Chapter 1

若到松江呼小渡，莫驚鴛鷺，
四橋盡是、老子經行處。
蘇軾《青玉案》

Normed Space: Examples

Generally speaking, in functional analysis we study infinite dimensional vector spaces of functions and the linear operators between them by analytic methods. This chapter is of preparatory nature. First, we use Zorn's lemma to prove there is always a basis for any vector space. It fills up a gap in elementary linear algebra where the proof was only given for finite dimensional vector spaces. The inadequacy of this notion of basis for infinite dimensional spaces motivates the introduction of analysis to the study of function spaces. Second, we discuss three basic inequalities, namely, Young's, Hölder's, and Minkowski's inequalities. We establish Young's inequality by elementary means, use it to deduce Hölder's inequality, and in term use Hölder's inequality to prove Minkowski's inequality. The latter will be used to introduce norms on some common vector spaces. As you will see, these spaces form our principal examples throughout this book.

1.1 Vector Spaces of Functions

Recall that a vector space is over a field $\mathbb{F}$. Throughout this book it is always assumed this field is either the real field $\mathbb{R}$ or the complex field $\mathbb{C}$. In the following $\mathbb{F}$ stands for $\mathbb{R}$ or $\mathbb{C}$.

It is true that many vector spaces can be viewed as vector spaces of functions. To describe this unified point of view, let S be a non-empty set and denote the collection of all functions from S to $\mathbb{F}$ by $F(S)$. It is routine to check that $F(S)$ forms a vector space over $\mathbb{F}$ under the obvious rules of addition and scalar multiplication for functions: For $f, g \in F(S)$ and $\alpha \in \mathbb{F}$,

$$(f+g)(p) \equiv f(p) + g(p), \qquad (\alpha f)(p) \equiv \alpha f(p).$$

In fact, these algebraic operations are inherited from the target $\mathbb{F}$.

First, take $S = \{p_1, \cdots, p_n\}$ a set consisting of n many elements. Every function $f \in F(S)$ is uniquely determined by its values at $p_1, \cdots, p_n$, so f can be identified with the n-triple $(f(p_1), \cdots, f(p_n))$. It is easy to see that $F(\{p_1, \cdots, p_n\})$ is linearly isomorphic to $\mathbb{F}^n$. More precisely, the mapping $f \mapsto (f(p_1), \cdots, f(p_n))$ is a linear bijection between $F(\{p_1, \cdots, p_n\})$ and $\mathbb{F}^n$.

Second, take $S = \{p_1, p_2, \cdots\}$. As above, any $f \in F(S)$ can be identified with the sequence $(f(p_1), f(p_2), f(p_3) \cdots)$. The vector space $F(\{p_j\}_{j=1}^{\infty})$ may be called the space of sequences over $\mathbb{F}$.

Finally, taking $S = [0,1]$, $F([0,1])$ consists of all $\mathbb{F}$-valued functions.

The vector spaces we are going to encounter are mostly these spaces and their subspaces.

1.2 Zorn's Lemma

In linear algebra, it was pointed out that every vector space has a basis no matter it is of finite or infinite dimension, but the proof was only given in the finite dimensional case. Here we provide a proof of the general case. The proof depends critically on Zorn's lemma, an assertion equivalent to the axiom of choice.

To formulate Zorn's lemma, we need to consider a partial order on a set.

A relation $\leq$ on a non-empty set X is called a **partial order** on X if it satisfies

(PO1) $x \leq x, \ \forall x \in X$;

(PO2) $x \leq y$ and $y \leq x$ implies $x = y$.

(PO3) $x \leq y, \ y \leq z$ implies $x \leq z$.

The pair $(X, \leq)$ is called a **partially ordered set** or a **poset** for short. A non-empty subset Y of X is called a **chain** or a **totally ordered set** if for any two $y_1, y_2 \in Y$, either $y_1 \leq y_2$ or $y_2 \leq y_1$ holds. In other words, every pair of elements in Y are related. An **upper bound** of a non-empty subset Y of X is an element u, which may or may not be in Y, such that $y \leq u$ for all $y \in Y$. Finally, a **maximal element** of $(X, \leq)$ is an element z in X such that $z \leq x$ implies $z = x$.

