □

However, we will need the following less obvious results.

Theorem 16.21. *Let X and Y be normed spaces and $T \in B(X,Y)$. Then*

$$\overline{\mathcal{R}(T)} = {}^\perp\mathcal{N}(T^*)\,.$$

Recall that

$${}^\perp\mathcal{N}(T^*) = \{y \in Y : \langle y^*, y\rangle = 0 \text{ whenever } T^*y^* = 0\}\,.$$

[52]We refer to Theorem 2.13 rather than to Theorem 5.5.

First we will prove that ${}^\perp\mathcal{N}(T^*) \subset \overline{\mathcal{R}(T)}$. If not, then there is $y_0 \in {}^\perp\mathcal{N}(T^*) \setminus \overline{\mathcal{R}(T)}$. The Hahn-Banach theorem (Theorem 11.6) yields the existence of $y^* \in Y^*$ such that

$$\langle y^*, y_0\rangle \neq 0 \quad \text{and} \quad \langle y^*, y\rangle = 0 \text{ for all } y \in \overline{\mathcal{R}(T)}.$$

The second condition implies that for all $x \in X$

$$\langle T^*y^*, x\rangle = \langle y^*, Tx\rangle = 0\,,$$

so $T^*y^* = 0$ and hence $\langle y^*, y_0\rangle \neq 0$ implies that $y_0 \notin {}^\perp\mathcal{N}(T^*)$ which is a contradiction.

It remains to prove that $\overline{\mathcal{R}(T)} \subset {}^\perp\mathcal{N}(T^*)$. Let $y \in \overline{\mathcal{R}(T)}$. Then there is a sequence $x_n \in X$ such that $Tx_n \to y$. If $T^*y^* = 0$, then

$$\langle y^*, y\rangle = \lim_{n\to\infty} \langle y^*, Tx_n\rangle = \lim_{n\to\infty}\langle T^*y^*, x_n\rangle = 0$$

which proves $y \in {}^\perp\mathcal{N}(T^*)$. □

Theorem 16.22. *Let X and Y be Banach spaces and $T \in B(X,Y)$. If $\mathcal{R}(T)$ is closed, then*

$$\mathcal{R}(T^*) = \mathcal{N}(T)^\perp\,.$$

Proof. First we prove that

$$\mathcal{R}(T^*) \subset \mathcal{N}(T)^\perp = \{x^* \in X^* :\, \langle x^*, x\rangle = 0 \text{ whenever } Tx = 0\}\,.$$

If $x^* \in \mathcal{R}(T^*)$, then $x^* = T^*y^*$ for some $y^* \in Y^*$. Then for any $x \in X$ satisfying $Tx = 0$ we have

$$\langle x^*, x\rangle = \langle T^*y^*, x\rangle = \langle y^*, Tx\rangle = 0$$

so $x^* \in \mathcal{N}(T)^\perp$.

The proof of $\mathcal{N}(T)^\perp \subset \mathcal{R}(T^*)$ is more difficult.

$\mathcal{R}(T)$ is a Banach space as a closed subspace of Y. The mapping

$$\tilde{T} : X/\ker T \to \mathcal{R}(T), \quad \tilde{T}([x]) = Tx$$

is a well defined bounded bijection. Hence it is an isomorphism. Thus the inverse mapping

$$\tilde{T}^{-1}(Tx) = [x]$$

is also bounded. That easily implies that there is $C > 0$ such that for every $y \in \mathcal{R}(T)$ we can find $x \in X$ such that $Tx = y$ and $\|x\| \leq C\|y\|$.

Let $x^* \in \mathcal{N}(T)^\perp$. We have to find $y^* \in Y^*$ such that $T^*y^* = x^*$. Since x^* vanishes on all $x \in X$ such that $Tx = 0$ it easily follows that

$$\langle \Lambda, Tx\rangle = \langle x^*, x\rangle$$

is a well defined linear functional on $\mathcal{R}(T)$. This functional is bounded. Indeed, if $y \in \mathcal{R}(T)$, then there is $x \in X$ such that $Tx = y$, $\|x\| \leq c\|y\|$ and hence

$$|\langle \Lambda, y\rangle| = |\langle \Lambda, Tx\rangle| = |\langle x^*, x\rangle| \leq C\|x^*\|\|y\|\,.$$

By the Hahn-Banach theorem Λ can be extended to $y^* \in Y^*$. Now for any $x \in X$ we have

$$\langle T^*y^*, x\rangle = \langle y^*, Tx\rangle = \langle \Lambda, Tx\rangle = \langle x^*, x\rangle$$

and hence $T^*y^* = x^*$. □

Proof of Theorem 16.18. If $T = A - \lambda I$, then $\mathcal{R}(T) = \mathcal{R}(A - \lambda I)$ is closed and hence

$$\mathcal{R}(A - \lambda I) = \overline{\mathcal{R}(T)} = {}^{\perp}\mathcal{N}(T^*) = {}^{\perp}\mathcal{N}(A^* - \lambda I)$$

by Theorem 16.21. The second equality in (16.11) follows from Theorem 16.22 since

$$\mathcal{R}(A^* - \lambda I) = \mathcal{R}(T^*) = \mathcal{N}(T)^{\perp} = \mathcal{N}(A - \lambda I)^{\perp}\,.$$

We are left with the proof of (16.10). By the Schauder theorem (Theorem 16.8) $A^* \in B(X^*)$ is compact, so both operators $A - \lambda I$ and $A^* - \lambda I$ are Fredholm and hence the null spaces are finitely dimensional. We need to prove that the dimensions are equal.

We start with a general observation. If $x_1, \ldots, x_n$ are linearly independent elements in a normed space X, then it follows from the Hahn-Banach theorem (Theorem 11.6) that there are functionals $x_1^*, \ldots, x_n^* \in X^*$ such that $\langle x_i^*, x_j\rangle = \delta_{ij}$. The converse result is also true.[53]

Lemma 16.23. *If $x_1^*, \ldots, x_n^*$ are linearly independent elements in X^*, then there exist points $x_1, \ldots, x_n \in X$ such that $\langle x_i^*, x_j\rangle = \delta_{ij}$.*

Proof. Let $N_j = \ker x_j^*$ and

$$M_j = N_1 \cap \ldots \cap N_{j-1} \cap N_{j+1} \cap \ldots \cap N_n\,.$$

It suffices to prove that $M_j \setminus N_j \neq \emptyset$ for $j = 1, 2, \ldots, n$. We will prove that $M_1 \setminus N_1 \neq \emptyset$ as the argument for $j \neq 1$ is the same. Suppose by contradiction that $M_1 \subset N_1$, so

$$\langle x_1^*, x\rangle = 0 \quad \text{whenever } \langle x_j^*, x\rangle = 0 \text{ for } j = 2, \ldots, n. \tag{16.16}$$

Consider the linear mapping

$$A : X \to \mathbb{C}^{n-1}, \quad Ax = [\langle x_2^*, x\rangle, \ldots, \langle x_n^*, x\rangle]$$

and define on $A(X)$ a linear functional Λ

$$\langle \Lambda, Ax\rangle = \langle x_1^*, x\rangle\,.$$

[53] It holds in any linear space, not only in a normed space as the proof is based on linear algebra only.

Note that this functional is well defined, because if $Ax = A\tilde{x}$, then $\langle x_1, x\rangle = \langle x_1^*, \tilde{x}\rangle$ by (16.16). The functional Λ can be extended linearly to a functional[54] on $\mathbb{C}^{n-1}$, so it is of the form

$$\langle \Lambda, [y_2, \ldots, y_n]\rangle = \sum_{j=2}^{n} a_j y_j$$

for some scalars $a_j \in \mathbb{C}$. This, however gives for any $x \in X$

$$\langle x_1^*, x\rangle = \Big\langle \sum_{j=2}^{n} a_j x_j^*, x\Big\rangle$$

which contradicts linear independence of functionals. $\square$

Now we can complete the proof of the theorem. Let $x_1, \ldots, x_n$ be a Hamel basis of $\mathcal{N}(A - \lambda I)$ and $y_1^*, \ldots, y_m^*$ be a Hamel basis of $\mathcal{N}(A^* - \lambda I)$.

As we observed above there are elements $x_1^*, \ldots, x_n^* \in X^*$ and $y_1, \ldots, y_m \in X$ such that

$$\langle x_i^*, x_j\rangle = \delta_{ij}, \quad \langle y_i^*, y_j\rangle = \delta_{ij}\,.$$

Suppose that $n < m$. Consider the operator

$$Bx = Ax + \sum_{i=1}^{n} \langle x_i^*, x\rangle y_i\,.$$

This is a sum of a compact operator and an operator with a finitely dimensional range, so B is compact. We claim that $\mathcal{N}(B - \lambda I) = \{0\}$. Indeed, if $x \in \mathcal{N}(B - \lambda I)$, then $Bx = \lambda x$, i.e.

$$\lambda x - Ax = \sum_{i=1}^{n} \langle x_i^*, x\rangle y_i\,.$$

Since $y_j^* \in \mathcal{N}(A^* - \lambda I)$ for $j = 1, 2, \ldots, n$ we have

$$0 = \langle \lambda y_j^* - A^* y_j^*, x\rangle = \langle y_j^*, \lambda x - Ax\rangle = \langle x_j^*, x\rangle\,,$$

so $\lambda x - Ax = 0$, $x \in \mathcal{N}(A - \lambda I)$. Thus $x = \sum_{i=1}^{n} a_i x_i$, but $0 = \langle x_j^*, x\rangle = a_j$ and hence $x = 0$. We prove that $\mathcal{N}(B - \lambda I) = \{0\}$. Since $B - \lambda I$ is Fredholm of index 0 we conclude that $\mathcal{R}(B - \lambda I) = X$. In particular there is $x \in X$ such that

$$Bx - \lambda x = y_{n+1}\,,$$

but this yields

$$1 = \langle y_{n+1}^*, y_{n+1}\rangle = \langle y_{n+1}^*, Bx - \lambda x\rangle = \Big\langle y_{n+1}^*, Ax - \lambda x + \sum_{i=1}^{n} \langle x_i^*, x\rangle y_i \Big\rangle = 0$$

[54] Linear algebra.

since $y^*_{n+1} \in \mathcal{N}(A^* - \lambda I)$ and $\langle y^*_{n+1}, y_i \rangle = 0$ for $i = 1, 2, \ldots, n$. The contradiction proves that $n \geq m$, i.e.

$$\dim \mathcal{N}(A - \lambda I) \geq \dim \mathcal{N}(A^* - \lambda I) .$$

Applying this result to A^* in place of A we have

$$\dim \mathcal{N}(A^* - \lambda I) \geq \dim \mathcal{N}(A^{**} - \lambda I) .$$

If $x \in \mathcal{N}(A - \lambda I)$, then it easily follows that $\kappa(x) \in \mathcal{N}(A^{**} - \lambda I)$ and hence

$$\dim \mathcal{N}(A^{**} - \lambda I) \geq \dim \mathcal{N}(A - \lambda I) ,$$

so the above inequalities give (16.10).

