B

Background Mathematics

B.1 Vector Spaces

We begin with the definition of a vector space. Where appropriate, we will give simpler definitions, which at the expense of some generality will be sufficient for the use made of them in the text. For example a vector space can be defined over any field, but we will consider vector spaces over the real numbers, so that what we now introduce are sometimes called 'real vector spaces'.

Definition B.1 A set X is a *vector space (VS)* if two operations (addition, and multiplication by scalar) are defined on X such that, for $\mathbf{x}$, $\mathbf{y} \in X$, and $\alpha \in \mathbb{R}$,

$$\begin{aligned} \mathbf{x}+\mathbf{y} &\in X, \\ \alpha\mathbf{x} &\in X, \\ 1\mathbf{x} &= \mathbf{x}, \\ 0\mathbf{x} &= \mathbf{0}, \end{aligned}$$

and such that in addition X is a commutative group with identity $\mathbf{0}$ under the addition operation and satisfies the distributive laws for scalar multiplication

$$\alpha(\mathbf{x}+\mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y},$$

and

$$(\alpha+\beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x},$$

The elements of X are also called *vectors* while the real numbers are referred to as *scalars*.

Example B.2 The standard example of a VS is the set $\mathbb{R}^n$ of real column vectors of fixed dimension n. We use a dash to denote transposition of vectors (and matrices) so that a general column vector can be written as

$$\mathbf{x} = (x_1, \ldots, x_n)',$$

where $x_i \in \mathbb{R}$, $i = 1, \ldots, n$.

Definition B.3 A non-empty subset M of a vector space X is a *subspace* of X, if the restriction of the operations to M makes it a vector space.

Definition B.4 A *linear combination* of the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ in a vector space is a sum of the form $\alpha_1 \mathbf{x}_1 + \cdots + \alpha_n \mathbf{x}_n$ where $\alpha_i \in \mathbb{R}$. If the α_i are positive and $\sum_{i=1}^n \alpha_i = 1$, the sum is also called a *convex combination*.

It is easy to see that a linear combination of vectors from a subspace is still in the subspace, and that linear combinations can actually be used to build subspaces from an arbitrary subset of a VS. If S is a subset of X, we denote by $\mathrm{span}(S)$ the subspace of all possible linear combinations of vectors in S.

Definition B.5 A finite set of vectors $S = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is *linearly dependent* if it is possible to find constants $\alpha_1, \ldots, \alpha_n$ not all zero such that

$$\alpha_1 \mathbf{x}_1 + \cdots + \alpha_n \mathbf{x}_n = \mathbf{0}.$$

If this is not possible, the vectors are called *linearly independent*.

This implies that a vector $\mathbf{y}$ in $\mathrm{span}(S)$,where S is a set of linearly independent vectors, has a unique expression of the form

$$\mathbf{y} = \alpha_1 \mathbf{x}_1 + \ldots + \alpha_n \mathbf{x}_n,$$

for some n and $\mathbf{x}_i \in S$, $i = 1, \ldots, n$.

Definition B.6 A set of vectors $S = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is said to form a *basis* for X if S is linearly independent, and if every $\mathbf{x} \in X$ can be uniquely expressed as a linear combination of the vectors in the set S. Though there will always be many different bases, their sizes are always the same. The size of a basis of a vector space is known as its *dimension*.

Finite dimensional spaces are easier to study, as fewer pathological cases arise. Special requirements have to be added in order to extend to the infinite dimensional case some basic properties of finite dimensional VSs. We consider introducing a measure of distance in a VS.

Definition B.7 A *normed linear space* is a VS X together with a real-valued function that maps each element $x \in X$ into a real number $\|\mathbf{x}\|$ called the *norm* of $\mathbf{x}$, satisfying the following three properties:

1. **Positivity** $\|\mathbf{x}\| \geq 0$, $\forall \mathbf{x} \in X$, equality holding if and only if $\mathbf{x} = \mathbf{0}$;

2. **Triangle inequality** $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$, $\forall \mathbf{x}, \mathbf{y} \in X$;

3. **Homogeneity** $\|\alpha \mathbf{x}\| = |\alpha| \, \|\mathbf{x}\|$, $\forall \alpha \in \mathbb{R}$, and $\forall \mathbf{x} \in X$.