Example 1.1. Let S be a set and consider $X = \mathscr{P}(S)$, the power set of S. It is clear that the relation "set inclusion" $A \subset B$ is a partial order on $\mathscr{P}(S)$. It has a unique maximal element given by S itself.

Example 1.2. Let $X = \mathbb{R}^2$ and define $x \prec y$ if and only if $x_1 \leq y_1$ and $x_2 \leq y_2$. For instance, $(-1, 5) \prec (0, 8)$ but $(-2, 3)$ and $(35, -1)$ are unrelated.

Then $(X, \prec)$ forms a poset without any maximal element.

Zorn's Lemma. *Let $(X, \leq)$ be a poset. If every chain in X has an upper bound, then X has at least one maximal element.*

Although called a lemma by historical reason, Zorn's lemma, a constituent in the Zermelo-Fraenkel set theory, is an axiom in nature. It is equivalent to the axiom of choice as well as the Hausdorff maximality principle. You may look up Hewitt-Stromberg's "Real and Abstract Analysis" for further information. A readable account on this "lemma" can also be found in Wikipedia.

1.3 Existence of Basis

As a standard application of Zorn's lemma, we show there is a basis in any vector space. To refresh your memory, let's recall that a subset S in a vector space X is called a linearly independent set if any finite number of vectors in S are linearly independent. In other words, letting $\{x_1, \cdots, x_n\}$ be any subset of S, if $\alpha_1 x_1 + \cdots + \alpha_n x_n = 0$ for some scalars $\alpha_i,\ i = 1, \cdots, n$, then $\alpha_i = 0$ for all i. On the other hand, given any subset S, denote all linear combinations of vectors from S by $\langle S \rangle$. It is easy to check that $\langle S \rangle$ forms a subspace of X called the subspace spanned by S. A subset S is called a spanning set of X if $\langle S \rangle$ is X, and it is called a basis of X if it is also a linearly independent spanning set. When X admits a finite spanning set, it has a basis consisting of finitely many vectors. Moreover, all bases have the same number of vectors and we call this number the dimension of the space X. The space X is of infinite dimension if it does not have a finite spanning set.

Theorem 1.1. *Every non-zero vector space has a basis.*

This basis is sometimes called a **Hamel basis**.

Proof. Let $\mathcal{X}$ be the set of all linearly independent subsets of a given vector space V. Since V is non-zero, $\mathcal{X}$ is a non-empty set. Clearly the set inclusion $\subset$ makes it into a poset. To apply Zorn's lemma, let's verify that every chain in it has an upper bound. Let $\mathcal{Y}$ be a chain in $\mathcal{X}$, consider the following subset of V,

$$S = \bigcup_{C \in \mathcal{Y}} C.$$

We claim that (i) $S \in \mathcal{X}$, that's, S is a linearly independent set, (ii) $C \subset S$, $\forall C \in \mathcal{Y}$, that's, S is an upper bound of $\mathcal{Y}$. Since (ii) is obvious, it is sufficient to verify (i).

To this end, pick $v_1, \cdots, v_n \in S$. By definition, we can find $C_1, \cdots, C_n$ in $\mathcal{Y}$ such that $v_1 \in C_1, \cdots, v_n \in C_n$. As $\mathcal{Y}$ is a chain, $C_1, \cdots, C_n$ satisfy $C_i \subset C_j$ or $C_j \subset C_i$ for any i, j. After rearranging the indices, one may assume $C_1 \subset C_2 \subset \cdots \subset C_n$, and so $\{v_1, \cdots, v_n\} \subset C_n$. Since C_n is a linearly independent set, $\{v_1, \cdots, v_n\}$ is linearly independent. This shows that S is a linearly independent set.