16.4. **Spectral theorem.**

Definition. Let H be a Hilbert space. We say that a linear operator $T : H \to H$ is *self-adjoint* if

$$\langle Tx, y \rangle = \langle x, Ty \rangle \quad \text{for all } x, y \in H.$$

It immediately follows from the Hellinger-Toeplitz theorem (Theorem 10.5) that self-adjoint operators are bounded.

Theorem 16.24. *If $T \in B(H)$ is self-adjoint, then*

(a) *$\langle Ax, x \rangle$ is real for all $x \in H$,*
(b) *Eigenvalues of T are real,*
(c) *Eigenspaces of T corresponding to distinct eigenvalues are orthogonal.*

Proof. Since $\langle Ax, x \rangle = \langle x, Ax \rangle = \overline{\langle Ax, x \rangle}$, (a) follows. If $Tx = \lambda x$, $x \neq 0$, then

$$\lambda \langle x, x \rangle = \langle Tx, x \rangle = \langle x, Tx \rangle = \overline{\lambda} \langle x, x \rangle$$

so $\lambda \in \mathbb{R}$ which is (b). If $\lambda_1 \neq \lambda_2$ are eigenvalues and $x_1, x_2 \neq 0$ are corresponding eigenvectors, then using the fact that $\lambda_2 \in \mathbb{R}$ we have

$$\lambda_1 \langle x_1, x_2 \rangle = \langle Tx_1, x_2 \rangle = \langle x_1, Tx_2 \rangle = \lambda_2 \langle x_1, x_2 \rangle ,$$

so $\langle x_1, x_2 \rangle = 0$ which is (c). □

The spectrum of a bounded operator is bounded. more precisely if $\lambda \in \sigma(T)$, then $|\lambda| \leq \|T\|$ (see a remark proceeding Theorem 16.14).

Theorem 16.25. *If $A \in B(H)$ is compact and self-adjoint, then $\|A\|$ or $-\|A\|$ is an eigenvalue of A.*

Proof. Let $x_n \in H$ be a sequence such that $\|x_n\| = 1$, $\|Ax_n\| \to \|A\|$. After passing to a subsequence we may assume that $Ax_n \to y \in H$, so $\|y\| = \|A\|$ and $A^2 x_n \to Ay$. We have

$$\|Ay\| = \lim_{n\to\infty} \|A^2 x_n\| \overset{\text{Schwarz}}{\geq} \lim_{n\to\infty} \langle A^2 x_n, x_n \rangle = \lim_{n\to\infty} \langle Ax_n, Ax_n \rangle = \|A\|^2 .$$

Hence

$$\|A^2y\|\,\|y\| \geq \langle A^2y, y\rangle = \|Ay\|^2 \geq \|A\|^4 = \|A\|^2\|y\|^2 \geq \|A^2y\|\,\|y\|\,,$$

so

$$\langle A^2y, y\rangle = \|A^2y\|\,\|y\|.$$

This is possible only if the vectors A^2y and y are parallel, $A^2y = cy$. We have

$$c = \frac{\langle A^2y, y\rangle}{\langle y, y\rangle} = \frac{\|A\|^4}{\|A\|^2} = \|A\|^2\,.$$

Let $x = y + \|A\|^{-1}\,Ay$. If $x = 0$, then $Ay = -\|A\|\,y$ and hence $-\|A\|$ is an eigenvalue of A. If $x \neq 0$, then

$$Ax = Ay + \|A\|^{-1}A^2y = Ay + \|A\|^{-1}\|A\|^2y = Ay + \|A\|\,y = \|A\|\,x\,,$$

so $\|A\|$ is an eigenvalue. □

Let $A \in B(H)$ be compact and self-adjoint. According to Theorem 16.16 all nonzero eigenvalues form a finite or infinite sequence $\{\lambda_i\}_{i=1}^N$ (N is finite or $N = \infty$). If $N = \infty$ and we order the eigenvalues in such a way that $|\lambda_1| \geq |\lambda_2| \geq |\lambda_3| \geq \ldots$, then

$$\lim_{i\to\infty} \lambda_i = 0.$$

Let

$$E_{\lambda_i} = \{x \in H : \ Ax = \lambda_i x\}\,.$$

E_{λ_i} is the eigenspace corresponding to the eigenvalue λ_i. Since $E_{\lambda_i} = \mathcal{N}(A - \lambda_i I)$,

$$\dim E_{\lambda_i} < \infty.$$

Moreover Theorem 16.24 yields

$$E_{\lambda_i} \perp E_{\lambda_j} \quad \text{for } i \neq j.$$

Theorem 16.26 (Spectral theorem)**.** *Let $A \in B(H)$ be compact and self-adjoint. Then*

$$H = \ker A \oplus \bigoplus_{i=1}^N E_{\lambda_i}\,.$$

Proof. Let

$$Y = \bigoplus_{i=1}^N E_{\lambda_i}.$$

Then $H = Y \oplus Y^\perp$ and it remains to show that $Y^\perp = \ker A$. It is easy to see that $A(Y) \subset Y$ and hence for every $x \in Y^\perp$ and $y \in Y$ we have

$$\langle Ax, y\rangle = \langle x, Ay\rangle = 0\,,$$

i.e. $Ax \in Y^\perp$. Hence $A|_{Y^\perp} : Y^\perp \to Y^\perp$ is a compact self-adjoint operator. According to Theorem 16.25 at least one of the numbers $\pm\|A|_{Y^\perp}\|$ is its eigenvalue. It cannot be a nonzero eigenvalue, because all eigenvectors with nonzero eigenvalues are contained in Y. Thus $\|A|_{Y^\perp}\| = 0$, $A|_{Y^\perp} = 0$, i.e. $Y^\perp = \ker A$. $\square$

Note that $\ker A = E_0$ is the eigenspace corresponding to the zero eigenvalue and hence the spectral theorem says that the Hilbert space is the direct sum of eigenspaces.

Example. Let $\{\lambda_i\}_{i=1}^\infty \subset \mathbb{C}$ be any sequence such that $\lambda_i \to 0$

EXERCISE. *Prove that the operator* $A : \ell^2 \to \ell^2$,

$$A(x_1, x_2, x_3, \ldots) = (\lambda_1 x_1, \lambda_2 x_2, \lambda_3 x_3, \ldots)$$

is compact.

Hence we can find a compact operator with a prescribed set of eigenvalues. If $\lambda_i \neq 0$ for all i, then $\ker A = \{0\}$, so compact operators can be one-to-one even in the infinitely dimensional spaces. If $\lambda_i \in \mathbb{R}$ for all i, then A is compact self-adjoint.

EXERCISE. *Let* $K \in L^2(\Omega\times\Omega)$. *Prove that the integral operator* $K : L^2(\Omega) \to L^2(\Omega)$,

$$Kf(x) = \int_\Omega K(x,y) f(y)\, dy$$

is self-adjoint if and only if

$$K(x,y) = \overline{K(y,x)} \quad \textit{a.e. in } \Omega\times\Omega.$$

16.5. **Sobolev spaces and the eigenfunctions od the Laplace oprtator.** In Section 7 we proved that the eigenfunctions of the Laplace operator on the sphere[55] form an orthonormal basis in $L^2(S^{n-1})$ (Theorem 7.1 and Theorem 7.5). In this section we will prove s similar result: suitable eigenfunctions of the Laplace operator form an orthonormal basis in $L^2(\Omega)$, where $\Omega \subset \mathbb{R}^n$ is any bounded open set. This will be a consequence of the spectral theorem from the previous section. To prove, or even to state the result, we need to introduce Sobolev spaces which is one of the main tools in the theory of partial differential equations and calculus of variations.

DEFINITION. Let $\Omega \subset \mathbb{R}^n$ be an open set, $u, v \in L^1_{\rm loc}(\Omega)$ and let α be a miltiindex. We say that $D^\alpha u = v$ in the *weak sense* if

$$\int_\Omega v\varphi = (-1)^{|\alpha|} \int_\Omega u D^\alpha \varphi \quad \text{for all } \varphi \in C_0^\infty(\Omega).$$

[55]Laplace-Beltrami.

The weak derivative (if it exists) is unique. Indeed, if $D^\alpha u = v_1$, $D^\alpha u = v_2$ weakly, then

$$\int_\Omega (v_1 - v_2)\varphi = 0 \quad \text{for all } \varphi \in C_0^\infty(\Omega)$$

and hence $v_1 = v_2$ a.e.

If $u \in C^\infty(\Omega)$, then the classical partial derivative $D^\alpha u$ equals to the weak one due to the integration by parts formula.

Definition. Let $1 \leq p \leq \infty$ and let $m \geq 1$ be an integer. The *Sobolev space* $W^{m,p}(\Omega)$ is the space of all functions $u \in L^p(\Omega)$ such that the weak derivatives $D^\alpha u$ exist for all $|\alpha| \leq m$ and belong to $L^p(\Omega)$. The Sobolev space is equipped with the norm

$$\|u\|_{m,p} = \sum_{|\alpha| \leq m} \|D^\alpha u\|_p \, .$$

Theorem 16.27. *$W^{m,p}(\Omega)$ is a Banach space.*

Proof. If $\{u_k\} \subset W^{m,p}(\Omega)$ is a Cauchy sequence, then for every α, $|\alpha| \leq m$, $\{D^\alpha u_k\}$ is a Cauchy sequence in $L^p(\Omega)$, so $D^\alpha u_k$ converges to some function $u_\alpha \in L^p(\Omega)$ (we will write$u$ instead of u_0). Since

$$\int_\Omega u D^\alpha \varphi \leftarrow \int_\Omega u_k D^\alpha \varphi = (-1)^{|\alpha|} \int_\Omega D^\alpha u_k \varphi \to (-1)^{|\alpha|} \int_\Omega u_\alpha \varphi$$

we conclude that $D^\alpha u = u_\alpha$ weakly, so $u \in W^{m,p}(\Omega)$ and $u_k \to u$ in the norm of $W^{m,p}(\Omega)$. $\Box$

Theorem 16.28 (Meyers-Serrin)**.** *For $1 \leq p < \infty$ smooth functions are dense in $W^{m,p}(\Omega)$, i.e. for every $u \in W^{m,p}(\Omega)$ there exist a sequence $u_k \in C^\infty(\Omega)$, $\|u_k\|_{m,p} < \infty$ such that $\|u - u_k\|_{m,p} \to 0$ as $k \to \infty$.*

We will not prove it. The space of functions $u \in C^\infty(\Omega)$ with $\|u\|_{m,p} < infty$ forms a normed space with respect to the Sobolev norm and the above theorem shows that the Sobolev space can be equivalently defined as its completion.

Theorem 16.29. *For $1 < p < \infty$ the Sobolev space $W^{m,p}(\Omega)$ us reflexive.*

Proof. Let N be the number of multiindices $|\alpha| \leq m$. The space

$$\underbrace{L^p(\Omega) \oplus \ldots \oplus L^p(\Omega)}_{N \text{ times}} = L^p(\Omega)^N$$

is reflexive, so is any of its closed subspaces. The mapping

$$T : W^{m,p}(\Omega) \to L^p(\Omega)^N, \quad T(u) = (D^\alpha u)_{|\alpha| \leq N}$$

is an isometric embedding, so its image $\mathcal{R}(T)$ is a closed subspace of $L^p(\Omega)^N$. Hence $W^{m,p}(\Omega)$ is reflexive as isomorphic to a reflexive space. $\Box$

The notion of weak derivative can be generalized as follows.