Example B.8 Consider countable sequences of real numbers and let $1 \leq p < \infty$. The space ℓ_p is the set of sequences $\mathbf{z} = \{z_1, z_2, \ldots, z_i, \ldots\}$ such that

$$\|\mathbf{z}\|_p = \left(\sum_{i=1}^{\infty} |z_i|^p\right)^{1/p} < \infty.$$

The space ℓ_∞ is formed of those sequences $\mathbf{z}$ that satisfy

$$\|\mathbf{z}\|_\infty = \max_{i \in \mathbb{N}} \left(|z_i|\right) < \infty.$$

The distance between two vectors $\mathbf{x}$ and $\mathbf{y}$ can be defined as the norm of their difference $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$.

Definition B.9 In a normed linear space an infinite sequence of vectors $\mathbf{x}_n$ is said to *converge* to a vector $\mathbf{x}$ if the sequence $\|\mathbf{x} - \mathbf{x}_n\|$ of real numbers converges to zero.

Definition B.10 A sequence $\mathbf{x}_n$ in a normed linear space is said to be a *Cauchy sequence* if $\|\mathbf{x}_n - \mathbf{x}_m\| \to 0$ as $n, m \to \infty$. More precisely, given $\varepsilon > 0$, there is an integer N such that $\|\mathbf{x}_n - \mathbf{x}_m\| < \varepsilon$ for all $n, m > N$. A space is said to be *complete* when every Cauchy sequence converges to an element of the space.

Note that in a normed space every convergent sequence is a Cauchy sequence, but the converse in not always true. Spaces in which every Cauchy sequence has a limit are said to be *complete*. Complete normed linear spaces are called *Banach spaces*.

B.2 Inner Product Spaces

The theory of inner product spaces is a tool used in geometry, algebra, calculus, functional analysis and approximation theory. We use it at different levels throughout this book, and it is useful to summarise here the main results we need. Once again, we give the definitions and basic results only for the real case.

Definition B.11 A function f from a vector space X to a vector space Y is said to be *linear* if for all $\alpha, \beta \in \mathbb{R}$ and $\mathbf{x}, \mathbf{y} \in X$,

$$f(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha f(\mathbf{x}) + \beta f(\mathbf{y}).$$

Note that we can view the real numbers as a vector space of dimension 1. Hence, a real-valued function is linear if is satisfies the same definition.

Example B.12 Let $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$. A linear function from X to Y can be denoted by an $m \times n$ matrix $\mathbf{A}$ with entries $\mathbf{A}_{ij}$ so that the vector $\mathbf{x} = (x_1, \ldots, x_n)'$ is mapped to the vector $\mathbf{y} = (y_1, \ldots, y_m)'$ where

$$y_i = \sum_{j=1}^{n} \mathbf{A}_{ij} x_j, \; i = 1, \ldots, m.$$

A matrix with entries $\mathbf{A}_{ij} = 0$, for $i \neq j$, is called diagonal.

Definition B.13 A vector space X is called an *inner product space* if there is a bilinear map (linear in each argument) that for each two elements $\mathbf{x}, \mathbf{y} \in X$ gives a real number denoted by $\langle \mathbf{x} \cdot \mathbf{y} \rangle$ satisfying the following properties:

- $\langle \mathbf{x} \cdot \mathbf{y} \rangle = \langle \mathbf{y} \cdot \mathbf{x} \rangle$,
- $\langle \mathbf{x} \cdot \mathbf{x} \rangle \geq 0$, and $\langle \mathbf{x} \cdot \mathbf{x} \rangle = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$.

The quantity $\langle \mathbf{x} \cdot \mathbf{y} \rangle$ is called the *inner product* of x and y, though it is also known as the dot product or scalar product.