After showing that every chain in $\mathcal{X}$ has an upper bound, we appeal to Zorn's lemma to conclude that $\mathcal{X}$ has a maximal element B. We claim that B is a basis for V. For, first of all, B belonging to $\mathcal{X}$ means that B is a linearly independent set. To show that it spans V, we pick $v \in V$. Suppose v does not belong to $\langle B \rangle$, so v is independent from all vectors in B. But then the set $\widetilde{B} = B \cup \{v\}$ is a linearly independent set which contains B as its proper subset, contradicting the maximality of B. We conclude that $\langle B \rangle = V$, so B forms a basis of V. □

The following example may help you in understanding the proof of Theorem 1.1.

Example 1.3. Consider the power set of $\mathbb{R}^3$ which is partially ordered by set inclusion. Let $\mathcal{X}$ be the subset of all linearly independent sets in $\mathbb{R}^3$. Then

$$\mathcal{Y}_1 \equiv \Big\{\{(1,0,0)\}, \{(1,0,0),(1,1,0)\}, \{(1,0,0),(1,1,0),(0,0,-3)\}\Big\}$$

and

$$\mathcal{Y}_2 \equiv \Big\{\{(1,3,5),(2,4,6)\}, \{(1,3,5),(2,4,6),(1,0,0)\}\Big\}$$

are chains but

$$\mathcal{Y}_3 \equiv \Big\{\{(1,0,0)\}, \{(1,0,0),(0,1,0)\}, \{(1,0,0),(0,-2,0),(0,0,1)\}\Big\}$$

is not a chain in $\mathcal{X}$.

For a finite dimensional vector space, it is relatively easy to find an explicit basis, and bases are used in many occasions such as in the determination of the dimension of the vector space and in the representation of a linear operator as a matrix. However, in contrast, the existence of a basis in infinite dimensional space is proved via a non-constructive argument. It is not easy to write down a basis. For example, consider the space of sequences $\mathcal{S} \equiv \{x = (x_1, x_2, \cdots, x_n \cdots) : \ x_i \in \mathbb{F}\}$. Letting $e_j = (0, \cdots, 1, \cdots)$ where "1" appears in the j-th place, it is tempting from the formula $x = \sum_{j=1}^{\infty} x_j e_j$ to assert that $\{e_j\}_1^\infty$ forms a basis for $\mathcal{S}$. But, this is not true. Why? It is because infinite sums are not linear combinations. Indeed, one cannot talk about infinite sums in a vector space as there is no means to measure convergence. According to Theorem 1.1, however, there is a rather mysterious basis. In general, a non-explicit basis is difficult to work with, and thus lessens its importance in the study of infinite dimensional spaces. To proceed further, analytical structures will be added to vector spaces. Later, we will see that for a reasonably nice infinite dimensional vector space, any basis must consist of uncountably many vectors (see Proposition 4.14). Suitable generalizations of this notion are needed. For an infinite dimensional normed space, one may introduce the so-called Schauder basis as a replacement. For a complete inner product spaces (a Hilbert space), an even more useful notion, a complete orthonormal set, will be much more useful.

Mathematics is a deductive science. A limited number of axioms is needed to build up the tower of mathematics, and Zorn's lemma is one of them. We will encounter this lemma again in later chapters. You may also google for more of its applications.

1.4 Three Inequalities

Now we come to Young's, Hölder's and Minkowski's inequalities.

Two positive numbers p and q are **conjugate** if $1/p+1/q=1$. Notice that they must be greater than one and q approaches infinity as p approaches 1. In the following paragraphs q is always conjugate to p.

Proposition 1.2 (Young's Inequality). *For any $a, b > 0$ and $p > 1$,*

$$ab \le \frac{a^p}{p} + \frac{b^q}{q},$$

and equality holds if and only if $a^p = b^q$.