DEFINITION. If $P = \sum_{|\alpha|\le m} a_\alpha D^\alpha$ is a differential operator, with smooth coefficients and $u, v \in L^1_{\mathrm{loc}}(\Omega)$, then we say that $Pu = v$ in Ω in the *weak sense* if

$$\int_\Omega v\varphi = \int_\Omega u \sum_{|\alpha|\le m} (-1)^{|\alpha|} D^\alpha(a_\alpha\varphi) \quad \text{for all } \varphi \in C_0^\infty(\Omega).$$

In particular $\Delta u = 0$ in weak sense if $\int_\Omega u\Delta\varphi = 0$ for all $\varphi \in C_0^\infty(\Omega)$.

As before Pu is unique (if it exists) and coincides with the classical differential operator applied to u is u is smooth.

DEFINITION. For $1 \le p < \infty$ and $m \ge 1$ we define $W_0^{m,p}(\Omega)$ as the closure of $C_0^\infty(\Omega)$ in the Sobolev norm.

Theorem 16.30 (Poincaré lemma)**.** *If $\Omega \subset \mathbb{R}^n$ is bounded, $1 \le p < \infty$ and $m \ge 1$, then there is a constant $C = C(m,p,\Omega)$ such that*

$$\|u\|_{m,p} \le C \sum_{|\alpha|=m} \|D^\alpha u\|_p \quad \textit{for all } u \in W_0^{m,p}(\Omega).$$

Proof. By the density argument it suffices to prove the result for $u \in C_0^\infty(\Omega)$ and it actually suffices to prove the inequality

$$\|u\| \le C \sum_{i=1}^n \Big\| \frac{\partial u}{\partial x_i} \Big\|_p$$

as the general case follows by iteration of this inequality.

Let $M > 0$ be such that $\Omega \subset [-M, M]^n$. Then for every $x \in [-M, M]^n$

$$u(x) = \int_{-M}^{x_1} \frac{\partial u}{\partial x_1}(t, x_2, \dots, x_n)\, dt \le \int_{-M}^{M} \Big| \frac{\partial u}{\partial x_1} \Big|\, dx\,.$$

Applying the Hölder inequality we get

$$|u(x)|^p \le 2^{p-1} M^{p-1} \int_{-M}^{M} \Big| \frac{\partial u}{\partial x_1} \Big|^p\, dt$$

and the result follows by integration with respect to $x \in \Omega$. □

The result shows that if Ω is abounded set, then

$$\|\nabla^m u\|_p := \sum_{|\alpha|=m} \|D^\alpha u\|_p$$

is an norm on $W_0^{m,p}$ equivalent to $\|\cdot\|_{m,p}$. Note that it is not a norm in $W^{m,p}(\Omega)$ because it vanishes on constant functions. In particular this shows that nonzero constant functions do not belong to $W_0^{m,p}(\Omega)$.

Theorem 16.31 (Rellich-Kondrachov)**.** *If $\Omega \subset \mathbb{R}^n$ is bounded and $1 \le p < \infty$, then the embedding, i.e. the inclusion*

$$W_0^{1,p}(\Omega) \subset L^p(\Omega)$$

is a compact operator.

We will not prove this result.

Consider the following *Dirichlet problem*. Given a bounded open set $\Omega \subset \mathbb{R}^n$, find $f \in W_0^{1,2}(\Omega)$ such that

$$-\Delta f = g \quad \text{weakly in } \Omega, \tag{16.17}$$

i.e.

$$\int_\Omega \nabla f \cdot \nabla \varphi = \int_\Omega g\varphi \quad \text{for all } \varphi \in C_0^\infty(\Omega).$$

Since $C_0^\infty(\Omega)$ is dense in $W_0^{1,2}(\Omega)$ this condition is equivalent to

$$\int_\Omega \nabla f \cdot \nabla h = \int_\Omega gh \quad \text{for all } h \in W_0^{1,2}(\Omega). \tag{16.18}$$

Theorem 16.32. *For every $g \in L^2(\Omega)$ there is a unique solution $f \in W_0^{1,2}(\Omega)$ of the Dirichlet problem (16.17).*

Proof. Since $f \mapsto \|\nabla f\|_2$ is an equivalent norm on $W_0^{1,2}(\Omega)$,

$$[f,h] = \int_\Omega \nabla f \cdot \nabla h$$

is an equivalent inner product. Since $h \mapsto \int_\Omega gh$ is a bounded linear functional on $W_0^{1,2}(\Omega)$ it follows from the Riesz representation theorem (Theorem 5.5) that there is unique $f \in W_0^{1,2}(\Omega)$ such that

$$[f,h] = \int_\Omega gh \quad \text{for all } h \in W_0^{1,2}(\Omega)\,,$$

i.e.

$$\int_\Omega \nabla f \cdot \nabla h = \int_\Omega gh \quad \text{for all } h \in W_0^{1,2}(\Omega)$$

which means f is a solution to (16.17). □

For each $g \in L^2(\Omega)$ denote by $Tg = f \in W_0^{1,2}(\Omega)$ the unique solution to (16.17). The operator

$$T : L^2(\Omega) \to W_0^{1,2}(\Omega)$$

is linear. It is also bounded, because (16.18) $h = f$ and the Poincaré inequality give

$$\int_\Omega |\nabla f|^2 = \int_\Omega gf \le \|g\|_2 \|f\|_2 \le C\|g\|_2 \|\nabla f\|_2\,,$$

so

$$\|Tg\|_{1,2} = \|f\|_{1,2} \approx \|\nabla f\|_2 \le c\|g\|_2\,.$$

We can regard T as an operator from $L^2(\Omega)$ into $L^2(\Omega)$

$$T : L^2(\Omega) \to W_0^{1,2}(\Omega) \subset L^2(\Omega)$$

and the Rellich-Kondrachv theorem implies that T is compact. Observe also that T is one-to-one.

Theorem 16.33. *$T : L^2(\Omega) \to L^2(\Omega)$ is compact and self-adjoint.*

Proof. We are left with the proof that T is self-adjoint. Let $g_1, g_2 \in L^2(\Omega)$ and $f_1 = Tg_1$, $f_2 = Tg_2$. We have to prove that

$$\langle Tg_1, g_2 \rangle = \langle g_1, Tg_2 \rangle\,,$$

i.e.

$$\int_\Omega f_1 g_2 = \int_\Omega g_1 f_2\,. \tag{16.19}$$

Since $f_1 = Tg_1$ is a weak solution to $-\Delta f_1 = g_1$ we have

$$\int_\Omega \nabla f_1 \cdot \nabla h = \int_\Omega g_1 h \quad \text{for } h \in W_0^{1,2}(\Omega) \tag{16.20}$$

and similarly $-\Delta f_2 = g_2$ yields

$$\int_\Omega \nabla f_2 \cdot \nabla h = \int_\Omega g_2 h \quad \text{for } h \in W_0^{1,2}(\Omega). \tag{16.21}$$

The two equalities give

$$\int_\Omega \underbrace{f_1}_{h} g_2 \overset{(16.21)}{=} \int_\Omega \nabla f_2 \cdot \underbrace{\nabla f_1}_{\nabla h} = \int_\Omega \underbrace{\nabla f_2}_{\nabla \tilde{h}} \cdot \nabla f_1 \overset{(16.20)}{=} \int_\Omega g_1 \underbrace{f_2}_{\tilde{h}}$$

which is (16.19). □

Since the operator $T : L^2 \to L^2$ is compact and self-adjoint we can apply the spectral theorem.

Definition. Let $\Omega \subset \mathbb{R}^n$ be open and bounded. We say that λ is an eigenvalue of $-\Delta$ in $W_0^{1,2}(\Omega)$ if there is $0 \neq g \in W_0^{1,2}(\Omega)$ such that

$$-\Delta g = \lambda g \quad \text{weakly in } \Omega.$$

Lemma 16.34. *If $\Omega \subset \mathbb{R}^n$ is bounded, the eigenvalues of $-\Delta$ in $W_0^{1,2}(\Omega)$ are strictly positive.*

Proof. Suppose that $-\Delta g = \lambda g$ weakly in Ω, where $0 \neq g \in W_0^{1,2}(\Omega)$, i.e.

$$\int_\Omega \nabla g \cdot \nabla h = \int_\Omega \lambda g h \quad \text{for } h \in W_0^{1,2}(\Omega).$$

Taking $h = g$ we have

$$0 < \int_\Omega |\nabla g|^2 = \lambda \int_\Omega |g|^2$$

and hence $\lambda > 0$. □

Note that $\mu = 0$ is not an eigenvalue of T since T is one-to-one. It follows immediately from the definition of the operator T that $\mu \neq 0$ is an eigenvalue of T if and only if $\lambda = \mu^{-1}$ is an eigenvalue of $-\Delta$ on $W_0^{1,2}(\Omega)$. Moreover the corresponding eigenspaces are the same.

Hence the lemma yields that the eigenvalues of T are strictly positive. According to the spectral theorem

$$L^2(\Omega) = \underbrace{\ker T}_{\{0\}} \oplus \bigoplus_{i=1}^{N} E_{\mu_i}\,.$$

Since $\dim E_{\mu_i} < \infty$ we conclude that $N = \infty$ and hence

$$L^2(\Omega) = \bigoplus_{i=1}^{\infty} E_{\mu_i}\,, \quad \mu_1 > \mu_2 > \mu_3 > \ldots, \quad \lim_{i\to\infty} \mu_i = 0\,.$$

As we are already observed the eigenspaces E_{μ_i} for T are equal to the eigenspaces $E_{\mu_i^{-1}}$ for $-\Delta$, so we have

Theorem 16.35. *If $\Omega \subset \mathbb{R}^n$ is open and bounded, then:*

(a) *The eigenvalues of the Laplace operator $-\Delta : W_0^{1,2}(\Omega) \to L^2(\Omega)$ are positive and form an increasing sequence*

$$\lambda_1 < \lambda_2 < \lambda_3 < \ldots, \quad \lim_{i\to\infty} \lambda_i = \infty\,.$$

(b) *The dimensions of the corresponding eigenspaces are finite, $\lim E_{\lambda_i} < \infty$, $i = 1, 2, 3 \ldots$*

(c)

$$L^2(\Omega) = \bigoplus_{i=1}^{\infty} E_{\lambda_i}$$

and hence we may choose an orthonormal basis of $L^2(\Omega)$ consisting of eigenfunctions of $-\Delta$.

Let us state an important deep regularity result.

Theorem 16.36. *If $\Omega \subset \mathbb{R}^n$ is open, $g \in C^\infty(\Omega \times \mathbb{R})$ and $f \in W^{1,2}_{\rm loc}(\Omega)$ is a weak solution to*

$$-\Delta f = g(x, f) \quad \textit{in } \Omega, \tag{16.22}$$

i.e.

$$\int_\Omega \nabla f \cdot \nabla \varphi = \int_\Omega f(x, f)\varphi \quad \textit{for } \varphi \in C_0^\infty(\Omega)\,,$$

then $f \in C^\infty(\Omega)$.