Proof. Consider the function

$$\varphi(x) = \frac{x^p}{p} + \frac{1}{q} - x, \quad x \in (0, \infty).$$

From the sign of $\varphi'(x) = x^{p-1} - 1$ we see that φ is strictly decreasing on $(0,1)$ and strictly increasing on $(1,\infty)$. It follows that $x = 1$ is the strict minimum of φ on $(0,\infty)$. So, $\varphi(x) \ge \varphi(1)$ and equality holds if and only if $x = 1$. In other words,

$$\frac{x^p}{p} + \frac{1}{q} - x \ge \frac{1}{p} + \frac{1}{q} - 1,$$

that is ,

$$\frac{x^p}{p} + \frac{1}{q} \ge x.$$

Letting $x = ab/b^q$, we get the Young's inequality. Equality holds if and only if $ab/b^q = 1$, i.e., $a^p = b^q$. □

Proposition 1.3 (Hölder's Inequality). *For $a, b \in \mathbb{R}^n$, $p > 1$,*

$$\sum_{k=1}^{n} |a_k||b_k| \leq \|a\|_p \|b\|_q,$$

where $\|a\|_p = (\sum_{k=1}^n |a_k|^p)^{\frac{1}{p}}$ and $\|b\|_q = (\sum_{k=1}^n |b_k|^q)^{\frac{1}{q}}$.

Proof. The inequality clearly holds when $a = (0, \cdots, 0)$. We may assume $a \neq (0, \cdots, 0)$ in the following proof. By Young's inequality, for each $\varepsilon > 0$ and k,

$$|a_k b_k| = |\varepsilon a_k||\varepsilon^{-1} b_k| \leq \frac{\varepsilon^p |a_k|^p}{p} + \frac{\varepsilon^{-q}|b_k|^q}{q}.$$

Thus

$$\begin{aligned} \sum_{k=1}^{n} |a_k||b_k| &= |a_1||b_1| + \cdots |a_n||b_n| \\ &\leq \frac{\varepsilon^p}{p} \sum_{k=1}^{n} |a_k|^p + \frac{\varepsilon^{-q}}{q} \sum_{k=1}^{n} |b_k|^q \\ &= \frac{\varepsilon^p}{p} \|a\|_p^p + \frac{\varepsilon^{-q}}{q} \|b\|_q^q, \qquad (1.1) \end{aligned}$$

for any $\varepsilon > 0$. To have the best choice of ε, we minimize the right hand side of this inequality. Taking derivative of the right hand side of (1.1) as a function of ε, we obtain

$$\varepsilon^{p-1} \|a\|_p^p - \varepsilon^{-q-1} \|b\|_q^q = 0,$$

that is,

$$\varepsilon = \frac{\|b\|_q^{\frac{q}{p+q}}}{\|a\|_p^{\frac{p}{p+q}}}.$$

is the minimum point. (Clearly this function has only one critical point and does not have any maximum.) Plugging this choice of ε into the inequality yields the Hölder's inequality after some manipulation. □

Proposition 1.4 (Minkowski's Inequality). *For $a, b \in \mathbb{F}^n$ and $p \geq 1$,*

$$\|a+b\|_p \leq \|a\|_p + \|b\|_p.$$

Proof. The inequality clearly holds when $p = 1$ or $\|a+b\| = 0$. In the following proof we may assume $p > 1$ and $\|a+b\| > 0$. For each k,

$$\begin{aligned} |a_k + b_k|^p &= |a_k + b_k||a_k + b_k|^{p-1} \\ &\leq |a_k||a_k + b_k|^{p-1} + |b_k||a_k + b_k|^{p-1}. \end{aligned} \tag{1.2}$$

Applying Hölder's inequality to the two terms on right hand side of (1.2) separately (more precisely, to the pairs of *real* vectors $(|a_1|, \cdots, |a_n|)$ and $(|a_1 + b_1|^{p-1}, \cdots, |a_n + b_n|^{p-1})$, and $(|b_1|, \cdots, |b_n|)$ and $(|a_1 + b_1|^{p-1}, \cdots, |a_n + b_n|^{p-1})$), we have

$$\begin{aligned} \sum_{k=1}^{n} |a_k + b_k|^p &\leq \|a\|_p \left(\sum_{k=1}^{n} |a_k + b_k|^{(p-1)q} \right)^{\frac{1}{q}} + \|b\|_p \left(\sum_{k=1}^{n} |a_k + b_k|^{(p-1)q} \right)^{\frac{1}{q}} \\ &= (\|a\|_p + \|b\|_p) \left(\sum_{k=1}^{n} |a_k + b_k|^p \right)^{\frac{1}{q}}, \end{aligned}$$

and Minkowski's inequality follows. □

Look up Wikipedia for the great mathematician Hermann Minkowski (1864-1909), the best friend of David Hilbert and a teacher of Albert Einstein, who died unexpectedly at forty-five. The biography "Hilbert" by C. Reid contains an interesting account on Minkowski and Hilbert.

The last two inequalities allow the following generalization.

Hölder's Inequality for Sequences. *For any two sequences a and b in $\mathbb{F}$, and $p > 1$,*

$$\sum_{k=1}^{\infty} |a_k||b_k| \leq \|a\|_p \|b\|_q,$$

where now the summation in the sums on the right runs from 1 *to* ∞*.*

Since the norms $\|a\|_p$ and $\|b\|_q$ are allowed to be zero or infinity, we adopt the convention $0 \times \infty = 0$ in the above inequality.

Minkowski's Inequality for Sequences. *For any two sequences* a *and* b *in* $\mathbb{F}$ *and* $p \geq 1$*,*

$$\|a+b\|_p \leq \|a\|_p + \|b\|_p,$$

where now the summation in the sums runs from 1 *to* ∞*.*

Hölder's Inequality for Functions. *For* $p > 1$ *and Riemann integrable functions* f *and* g *on* $[a,b]$*, we have*

$$\int_a^b |fg| \leq \left(\int_a^b |f|^p\right)^{\frac{1}{p}} \left(\int_a^b |g|^q\right)^{\frac{1}{q}}.$$

Minkowski's Inequality for Functions. *For* $p \geq 1$ *and Riemann integrable functions* f *and* g *on* $[a,b]$*, we have*

$$\left(\int_a^b |f+g|^p\right)^{\frac{1}{p}} \leq \left(\int_a^b |f|^p\right)^{\frac{1}{p}} + \left(\int_a^b |g|^p\right)^{\frac{1}{p}},$$

We leave the proofs of these generalizations as exercises.

1.5 Normed Vector Spaces

Let $(X, +, \cdot)$ be a vector space over $\mathbb{F}$. A **norm** on X is a function from X to $[0,\infty)$ satisfying the following three properties: For all $x, y \in X$ and $\alpha \in \mathbb{F}$,

(N1) $\|x\| \geq 0$ and "$=$" holds if and only if $x = 0$, (N2) $\|x+y\| \le \|x\| + \|y\|$,

(N3) $\|\alpha x\| = |\alpha|\|x\|$.

The vector space with a norm, $(X, +, \cdot, \|\cdot\|)$, or $(X, \|\cdot\|)$, or even stripped to a single X when the context is clear, is called a **normed vector space** or simply a **normed space**.

Here are some normed vector spaces.

Example 1.4. $(\mathbb{F}^n, \|\cdot\|_p), 1 \le p < \infty$, where

$$\|x\|_p = \left(\sum_{k=1}^{n} |x_k|^p\right)^{\frac{1}{p}}.$$

Clearly, (N1) and (N3) hold. According to the Minkowski's inequality (N2) holds too. When $p = 2$ and $\mathbb{F}^n = \mathbb{R}^n$ or $\mathbb{C}^n$, the norm is called the **Euclidean norm** or the **unitary norm**.

Example 1.5. $(\mathbb{F}^n, \|\cdot\|_\infty)$ where

$$\|x\|_\infty = \max_{k=1,\cdots,n} |x_k|.$$

is called the sup-norm.

Example 1.6. Let ℓ^p, $1 \le p < \infty$, be the collection of all $\mathbb{F}$-valued sequences $x = (x_1, x_2, \cdots)$ satisfying

$$\sum_{k=1}^{\infty} |x_k|^p < \infty.$$

First of all, from the Minkowski's inequality for sequences the sum of two sequences in ℓ^p belongs to ℓ^p. With the other easily checked properties, ℓ^p forms a vector space. The function $\|\cdot\|_p$, i.e.,

$$\|x\|_p = \left(\sum_{k=1}^{\infty} |x_k|^p\right)^{\frac{1}{p}}$$

clearly satisfies (N1) and (N3). Moreover, (N2) also holds by Minkowski's inequality for sequences. Hence it defines a norm on ℓ^p.