We will not prove this result.

If f is an eigenfunction of $-\Delta$, then $-\Delta f = \lambda f$ in Ω, so f is a solution to (16.22) with $g(x,\xi) = \lambda\xi$ and hence $f \in C^\infty(\Omega)$. That means eigenfunctions of $-\Delta$ in $W_0^{1,2}(\Omega)$ are smooth functions in the classical sense.

Example. If $n = 1$ and $\Omega = (0,1)$, then $-\Delta = -d^2/dx^2$ and it is easy to see that the functions

$$w_n(x) = \sqrt{2}\sin(n\pi x), \quad n = 1,2,3,\ldots$$

are eigenfunctions of $-\Delta$ and the corresponding eigenvalues are $(n\pi)^2$. The functions w_n form an orthonormal basis in $L^2(0,1)$. The corresponding expansion in this basis is the (sinusoidal) Fourier series.

Theorem 16.35 generalizes ro the case of any compact Riemannian manifold M: the eigenfunctions of the Laplace-Beltrami operator are smooth and they form an orthonormal basis in $L^2(M)$.[56]

17. Banach Algebras

Definition. A *complex algebra* is a complex linear space with a multiplication such that

(a) $x(yz) = x(yz)$;
(b) $x(y+z) = xy + xz$, $(x+y)z = xz + yz$;
(c) $\alpha(xy) = (\alpha x)y = x(\alpha y)$

for all $x, y, z \in A$ and $\alpha \in \mathbb{C}$.

If in addition A is a Banach space,

$$\|xy\| \leq \|x\|\|y\| \quad \text{for all } x, y \in A \tag{17.1}$$

and if there is a *unit element* $e \in A$ such that $\|e\| = 1$ and

$$ex = xe = x \quad \text{for all } x \in A,$$

then A is called a *Banach algebra.*

The unit element is uniquely determined. Indeed, if e, e' are unit elements, then $e = ee' = e'$. Moreover the multiplication is continuous, i.e. if $x_n \to x$, $y_n \to y$, then $x_n y_n \to xy$. It follows from $x_n y_n - xy = (x_n - x)y_n + x(y_n - y)$ and (17.1).

A complex algebra is *commutative* if

$$xy = yx \quad \text{for all } x, y \in A,$$

but there are many interesting examples of noncommutative algebras.

[56]For a proof see F.W. Warner, *Foundations of differentiable manifolds and Lie groups*, Chapter 6.

Examples. 1. If X is a compact metric space, $C(X)$, the space of complex valued continuous functions with the supremum norm is a Banach algebra.

2. If $U \subset \mathbb{C}$ is the unit disc, then the space of functions that are continuous on $\overline{U}$ and holomorphic in U is a Banach algebra. It is called the *disc algebra.*

3. If X is a Banach space, then the space of bounded operators $B(X)$ is a Banach algebra. If $\dim X > 1$, then $B(X)$ is noncommutative.

4. $L^1(\mathbb{R}^n)$ with the convolution

$$f * g(x) = \int_{\mathbb{R}^n} f(x-y)g(y)\, dy$$

is a commutative Banach algebra except for the fact that there is no unit element.

In the definition of the Banach algebra we require that $\|xy\| \leq \|x\|\|y\|$, however we may think of a continuous multiplication that does not satisfy this inequality, but it is not a problem, because we can always find an equivalent norm such that the inequality is granted.

Theorem 17.1. *Assume that A is a Banach space and a complex algebra with unit element $e \neq 0$ such that the multiplication is left and right continuous. Then there is an equivalent norm with respect to which A is a Banach algebra.*

Proof. For each $x \in A$ we define $M_x \in B(A)$ by $M_x(z) = xz$. It is a bounded operator since the multiplication is right continuous. The collection of all such operators

$$\tilde{A} = \{M_x : x \in A\} \subset B(A)$$

is a complex algebra with the unit element $M_e = I$. Moreover $\|M_e\| = \|I\| = 1$, $\|M_x M_y\| \leq \|M_x\|\|M_y\|$. It is a Banach algebra, because $\tilde{A}$ is a closed subspace of $B(X)$. Indeed, if $M_{x_i} \to T \in B(A)$ in norm of $B(A)$, then $M_{x_i}(e) \to T(e)$ and $M_{x_i}(z) \to T(z)$ for all $z \in A$ as $i \to \infty$ Hence

$$T(z) \leftarrow M_{x_i}(z) = M_{x_i}(e)z \to T(e)z \quad \text{as } i \to \infty$$

by left-continuity of the multiplication, so $T(z) = T(e)z$, $T = M_{T(e)} \in \tilde{A}$.

The mapping $\Phi : A \to \tilde{A}$, $\Phi(x) = M_x$ is an isomorphism of complex algebras, i.e. it is an isomorphism of linear spaces and $\Phi(e) = M_e$, $\Phi(xy) = \Phi(x)\Phi(y)$. Moreover, the inverse mapping is bounded, because for $x \in A$

$$\|x\| = \|M_x e\| \leq \|M_x\|\|e\|\,.$$

Thus by the open mapping theorem, Φ is an isomorphism of Banach spaces. Hence $\|x\|' = \|M_x\|$ is an equivalent norm on A which turns it into a Banach algebra. $\square$

Observe that a closed subalgebra of $B(X)$ that contains I is also a Banach algebra and as a corollary from the above proof we get

Corollary 17.2. *Every Banach algebra A is isomorphic to a closed subalgebra of $B(A)$.*

Definition. A *nonzero* linear functional $\phi : A \to \mathbb{C}$ on a complex algebra is called a *complex homomorphism* if

$$\phi(xy) = \phi(x)\phi(y) \quad \text{for all } x, y \in A.$$

An element $x \in A$ is called *invertible* if there is another element $x^{-1} \in A$ such that

$$x^{-1}x = xx^{-1} = e.$$

It is easy to see that the inverse element x^{-1}, if exists, is uniquely defined.

Proposition 17.3. *If ϕ is a complex homomorphism on a complex algebra with unit e, then $\phi(e) = 1$ and $\phi(x) \neq 0$ for every invertible $x \in A$.*

Proof. Since ϕ is nonzero, there is $y \in A$ such that $\phi(y) \neq 0$ and hence $\phi(y) = \phi(ey) = \phi(e)\phi(y)$, so $\phi(e) = 1$. Moreover, if $x \in A$ is invertible, then $\phi(x^{-1})\phi(x) = \phi(x^{-1}x) = \phi(e) = 1$, so $\phi(x) \neq 0$. □

Theorem 17.4. *Suppose A is a Banach algebra and $x \in A$, $\|x\| < 1$. Then*

(a) *$e - x$ is invertible.*
(b)
$$\|(e - x)^{-1} - e - x\| \leq \frac{\|x\|^2}{1 - \|x\|}.$$
(c) $|\phi(x)| < 1$ *for every complex homomorphism ϕ on A.*

Proof. Proof of the part (a) is very similar to that of Theorem 2.7. The series

$$\sum_{n=0}^{\infty} x^n = e + x + x^2 + x^3 + \ldots$$

converges absolutely, because $\|x^n\| \leq \|x\|^n$ and $\|x\| < 1$. Hence it converges to an element in A. It is easy to see that this element is an inverse of $e - x$. Since

$$\|(e - x)^{-1} - e - x\| = \|x^2 + x^3 + \ldots\| \leq \frac{\|x\|^2}{1 - \|x\|}$$

the part (b) follows. Finally if $\lambda \in \mathbb{C}$, $|\lambda| \geq 1$, then $e - \lambda^{-1}x$ is invertible by (a) and thus

$$1 - \lambda^{-1}\phi(x) = \phi(e - \lambda^{-1}x) \neq 0$$

by Proposition 17.3. Hence $\phi(x) \neq \lambda$ for any such λ, so $|\phi(x)| < 1$. □

Corollary 17.5. *Every complex homomorphism on a Banach algebra is continuous.*

Proof. It is a direct consequence of part (c) of the above theorem. □

It turns out that the property of complex homomorphisms described in Proposition 17.3 characterizes complex homomorphisms in the case of Banach algebras.

Theorem 17.6 (Gleason-Kahane-Żelazko). *A linear functional*[57] ϕ *on a Banach algebra* A *is a complex homomorphism if and only if* $\phi(e)=1$ *and* $\phi(x)\neq 0$ *for every noninvertible* $x\in A$.

Proof. The implication from left to right is contained in Proposition 17.3, so we are left with the proof of the implication from right to left.

Let $\mathcal{N}=\ker\phi$. If $x,y\in A$, then $x-\phi(x)e, y-\phi(y)e\in\mathcal{N}$ and hence

$$x=a+\phi(x)e,\quad y=b+\phi(y)e \tag{17.2}$$

for some $a,b\in\mathcal{N}$. Applying ϕ to the product of the two equations we get

$$\phi(xy)=\phi(ab)+\phi(x)\phi(y)\,. \tag{17.3}$$

Thus it remains to prove the implications

$$\text{If } a,b\in\mathcal{N}, \text{ then } ab\in\mathcal{N}. \tag{17.4}$$

Consider a seemingly weaker statement

$$\text{If } a\in\mathcal{N}, \text{ then } a^2\in\mathcal{N}. \tag{17.5}$$

Lemma 17.7. *Property (17.5) implies property (17.4).*

Proof. Suppose (17.5) is true. then (17.3) for $x=y$ gives

$$\phi(x^2)=\phi(x)^2\quad\text{for all } x\in A.$$

Replacement of x by $x+y$ gives

$$\phi(xy+yx)=2\phi(x)\phi(y)\quad\text{for } x,y\in A.$$

Thus

$$\text{If } x\in\mathcal{N},\ y\in A, \text{ then } xy+yx\in\mathcal{N}. \tag{17.6}$$

The next identity is easy to check

$$(xy-yx)^2+(xy+yx)^2=2[x(yxy)+(yxy)x]\,.$$

if $x\in\mathcal{N}$, then the right hand side belongs to $\mathcal{N}$ by (17.6) and since $(xy-yx)^2\in\mathcal{N}$ by (17.6) and (17.5), we conclude that $(xy+yx)^2\in\mathcal{N}$ and hence $xy+yx\in\mathcal{N}$ by another application of (17.5). Thus also $(xy+yx)+(xy-yx)=2xy\in\mathcal{N}$. We proved that if $x\in\mathcal{N}$ and $y\in A$, then $xy\in\mathcal{N}$ which implies (17.4). □

[57] We do not assume continuity of ϕ.

Thus we are left with the proof of the property (17.5). The above arguments were purely algebraic and would work in any complex algebra with unit, however, the proof of (17.5) is analytic.