Example 1.7. Let ℓ^∞ be the collection of all $\mathbb{F}$-valued bounded sequences. Define the sup-norm

$$\|x\|_\infty = \sup_k |x_k|.$$

Clearly ℓ^∞ forms a normed vector space over $\mathbb{F}$.

Example 1.8. Let $C[a,b]$ be the vector space of all continuous functions on the interval $[a,b]$. For $1 \leq p < \infty$, define

$$\|f\|_p = \left(\int_a^b |f(x)|^p dx\right)^{\frac{1}{p}}.$$

By the Minkowski's inequality for functions, one sees that $(C[a,b], \|\cdot\|_p)$ forms a normed space under this norm.

Example 1.9. Let $B([a,b])$ be the vector space of all bounded functions on $[a,b]$. The sup-norm

$$\|f\|_\infty = \sup_{x\in[a,b]} |f(x)|$$

defines a norm on $B([a,b])$.

Example 1.10. One may have already observed that the normed spaces in Examples 1.5, 1.7 and 1.9 are of the same nature. In fact, let $F_b(S)$ be the vector subspace of $F(S)$ consisting of all bounded functions from S to $\mathbb{F}$. The sub-norm can be defined on $F_b(S)$ and these examples are special cases obtained by taking different sets S.

Example 1.11. Any vector subspace of a normed vector space forms a normed vector space under the same norm. In this way we obtain many many normed vector spaces. Here are some examples: The space of all convergent sequences, $\mathcal{C}$, the space of all sequences which converges to 0, $\mathcal{C}_0$, and the space of all sequences which have finitely many non-zero terms, $\mathcal{C}_{00}$, are normed subspaces of ℓ^∞ under the sup-norm. The space of all continuous functions on $[a,b]$, $C[a,b]$, is an important normed subspace of $B([a,b])$. The spaces $\{f : f(a) = 0,\ f \in C[a,b]\}, \{f : f \text{ is differentiable}, f \in C[a,b]\}$ and $\{f : f \text{ is the restriction of a polynomial on } [a,b]\}$ are normed subspaces of $C[a,b]$ under the sup-norm. But the set $\{f :\ f(a) = 1,\ f \in C[a,b]\}$ is not a normed space because it is not a subspace.

To accommodate more applications, one needs to replace $[a,b]$ by more general sets in the examples above. For any closed and bounded subset K in $\mathbb{R}^n$, one may define $C(K)$ to be the collection of all continuous functions in K. As any continuous function in a closed and bounded set must be bounded (with its maximum attained at some point), its sup-norm is well-defined. Thus $(C(K), \|\cdot\|_\infty)$ forms a normed space. On the other hand, let R be any rectangular box in $\mathbb{R}^n$. We know that Riemann integration makes sense for bounded, continuous functions in R. Consequently, we may introduce the normed $\|\cdot\|_p = (\int_R |f|^p)^{1/p}$ to make all bounded, continuous functions in R a normed space. However, this p-norm does not form a norm on the space of Riemann integrable functions. Which axiom of the norm is not satisfied?

In addition to Example 10 where new normed spaces are found by restricting to subspaces, there are two more general ways to obtain them. For any two given normed spaces $(X, \|\cdot\|_1)$ and $(Y, \|\cdot\|_2)$ the function $\|(x,y)\| = \|x\|_1 + \|y\|_2$ defines a norm on the product space $X \times Y$ and thus makes $X \times Y$ the product normed space. On the other hand, to each subspace of a normed space one may form a corresponding quotient space and endow it the quotient norm. We will do this in the next chapter.

These examples of normed spaces will be used throughout this book. For simplicity the norm of the space will usually be suppressed. For instance, $\mathbb{F}^n$ always stands for the normed space under the Euclidean or the unitary norm,

ℓ^p and ℓ^∞ are always under the p-norms and sup-norm respectively and a single $C(K)$ refers to the space of continuous functions on the closed, bounded set K under the sup-norm.