By the assumptions, there are no invertible elements in $\mathcal{N}$, so $\|e-x\| \geq 1$ for every $x \in \mathcal{N}$ by Theorem 17.4. Hence

$$\|\lambda e - x\| \geq |\lambda| = \phi(\lambda e - x) \quad \text{for } x \in \mathcal{N}.$$

Since every element in A is of the form $\lambda e - x$, $x \in \mathcal{N}$ (see (17.2) we conclude that ϕ is a continuous linear functional of norm 1.

Lemma 17.8. *Suppose f is an entire function on $\mathbb{C}$ such that $f(0) = 1$, $f'(0) = 0$ and*

$$0 < |f(\lambda)| \leq e^{|\lambda|} \quad \textit{for } \lambda \in \mathbb{C}.$$

Then $f(\lambda) = 1$ for all $\lambda \in \mathbb{C}$.

Proof. Since f has no zeroes, there is another entire function g such that

$$f = e^g\,.$$

Clearly $g(0) = g'(0) = 0$ and

$$\operatorname{re} g(\lambda) \leq |\lambda|. \tag{17.7}$$

If $|\lambda| \leq r$, then $\operatorname{re} g(\lambda) \leq r$ and hence

$$|\operatorname{re} g(\lambda)| \leq |2r - \operatorname{re} g(\lambda)|\,,$$

so

$$|g(\lambda)|^2 = (\operatorname{re} g(\lambda))^2 + (\operatorname{im} g(\lambda))^2 \leq (2r - \operatorname{re} g(\lambda))^2 + (\operatorname{im} g(\lambda))^2 = |2r - g(\lambda)|^2\,.$$

Thus

$$|g(\lambda)| \leq |2r - g(\lambda)| \quad \text{for } |\lambda| \leq r. \tag{17.8}$$

Consider the function

$$h_r(\lambda) = \frac{r^2 g(\lambda)}{\lambda^2(2r - g(\lambda))}\,. \tag{17.9}$$

This fuinction is holomorphic in the disc $\{\lambda : |\lambda| < 2r\}$ by (17.7) and $|h_r(\lambda)| \leq 1$ if $|\lambda| = r$ by (17.8). Hence the maximum principle gives

$$|h_r(\lambda)| \leq 1 \quad \text{for } |\lambda| \leq r. \tag{17.10}$$

if we fix λ and let $r \to \infty$ (17.9) and (17.10) gives $g(\lambda) = 0$. Hence $f(\lambda) = e^0 = 1$. □

Now we are ready to complete the proof of (17.5). Let $a \in \mathcal{N}$. We can assume that $\|a\| = 1$. Since ϕ has norm 1, $|\phi(a^n)| \leq \|a^n\| \leq \|a\|^n = 1$ and hence the function

$$f(\lambda) = \sum_{n=0}^{\infty} \frac{\phi(a^n)}{n!}\,\lambda^n, \quad \lambda \in \mathbb{C}$$

is entire. Moreover $f(0) = \phi(e) = 1$, $f'(0) = \phi(a) = 0$. If we prove that $f(\lambda) \neq 0$ for every $\lambda \in \mathbb{C}$, then the lemma will imply that $f''(0) = 0$, i.e. $\phi(a^2) = 0$ which is (17.5).

The series

$$E(\lambda) = \sum_{n=0}^{\infty} \frac{a^n}{n!}\,\lambda^n$$

converges in norms for every $\lambda \in \mathbb{C}$ and hence continuity of ϕ shows that

$$f(\lambda) = \phi(E(\lambda)) \quad \text{for } \lambda \in \mathbb{C}.$$

The functional equation $E(\lambda + \mu) = E(\lambda)E(\mu)$ can be proves in the same way as in the scalar case. In particular $E(-\lambda)E(\lambda) = E(0) = e$, so $E(\lambda)$ is invertible for every λ and hence $f(\lambda) = \phi(E(\lambda)) \neq 0$, because by the assumption ϕ is nonzero on invertible elements. $\square$

Definition. Let A be a Banach algebra. By $G(A)$ we denote the set os all invertible elements in A. It is easy to see that $G(A)$ is a group with respect to the algebra multiplication.

For $x \in A$ the *spectrum of x* is the set

$$\sigma(x) = \{\lambda \in \mathbb{C} :\ \lambda e - x \text{ is not invertible}\}$$

and $\mathbb{C} \setminus \sigma(x)$ is called the *resolvent set of x*.

As we shall see $\sigma(x)$ is always nonempty and we define the *spectral radius of x* as

$$\rho(x) = \sup\{|\lambda| :\ \lambda \in \sigma(x)\}.$$

It easily follows from Theorem 17.4(a) that $\sigma(x)$ is bounded, i.e. $|\lambda$ $leq\|x\|$ for $\lambda \in \sigma(x)$, so $\rho(x) \leq \|x\|$.

Lemma 17.9. *If $x \in G(A)$ and $h \in A$, $\|h\| \leq \frac{1}{2}\|x^{-1}\|^{-1}$, then $x + h \in G(A)$ and*

$$\|(x+h)^{-1} - x^{-1} + x^{-1}hx^{-1}\| \leq 2\|x^{-1}\|^3\|h\|^2\,. \tag{17.11}$$

Proof. $x + h = x(e + x^{-1}h)$, $\|x^{-1}h\| < 1/2$, so invertibility of $x + h$ follows from Theorem 17.4(a), so $x + h \in G(A)$ and $(x + h)^{-1} = (e + x^{-1}h)^{-1}x^{-1}$. Hence

$$(x+h)^{-1} - x^{-1} - x^{-1}hx^{-1} = [(e + x^{-1}h)^{-1} - e - x^{-1}h]x^{-1}$$

and the inequality (17.11) follows from Theorem 17.4(b). $\square$

Theorem 17.10. *If A is a Banach algebra, then $G(A)$ is an open set and the mapping $x \mapsto x^{-1}$ is a homeomorphism of $G(A)$ onto $G(A)$.*

Proof. The lemma implies that $G(A)$ is open. It also implies that

$$\|(x+h)^{-1} - x^{-1}\| \leq 2\|x^{-1}\|^3\|h\|^2 + \|x^{-1}\|^2\|h\|$$

which yields continuity of $x \mapsto x^{-1}$. Since $x \mapsto x^{-1}$ maps $G(A)$ onto $G(A)$ and it is its inverse, it follows that $x \mapsto x^{-1}$ is a homeomorphism of $G(A)$ onto itself. $\square$

The following lemma is often useful.

Lemma 17.11. *let $\{c_n\}_{n=1}^{\infty}$ be a sequence of nonnegative numbers such that*

$$c_{m+n} \le c_m c_n \tag{17.12}$$

for all positive integers m, n. Then the limit

$$\lim_{n\to\infty} c_n^{1/n}$$

exists and equals $\inf_{n\ge 1} c_n^{1/n}$.

Proof. If $c_m = 0$ for some m, then $c_n = 0$ for $n \ge m$ and the lemma follows, so we may assume that $c_n > 0$ for all n. We put $c_0 = 1$. Fix m. Any integer n can be represented as

$$n = q(n)m + r(n), \quad 0 \le r(n) < m.$$

Thus (17.12) implies that

$$c_n^{1/n} \le c_m^{q(n)/n} c_{r(n)}^{1/n}.$$

Clearly $q(n)/n \to 1/m$ as $n \to \infty$ and since $c_{r(n)}$ attains only a finite number of positive values, $c_{r(n)}^{1/n} \to 1$ as $n \to \infty$. Hence

$$\limsup_{n\to\infty} c_n^{1/n} \le c_m^{1/m} \quad \text{for any integer } m.$$

Thus

$$\limsup_{n\to\infty} c_n^{1/n} \le \inf_m c_m^{1/m} \le \liminf_{n\to\infty} c_n^{1/n}$$

and the lemma follows. $\square$

Let x be an element of a Banach algebra and $c_n = \|x^n\|$, then $c_{n+m} \le c_n c_m$ and hence the lemma shows that

$$\lim_{n\to\infty} \|x^n\|^{1/n} = \inf_{n\ge 1} \|x^n\|^{1/n}. \tag{17.13}$$

Theorem 17.12. *If A is a Banach algebra and $x \in A$, then*

(a) *The spectrum $\sigma(x)$ is compact and nonempty.*
(b) *The spectral radius satisfies*

$$\rho(x) = \lim_{n\to\infty} \|x^n\|^{1/n} = \inf_{n\ge 1} \|x^n\|^{1/n}. \tag{17.14}$$

Proof. Although the second equality in (17.14) follows from (17.13) we will not use this fact and we will conclude (17.14) from different arguments.

As we know $\sigma(x)$ is a bounded set. To prove that $\sigma(x)$ is compact define $g:\mathbb{C}\to A$ by $g(\lambda)=\lambda e - x$. Clearly g is continuous and

$$\Omega := \mathbb{C}\setminus\sigma(x) = g^{-1}(G(A))\,.$$

The set Ω is open as a preimage of an open set and hence $\sigma(x)$ is compact as bounded and closed. Now we define $f:\Omega\to G(A)$ by

$$f(\lambda) = (\lambda e - x)^{-1}\,.$$

We will prove that $f(\lambda)$ is a homomorphic A-valued function. If we replace x and h by $\lambda e - x$ and $(\mu-\lambda)e$ in Lemma 17.9, then for μ sufficiently close to λ inequality (17.11) reads as

$$\Vert f(\mu) - f(\lambda) - (\mu-\lambda)f^2(\lambda)\Vert \leq 2\Vert f(\lambda)\Vert^3|\mu-\lambda|^2$$

and hence

$$\lim_{\mu\to\lambda}\frac{f(\mu)-f(\lambda)}{\mu-\lambda} = -f^2(\lambda)\,,$$

so f is holomorphic.

If $|\lambda| > \Vert x\Vert$, then it is easy to see that

$$f(\lambda) = (\lambda e - x)^{-1} = \sum_{n=0}^\infty \lambda^{-n-1}x^n\,. \tag{17.15}$$

The series converges uniformly on any circle Γ_r centered at 0 of radius $r>\Vert x\Vert$ and hence Theorem 12.3 allows us to integrate it term by term

$$\frac{1}{2\pi i}\int_{\Gamma_r} f(\lambda)\,d\lambda = \sum_{n=0}^\infty x^n\,\frac{1}{2\pi i}\int_{\Gamma_r}\lambda^{-n-1}\,d\lambda = e. \tag{17.16}$$

Suppose that $\sigma(x)=\emptyset$, i.e. $\Omega=\mathbb{C}$. Then f is an entire function and the Cauchy formula (12.6) implies that the left hand side of (17.16) equals zero which is a contradiction. Thus $\sigma(x)\neq\emptyset$ which completes the proof of the first part of the theorem.

The same term by term integration as in (17.16) gives

$$\frac{1}{2\pi i}\int_{\Gamma_r}\lambda^n f(\lambda) = x^n, \quad r>\Vert x\Vert,\ n=0,1,2,\ldots \tag{17.17}$$

Since $\lambda^n f(\lambda)$ is holomorphic for all $\lambda>\rho(x)$, the Cauchy theorem (12.6) implies that (17.17) holds for all $r>\rho(x)$ and thus

$$\Vert x^n\Vert \leq r^{n+1}\sup_{\theta\in[0,1]}\Vert f(re^{i\theta})\Vert, \quad r>\rho(x).$$

Hence

$$\limsup_{n\to\infty}\Vert x^n\Vert^{1/n}\leq r \quad \text{for all } r>\rho(x),$$

i.e.

$$\limsup_{n\to\infty}\Vert x^n\Vert^{1/n}\leq\rho(x)\,. \tag{17.18}$$

Lemma 17.13. *If ab and ba are invertible elements in a complex algebra, then a and b are invertible.*

Proof. We will prove that a is invertible as the proof for b is the same. Since $ba(ba)^{-1}b = bab(ab)^{-1}$ we have $(ba)^{-1}b = b(ab)^{-1}$ and thus

$$a[b(ab)^{-1}] = e, \quad [b(ab)^{-1}]a = (ba)^{-1}ba = e$$

so $a^{-1} = b(ab)^{-1}$. □

Lemma 17.14. *If $\lambda \in \sigma(x)$, then $\lambda^n \in \sigma(x^n)$.*

Proof. We have

$$\lambda^n e - x^n = (\lambda e - x)(\lambda^{n-1}e + \ldots + x^{n-1}) = (\lambda^{n-1}e + \ldots + x^{n-1})(\lambda e - x)\,.$$

If $\lambda^n \notin \sigma(x^n)$, then according to the previous lemma $\lambda e - x$ is invertible, so $\lambda \notin \sigma(x)$. □

If $\lambda \in \sigma(x)$, then $\lambda^n \in \sigma(x^n)$ and hence $|\lambda|^n = |\lambda^n| \leq \|x^n\|$, so $\rho(x) \leq \|x^n\|^{1/n}$ for any $n = 1, 2, \ldots$, i.e.

$$\rho(x) \leq \inf_{n\geq 1} \|x^n\|^{1/n}$$

which together with (17.18) implies (17.14). □

Theorem 17.15 (Gelfand-Mazur)**.** *If A is a Banach algebra in which every nonzero element is invertible, then A is isometrically isomorphic to $\mathbb{C}$.*

Proof. let $x \in A$ and $\lambda \in \sigma(x)$ ($\sigma(x)$ is nonempty). Hence $\lambda e - x = 0$, so $x = \lambda e$. Thus $\sigma(x)$ consists of exactly one point. Denote this point by $\lambda(x)$. The mapping $x \mapsto \lambda(x)$ is an isometric isomorphism of A onto $\mathbb{C}$. □

18. Commutative Banach algebras

Definition. A subset J of a commutative complex algebra is called an *ideal* if J is a linear subspace of A and $xy \in J$ whenever $x \in A$ and $y \in J$. J is a *proper* ideal if $J \neq A$ and it is a *maximal* ideal if it is proper and is not contained in any larger ideal.

The ideals are interesting, because they are kernels of homomorphisms of commutative Banach algebras and closed ideals are particularly interesting. Indeed, if : $\phi : A \to B$ is a homomorphism of commutative Banach algebras (i.e. $\phi(xy) = \phi(x)\phi(y)$)), then $\ker \phi$ is an ideal. Moreover the ideal is closed if ϕ is continuous.

Conversely every closed ideal J in a Banach algebra A it is a kernel of a continuous homomorphism. Indeed, let $\pi : A \to A/J$ be the quotient mapping. A/J is a Banach space, but as we shall see it is also a Banach algebra and π is a homomorphism with the kernel J.

The multiplication in A/J is defined by

$$[x][y] = [xy] \quad (\text{i.e. } \pi(x)\pi(y) = \pi(xy)).$$

This multiplication is well defined because if $[x] = [x']$, $[y] = [y']$, then $x' - x, y' - y \in J$,

$$x'y' - xy = (x' - x)y' + x(y' - y) \in J$$

and hence $[x'y'] = [xy]$. Thus A/J is a complex algebra with unit $[e] \neq 0$ and π is a homomorphism with $\ker \pi = J$. π is continuous since $\|\pi(x)\| \leq \|x\|$ by the definition of the norm in A/J. It remains to prove that A/J is a Banach algebra, i.e.

$$\|[e]\| = 1, \quad \|[x][y]\| \leq \|[x]\|\|[y]\|.$$

Let $x, y \in A$. Then by the definition of the quotient norm for every $\varepsilon > 0$ there are $\tilde{x}, \tilde{y} \in J$ such that

$$\|x + \tilde{x}\| \leq \|[x]\| + \varepsilon, \quad \|y + \tilde{y}\| \leq \|[y]\| + \varepsilon\,.$$

Since

$$xy = (x + \tilde{x})(y + \tilde{y}) - \underbrace{(x\tilde{y} + \tilde{x}y + \tilde{x}\tilde{y})}_{\in J}$$

we have

$$\|[xy]\| \leq \|(x + \tilde{x})(y + \tilde{y})\| \leq \|x + \tilde{x}\|\|y + \tilde{y}\| \leq (\|[x]\| + \varepsilon)(\|[y]\| + \varepsilon)\,,$$

and thus

$$\|[xy]\| \leq \|[x]\|\|[y]\|\,. \tag{18.1}$$

Clearly $[e] \neq 0$ and hence (18.1) gives $\|[e]\| \leq \|[e]\|^2$, $\|[e]\| \geq 1$, but on the other hand $\|[e]\| \leq \|e\| = 1$, so $\|[e]\| = 1$.

Proposition 18.1.

(a) *No proper ideal of A contains an invertible element.*
(b) *If J is an ideal in a commutative Banach algebra, then its closure $\overline{J}$ is also an ideal.*

The proof is left to the reader as a simple exercise.

Theorem 18.2.

(a) *If A is a commutative complex algebra with unit, then every proper ideal is contained in a maximal ideal.*
(b) *If A is a commutative Banach algebra, then every maximal ideal is closed.*

Proof. (a) Let $J \subset A$ be a proper ideal. Consider the family of all proper ideals that contain J. The family is partially ordered by inclusion. According to the Hausdorff maximality theorem there is a maximal totally ordered subfamily. The union of this subfamily is also an ideal. It is proper since the unit element does not belong to any of the ideals in the family. Hence the union is a maximal ideal.

(b) Let M be a maximal ideal. Then $\overline{M}$ is an ideal. It is proper since M has no invertible elements and invertible elements form an open set. Hence $M = \overline{M}$ by maximality of M. □

Theorem 18.3. *Let A be a commutative Banach algebra, and let Δ be the set of all complex homomorphisms of A.*

(a) *If $\phi \in \Delta$, the kernel of ϕ is a maximal ideal of A.*
(b) *Every maximal ideal of A is the kernel of some $\phi \in \Delta$.*
(c) *An element $x \in A$ is invertible if and only if $\phi(x) \neq 0$ for all $\phi \in \Delta$.*
(d) *An element $x \in A$ is invertible if and only if x does not belong to any proper ideal of A.*
(e) *$\lambda \in \sigma(x)$ if and only if $\phi(x) = \lambda$ for some $\phi \in \Delta$.*

Proof. (a) If $\phi \in \Delta$, then $\ker \phi$ is a proper ideal. It is maximal because it has codimension 1.

(b) Let M be a maximal ideal. Hence M is closed, A/M is a Banach algebra and $\pi : A \to A/M$ is a homomorphism with $\ker \pi = M$. It remains to prove that A/M is isomorphic to $\mathbb{C}$. Any nonzero element in A/M is of the form $\pi(x)$ for some $x \in A \setminus M$. Let

$$J = \{ax + y : a \in A, y \in M\}.$$

Clearly J is an ideal that contains M as a proper subspace. Hence $J = A$ by the maximality of M. In particular $ax + y = e$ for some $a \in A$, $y \in M$. Thus $\pi(e) = \pi(ax + y) = \pi(a)\pi(x)$, so $\pi(x) \in A/M$ is invertible and according to the Gelfand-Mazur theorem A/M is isometrically isomorphic to $\mathbb{C}$.

(c) If x is invertible and $\phi \in \Delta$, then $1 = \phi(e) = \phi(x^{-1})\phi(x)$, so $\phi(x) \neq 0$. If x is not invertible, then $\{ax : a \in A\}$ is an ideal that does not contain e, and hence it is contained in a maximal ideal M. Now according to (b) there is $\phi \in \Delta$ with $\ker \phi = M$, so $\phi(x) = 0$.

(d) If x is invertible, then x is not contained in any proper ideal by Proposition 18.1(a). The converse implication was proved in (c).

(e) It follows from (c) applied to $\lambda e - x$ in place of x. □

Theorem 18.4. *Suppose $f_1, f_2, \ldots, f_n \in A(U)$ (disc algebra) are such that*

$$|f_1|^2 + \ldots + |f_n|^2 > 0 \quad \text{on } \overline{U}, \tag{18.2}$$

i.e. the functions $f_1, \ldots, f_n$ do not have a common zero in $\overline{U}$. Then there are functions $\phi_1, \ldots, \phi_n \in A(U)$ such that

$$f_1(z)\phi_1(z) + \ldots + f_n(z)\phi_n(z) = 1 \quad \text{for all } x \in \overline{U}.$$

Proof. $A(U)$ is a commutative Banach algebra with pointwise multiplication and the supremum norm.

$$J = \{f_1\phi_1 + \ldots + f_n\phi_n : \phi_1, \ldots, \phi_n \in A(U)\}$$

is an ideal in $A(U)$ and it suffices to prove that $J = A(U)$. If $J \neq A(U)$, then J is contained in a maximal ideal and hence there is a complex homomorphism $\phi \in \Delta$ such that ϕ vanishes on J. Let $g(z) = z \in A(U)$, $\|g\| = 1$. Since complex homomorphisms have norm 1 (Theorem 17.4(c)), $\phi(g) = w$ for some $\|w\| \leq 1$. Now from a multiplicative property of ϕ it easily follows that for every complex polynomial P, $\phi(P) = P(w)$. The density of polynomials in $A(U)$[58] and continuity of ϕ imply that $\phi(f) = f(w)$ for every $f \in A(U)$. Since f vanishes on J, $0 = \phi(f_i) = f_i(w)$ for $i = 1, 2, \ldots, n$ which contradicts (18.2). $\square$

Theorem 18.5 (Wiener's lemma)**.** *Let $f : \mathbb{R}^n \to \mathbb{C}$ be the sum of absolutely convergent multidimensional Fourier series*[59]

$$f(x) = \sum_{m\in\mathbb{Z}^n} a_m e^{2\pi i m\cdot x}, \qquad \sum_{m\in\mathbb{Z}} |a_m| < \infty\,. \tag{18.3}$$

If $f(x) \neq 0$ for all $x \in \mathbb{R}^n$, then

$$\frac{1}{f(x)} = \sum_{m\in\mathbb{Z}^n} b_m e^{2\pi i m\cdot x}, \qquad \sum_{m\in\mathbb{Z}^n} |b_m| < \infty\,.$$

Proof. One can easily check that the functions of the form (18.3) form a Banach algebra A with respect to the pointwise multiplication and the supremum norm. For each $x \in \mathbb{R}^n$,

$$A \ni f \mapsto f(x) \tag{18.4}$$

is a complex homomorphism. If we prove that all complex homomorphisms are of that form, it will follow that the function f satisfies $\phi(f) \neq 0$, for all $\phi \in \Delta$ and hence f is invertible by Theorem 18.3(c) which is what we want to prove.

For $k = 1, 2, \ldots, n$ put $g_k(x) = e^{2\pi i x_k}$, where x_k is kth coordinate of $x \in \mathbb{R}^n$. Clearly $g_k, 1/g_k \in A$ and both functions have norm 1. If $\phi \in \Delta$, then

$$|\phi(g_k)| \leq 1 \quad \text{and} \quad \Big|\frac{1}{\phi(g_k)}\Big| = \Big|\phi\Big(\frac{1}{g_k}\Big)\Big| \leq 1\,.$$

Hence there is $y \in \mathbb{R}^n$ such that

$$\phi(g_k) = e^{2\pi i y_k} = g_k(y), \quad k = 1, 2, \ldots, n. \tag{18.5}$$

Every trigonometric polynomial P is a linear combination of products of the functions g_k and $1/g_k$ and since ϕ is a complex homomorphism (18.5) implies that $\phi(P) = P(y)$. Since ϕ is continuous and trigonometric polynomials are

[58] Prove it.

[59] $m \cdot x = m_1x_1 + \ldots + m_nx_n$.

dense in A, we get $\phi(f) = f(y)$ for all $f \in A$ and hence ϕ is of the form (18.4). $\square$

Definition. Let Δ be the set of all complex homomorphisms of a commutative Banach algebra A. The formula

$$\hat{x}(\phi) = \phi(x) \quad \text{for } \phi \in \Delta$$

assigns to each $x \in A$ a function $\hat{x} : \Delta \to \mathbb{C}$ called the *Gelfand transform* of x.

Let $\hat{A}$ be the space of all functions $\hat{x}$ on Δ. The *Gelfand topology* on Δ is the weakest topology for which all the functions $\hat{x}$ are continuous. Obviously $\hat{A} \subset C(\Delta)$.

Since there is a one-to-one correspondence between Δ and the maximal ideals in A, Δ equipped with the Gelfand topology is called the *maximal ideal space* of A.

Theorem 18.6. *If A is a commutative Banach algebra, then*

(a) *Δ is a compact Hausdorff space;*
(b) *For each $x \in A$, the range of $\hat{x}$ is $\sigma(x)$ and hence*

$$\|\hat{x}\|_\infty = \rho(x) \leq \|x\|.$$

Proof. (a) Let A^* be the dual Banach space and $\overline{B}$ the closed unit ball in A^*. By the Banach-Alaoglu theorem, $\overline{B}$ is compact in the weak-$*$ topology. It is Hausdorff, because the weak-$*$ topology is Hausdorff. Clearly $\Delta \subset \overline{B}$ and the Gelfand topology is the restriction of the weak-$*$ topology to Δ. Thus it remains to show that Δ is a closed subset of $\overline{B}$ in the weak-$*$ topology.[60]

Let ϕ_0 belongs to the closure of Δ in the weak-$*$ topology. We have to prove that

$$\phi_0(xy) = \phi_0(x)\phi_0(y), \quad \phi_0(e) = 1\,.$$

Fix $x, y \in A$ and $\varepsilon > 0$. The set

$$\mathcal{V} = \{x^* \in A^* : |\langle x^*, z_i\rangle - \phi_0(z_i)| < \varepsilon,\ i = 1, 2, 3, 4\}$$

where $z_1 = e$, $z_2 = x$, $z_3 = y$, $z_4 = xy$ is open in the weak-$*$ topology and $\phi_0 \in \mathcal{V}$. It follows from the definition of the closure that there is $\phi \in \Delta \cap \mathcal{V}$, so $|\phi(z_i) - \phi_0(z_i)| < \varepsilon$ for $i = 1, 2, 3, 4$. In particular

$$|1 - \phi_0(e)| = |\phi(e) - \phi_0(e)| < \varepsilon$$

gives $\phi_0(e) = 1$. Moreover

$$\begin{aligned} \phi_0(xy) - \phi_0(x)\phi_0(y) &= (\phi_0(xy) - \phi(xy)) + (\phi(x)\phi(y) - \phi_0(x)\phi_0(y)) \\ &= \quad (\phi_0(xy) - \phi(xy)) + \phi(x)(\phi(y) - \phi_0(y)) + \phi_0(y)(\phi(x) - \phi_0(x)) \end{aligned}$$

[60] The proof of this part is very similar to a corresponding argument in the proof of the Banach-Alaoglu theorem.

gives

$$|\phi_0(xy) - \phi_0(x)\phi_0(y)| \leq (1 + |\phi(x)| + |\phi_0(y)|)\,\varepsilon$$

and hence

$$\phi_0(xy) = \phi_0(x)\phi_0(y)\,.$$

(b) λ is in the range of $\hat{x}$ if $\hat{x}(\phi) = \phi(x) = \lambda$ for some $\phi \in \Delta$ which is equivalent to $\lambda \in \sigma(x)$ by Theorem 18.3(e). □

18.1. C^*-algebras.

Definition. Let A be a complex algebra (not necessarily commutative). By an *involution* on A we mean a map $x \mapsto x^*$ of A onto itself such that

$$(x+y)^* = x^* + y^*, \quad (\alpha x)^* = \overline{\alpha} x^*,$$

$$(xy)^* = y^* x^*, \quad x^{**} = x.$$

If e is a unit element, then one easily verifies that $e^* = e$ and if x is invertible, then $(x^{-1})^* = (x^*)^{-1}$.

A Banach algebra A is called a *C^*-algebra* if it is an algebra with involution that satisfies

$$\|x^* x\| = \|x\|^2 \quad \text{for all } x \in A\,.$$

An element $x \in A$ satisfying $x = x^*$ is called *hermitian* or *self-adjoint*.

Note that $\|x\|^2 = \|x^* x\| \leq \|x\|\|x^*\|$ implies $\|x\| \leq \|x^*\|$. Hence also $\|x^*\| \leq \|x^{**}\| = \|x\|$. Thus

$$\|x^*\| = \|x\|. \tag{18.6}$$

Example. If X is a compact Hausdorff space, then $C(X)$ is a commutative C^*-algebra with the involution $f \mapsto \overline{f}$. The Gelfand-Najmark theorem (Theorem 18.9) states that every commutative C^*-algebra is isometrically isomorphic to $C(X)$ for some X.

Example. Let H be a Hilbert space. For $A \in B(H)$ we define the *adjoint operator* $A^* \in B(H)$ to the the unique operator such that

$$\langle Ax, y\rangle = \langle x, A^* y\rangle \quad \text{for all } x, y \in H.$$

Clearly

$$(A+B)^* = A^* + B^*, \quad (\alpha A)^* = \overline{\alpha} A^*$$

$$(AB)^* = B^* A^*, \quad A^{**} = A\,.$$

Moreover

$$\|A^* A\| = \|A\|^2\,.$$

To prove the last equality observe that

$$\|A^* A\| = \sup_{\substack{\|x\|\leq 1\\ \|y\|\leq 1}} \langle A^* Ax, y\rangle\,.$$

Since

$$|\langle A^*Ax, y\rangle| = |\langle Ax, Ay\rangle| \leq \|A\|^2\|x\|\|y\|$$

we conclude that $\|A^*A\| \leq \|A\|^2$. On the other hand

$$\|Ax\|^2 = \langle Ax, Ax\rangle = \langle A^*Ax, x\rangle \leq \|A^*A\|\|x\|^2$$

implies $\|A\|^2 \leq \|A^*A\|$. Thus $\|A^*A\| = \|A\|^2$.

Therefore $B(H)$ with the adjoint operator as involution is a C^*-algebra. As a direct consequence of (18.6) we have

$$\|A^*\| = \|A\|.$$

Theorem 18.7. *Let A be a C^*-algebra. If $x \in A$ is self-adjoint, then $\sigma(x) \subset \mathbb{R}$ and $\rho(x) = \|x\|$.*

Proof. Since $\|x^2\| = \|x^*x\| = \|x\|^2$ it easily follows that

$$\|x^{2^n}\| = \|x\|^{2^n}, \quad \rho(x) = \lim_{n\to\infty} \|x^{2^n}\|^{2^{-n}} = \|x\|\,.$$

Let $\lambda \in \sigma(x)$. Then for any $t \in \mathbb{R}$, $\lambda + it \in \sigma(x + ite)$. We have

$$\begin{aligned} |\lambda + it|^2 &\leq \|x + ite\|^2 = \|(x + ite)(x + ite)^*\| = \|(x + ite)(x - ite)\| \\ &= \|x^2 + t^2e\| \leq \|x^2\| + t^2\,. \end{aligned}$$

If $\lambda = a + bi$, then the inequality reads as

$$a^2 + b^2 + 2bt \leq \|x^2\|$$

which is possible only if $b = 0$, so $\lambda \in \mathbb{R}$. □

Lemma 18.8. *Let A be a commutative C^*-algebra. If $\phi \in \Delta$ is a complex homomorphism, then*

$$\phi(x^*) = \overline{\phi(x)} \quad \textit{for all } x \in A.$$

Proof. If $x = x^*$, then $\phi(x^*) = \phi(x) \in \sigma(x) \subset \mathbb{R}$ by Theorem 18.3(e) and Theorem 18.7. If x is arbitrary, then

$$x = u + iv, \quad \text{where} \quad u = \frac{1}{2}(x + x^*),\ v = \frac{1}{2i}(x - x^*)\,.$$

Since $u = u^*$, $v = v^*$ we have

$$\phi(x^*) = \phi(u) - i\phi(v) = \overline{\phi(u) + i\phi(v)} = \overline{\phi(x)}\,.$$

The proof is complete. □

Theorem 18.9 (Gelfand-Najmark)**.** *Suppose A is a commutative C^*-algebra, with maximal ideal space Δ. Then the Gelfand transform is an isometric isomorphism of A onto $C(\Delta)$ such that*

$$(x^*)\hat{} = \overline{\hat{x}} \quad \textit{for all } x \in A. \tag{18.7}$$

In particular x is self-adjoint if and only if $\hat{x}$ is a real valued function, i.e. $\sigma(x) \subset \mathbb{R}$.

Proof. Equality (18.7) means that
$$\phi(x^*) = \overline{\phi(x)} \quad \text{for all } \phi \in \Delta.$$
and hence it follows from Lemma 18.8.

Let $\hat{A} \subset C(\Delta)$ be the space of all functions on Δ of the form $\hat{x}$. $\hat{A}$ is self-adjoint, i.e. it is closed under the complex conjugation, it separates points and it does not vanish at any point $\phi \in \Delta$. Hence $\hat{A}$ is dense in $C(\Delta)$ by the Stone-Weierstrass theorem.

Let $x \in A$. Then $y = x^*x$ is self-adjoint. For $\phi \in \Delta$ we have
$$\hat{y}(\phi) = \phi(x^*x) = \phi(x^*)\phi(x) = |\phi(x)|^2 = |\hat{x}(\phi)|^2$$
and hence
$$\|\hat{y}\|_\infty = \|\hat{x}\|_\infty^2 .$$
Since y is self-adjoint, Theorem 18.6(b) and Theorem 18.7 yield
$$\|\hat{x}\|_\infty^2 = \|\hat{y}\|_\infty = \rho(y) = \|y\| = \|x^*x\| = \|x\|^2 ,$$
i.e. $\|\hat{x}\|_\infty = \|x\|$. Thus $A \ni x \mapsto \hat{x} \in \hat{A}$ is an isometry. Hence $\hat{A}$ is complete and thus closed in $C(\Delta)$. Since $\hat{A}$ is dense, $\hat{A} = C(\Delta)$. □

Let A be a C^*-algebra and $x \in A$. We say that x is *normal* is $x^*x = xx^*$. Note that self-adjoint elements are normal. Let B be the closure of the space of all polynomials in x and x^*, i.e. polynomials of the form $\sum_{i,j} a_{ij}x^i(x^*)^j$. Clearly B is a commutative C^*-subalgebra of B generated by x and polynomials in x and x^* are dense in B.

Denote the spectrum of x with respect to A and B by $\sigma_A(x)$ and $\sigma_B(x)$. Clearly $\sigma_A(x) \subset \sigma_B(x)$.

Theorem 18.10. *If A is a C^*-algebra, $x \in A$ is normal and B is defined as above, then*
$$\sigma_A(x) = \sigma_B(x).$$

Proof. We have to prove that if $y = \lambda e - x$ is invertible in A, then $y^{-1} \in B$. yy^* is self-adjoint and invertible in A. Hence $\sigma_A(yy^*) \subset \mathbb{R}$ is compact, so $\Omega = \mathbb{C} \setminus \sigma_A(yy^*)$ is connected. The function
$$\Omega \ni \lambda \mapsto (\lambda e - yy^*)^{-1} \in A$$
is holomorphic and on the open set $\mathbb{C} \setminus \sigma_B(yy^*) \subset \Omega$ it takes values into B. Hence it follows from Proposition 12.4 that it takes values into B on all of Ω. Since yy^* is invertible in A, $0 \in \Omega$ and hence $(yy^*)^{-1} = -(0 \cdot e - yy^*)^{-1} \in B$. Since $yy^* = y^*y$ is invertible in B, Lemma 17.13 implies that y is invertible in B. □

Theorem 18.11 (Spectral mapping theorem)**.** *Let B be a commutative C^*-algebra and $x \in B$ be such that polynomials in x and x^* are dense in B. Then the mapping $\hat{x} : \Delta \to \sigma(x)$ is a homeomorphism. Hence $C(\sigma(x)) \ni$*

$f \mapsto f \circ \hat{x} \in C(\Delta)$ is an isometric isomorphism of algebras. Thus for every $f \in C(\sigma(x))$ there is unique element $\Psi f \in B$ such that

$$(\Psi f)\hat{} = f \circ \hat{x}$$

and hence Ψ defines an isometric isomorphism of $C(\sigma(x))$ onto B that satisfies

$$\Psi \overline{f} = (\Psi f)^* \quad \text{for all } f \in C(\sigma(x)). \tag{18.8}$$

Moreover, if $f(\lambda) = \lambda$ on $\sigma(x)$, then $\Psi f = x$.

Proof. The mapping $\hat{x} : \Delta \to \sigma(x)$ is a continuous[61] surjection by Theorem 18.6(b). In order to prove that $\hat{x}$ is a homeomorphism it suffices to prove that $\hat{x}$ is one-to-one.[62] Suppose $\phi_1, \phi_2 \in \Delta$ and $\hat{x}(\phi_1) = \hat{x}(\phi_2)$, i.e. $\phi_1(x) = \phi_2(x)$. Hence $\phi_1(x^*) = \phi_2(x^*)$ by Lemma 18.8. Since ϕ_1 and ϕ_2 are homomorphisms it follows that $\phi_1(P(x, x^*)) = \phi_2(P(x, x^*))$ for any polynomial P. Now the density of such polynomials in B and continuity of ϕ_1, ϕ_2 imply that $\phi_1(y) = \phi_2(y)$ for all $y \in B$, i.e. $\phi_1 = \phi_2$. Thus $\hat{x} : \Delta \to \sigma(x)$ is one-to-one and hence homeomorphism. Equality (18.8) follows from (18.7). If $f(\lambda) = \lambda$, then $(\Psi f)\hat{} = f \circ \hat{x} = \hat{x}$, so $\Psi f = x$. □

Corollary 18.12. *Let A be a C^*-algebra and let $x \in A$ be normal. Then*

(a) $\|x\| = \rho(x)$.
(b) *x is self-adjoint if and only if $\sigma(x) \subset \mathbb{R}$.*

Proof. Let B be a commutative C^*-algebra defined as above. Recall that

$$\sigma_A(x) = \sigma_B(x) \ (= \sigma(x)) \quad \text{and hence} \quad \rho_A(x) = \rho_B(x) \ (= \rho(x)).$$

(a) Let $f \in C(\sigma(x))$, $f(\lambda) = \lambda$. Then

$$\rho(x) = \|f\|_\infty = \|\Psi f\| = \|x\|.$$

(b) follows directly from the last statement in the Gelfand-Najmark theorem.
□

In the situation described in Theorem 18.11 we write

$$\Psi f = f(x)\,. \tag{18.9}$$

If $P(\lambda) = \sum_{i,j} a_{ij}\lambda^i\overline{\lambda}^j$ is a polynomial, then it is easy to see that $P(x)$ defined by (18.9) coincides with[63]

$$P(x) = \sum_{i,j} a_{ij} x^i (x^*)^j$$

Note that such polynomials $P(\lambda)$ are dense in $C(\sigma(x))$. Theorem 18.11 allows us to define not only polynomial functions of x, but any continuous functions

[61]By the definition of the Gelfand topology.
[62]Because Δ is compact.
[63]Prove it.

of x for $f \in C(\sigma(x))$. Since $\Psi : C(\sigma(x)) \to B$ is an isometric isomorphism of algebras we have that

$$(f+g)(x) = f(x) + g(x), \quad (fg)(x) = f(x)g(x), \quad \overline{f}(x) = (f(x))^*$$
$$\|f(x)\| = \|f\|_\infty.$$

Moreover

$$\sigma(f(x)) = f(\sigma(x)).$$

Indeed, according to Theorem 18.3(e), for any $y \in B$, $\sigma(y) = \hat{y}(\Delta)$, so

$$\begin{aligned}\sigma(f(x)) &= (f(x))\hat{}\,(\Delta) = (\Psi f)\hat{}\,(\Delta) = (f \circ \hat{x})(\Delta) \\ &= f(\hat{x}(\Delta)) = f(\sigma(x)).\end{aligned}$$

Corollary 18.13. *Let A be a C^*-algebra and let $x \in A$ be self-adjoint with $\sigma(x) \subset [0, \infty)$. Then there is another self-adjoint element $y \in A$ such that $y^2 = x$.*

Proof. Let B be the closure of polynomials in x and x^*. Then B is commutative C^*-algebra. We will apply Theorem 18.11 to the algebra B. We have $\sigma_B(x) = \sigma_A(x) \subset [0, \infty)$ by Theorem 18.10. The function $f(t) = t^{1/2}$ is continuous on $\sigma(x)$ and hence $y = f(x)$ is well defined. We have

$$y^2 = f(x)f(x) = (f \cdot f)(x) = x$$

Moreover

$$y^* = (f(x))^* = \overline{f}(x) = f(x) = y,$$

so y is self-adjoint. □

18.2. **Applications to the spectral theory.** Let H be a Hilbert space and $T \in B(H)$ be a normal operator, i.e. $TT^* = T^*T$. With each polynomial

$$P(\lambda) = \sum_{i,j} a_{ij} \lambda^i \overline{\lambda}^j \tag{18.10}$$

we associate an operator

$$P(T) = \sum_{i,j} a_{ij} T^i (T^*)^j \tag{18.11}$$

Let $A \subset B(H)$ be the closure of the space of all operators of the form (18.11). Clearly A is a commutative C^*-subalgebra of $B(H)$ generated by T. Note that polynomials (18.10) are dense in $C(\sigma(T))$. As a consequence of Theorem 18.11 we have

Theorem 18.14 (Spectral mapping theorem)**.** *The mapping $P \mapsto P(T)$ uniquely extends to the isometric isomorphism of algebras*

$$C(\sigma(T)) \ni f \mapsto f(T) \in A.$$

Moreover $\overline{f}(T) = f(T)^*$ and $\sigma(f(T)) = f(\sigma(T))$.

Also Corollary 18.13 can be stated now as

Theorem 18.15. *If $T \in B(H)$ is self-adjoint and $\sigma(T) \subset [0, \infty)$, then there is a self-adjoint operator S such that $S^2 = T$.*

It is very useful to identify the class of operators mentioned in Theorem 18.15.

Theorem 18.16. *Suppose $T \in B(H)$. Then the following conditions are equivalent:*

(a) $\langle Tx, x\rangle \geq 0$ *for all* $x \in H$.
(b) $T = T^*$ *and* $\sigma(T) \subset [0, \infty)$.

Operators satisfying condition (a) are called *positive*.

Proof. Suppose (a). Then
$$\langle Tx, x\rangle = \overline{\langle x, Tx\rangle} = \langle x, Tx\rangle = \langle T^*x, x\rangle,$$
so $\langle (T - T^*)x, x\rangle = 0$ for all $x \in H$ and it easily follows from the polarization identity that $T - T^* = 0$. Thus T is self-adjoint and hence $\sigma(T) \subset \mathbb{R}$ by Theorem 18.7. In order to prove that $\sigma(T) \subset [0, \infty)$ it suffices to prove that for any $\lambda > 0$, $T + \lambda I$ is invertible. The inequality at (a) yields
$$\lambda\|x\|^2 = \langle \lambda x, x\rangle \leq \langle (T + \lambda I)x, x\rangle \leq \|(T + \lambda I)x\|\|x\|,$$
so
$$\|(T + \lambda I)x\| \geq \lambda\|x\|. \tag{18.12}$$
This inequality easily implies that the operator $T + \lambda I$ is one-to-one and that the range $\mathcal{R}(T + \lambda I)$ is closed. It remains to prove that the operator is surjective, i.e. $\mathcal{R}(T + \lambda I)^\perp = \{0\}$. If $y \in \mathcal{R}(T + \lambda I)^\perp$, the for any $x \in H$,
$$0 = \langle (T + \lambda I)x, y\rangle = \langle x, (T + \lambda I)y\rangle$$
and hence $(T + \lambda I)y = 0$, so $y = 0$ by (18.12).

Suppose now (b). Let S be a self-adjoint operator such that $S^2 = T$. We have
$$\langle Tx, x\rangle = \langle S^2x, x\rangle = \langle Sx, Sx\rangle \geq 0$$
and hence (a) follows. □

Department of Mathematics, University of Pittsburgh, 301 Thackeray Hall, Pittsburgh, PA 15260, USA, hajlasz@pitt.edu