Example B.14 Let $X = \mathbb{R}^n$, $\mathbf{x} = (x_1, \dots, x_n)'$, $\mathbf{y} = (y_1, \dots, y_n)'$. Let λ_i be fixed positive numbers. The following defines a valid inner product:

$$\langle \mathbf{x} \cdot \mathbf{y} \rangle = \sum_{i=1}^{n} \lambda_i x_i y_i = \mathbf{x}' \Lambda \mathbf{y},$$

where Λ is the $n \times n$ diagonal matrix with non zero entries $\Lambda_{ii} = \lambda_i$.

Example B.15 Let $X = C[a, b]$ be the vector space of continuous functions on the interval $[a, b]$ of the real line with the obvious definitions of addition and scalar multiplication. For $f, g \in X$, define

$$\langle f \cdot g \rangle = \int_a^b f(t) g(t) dt.$$

From the definition of an inner product, two further properties follow:

- $\langle \mathbf{0} \cdot \mathbf{y} \rangle = 0$,
- X is automatically a normed space, with the norm

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x} \cdot \mathbf{x} \rangle}.$$

Definition B.16 Two elements $\mathbf{x}$ and $\mathbf{y}$ of X are called *orthogonal* if $\langle \mathbf{x} \cdot \mathbf{y} \rangle = 0$. A set $S = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ of vectors from X is called *orthonormal* if $\langle \mathbf{x}_i \cdot \mathbf{x}_j \rangle = \delta_{ij}$, where $\delta_{ij} = 1$ if $i = j$, and 0 otherwise. For an orthonormal set S, and a vector $\mathbf{y} \in X$, the expression

$$\sum_{i=1}^{n} \langle \mathbf{x}_i \cdot \mathbf{y} \rangle \, \mathbf{x}_i$$

is said to be a *Fourier series* for $\mathbf{y}$.

If S forms an orthonormal basis each vector $\mathbf{y}$ is equal to its Fourier series.

Theorem B.17 *(Schwarz inequality) In an inner product space*

$$|\langle \mathbf{x} \cdot \mathbf{y} \rangle|^2 \leq \langle \mathbf{x} \cdot \mathbf{x} \rangle \, \langle \mathbf{y} \cdot \mathbf{y} \rangle$$

and the equality sign holds if and only if $\mathbf{x}$ *and* $\mathbf{y}$ *are dependent.*

Theorem B.18 *For* **x** *and* **y** *vectors of an inner product space X*

$$\begin{aligned} \|\mathbf{x}+\mathbf{y}\|^2 &= \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2\langle \mathbf{x}\cdot\mathbf{y}\rangle, \\ \|\mathbf{x}-\mathbf{y}\|^2 &= \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\langle \mathbf{x}\cdot\mathbf{y}\rangle. \end{aligned}$$

Definition B.19 The *angle* θ between two vectors **x** and **y** of an inner product space is defined by

$$\cos\theta = \frac{\langle \mathbf{x}\cdot\mathbf{y}\rangle}{\|\mathbf{x}\|\,\|\mathbf{y}\|}.$$

If $|\langle \mathbf{x}\cdot\mathbf{y}\rangle| = \|\mathbf{x}\|\,\|\mathbf{y}\|$, the cosine is 1, $\theta = 0$, and **x** and **y** are said to be *parallel*. If $\langle \mathbf{x}\cdot\mathbf{y}\rangle = 0$, the cosine is 0, $\theta = \frac{\pi}{2}$ and the vectors are said to be *orthogonal*.

Definition B.20 Given a set $S = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ of vectors from an inner product space X, the $n \times n$ matrix $\mathbf{G}$ with entries $\mathbf{G}_{ij} = \langle \mathbf{x}_i \cdot \mathbf{x}_j\rangle$ is called the *Gram matrix* of S.

B.3 Hilbert Spaces

Definition B.21 A space H is *separable* if there exists a countable subset $D \subseteq H$, such that every element of H is the limit of a sequence of elements of D. A *Hilbert space* is a complete separable inner product space.

Finite dimensional vector spaces such as $\mathbb{R}^n$ are Hilbert spaces.

Theorem B.22 *Let H be a Hilbert space, M a closed subspace of H, and* $\mathbf{x} \in H$. *There is a unique vector* $\mathbf{m}_0 \in M$, *known as the* projection of **x** onto *M*, *such that*

$$\|\mathbf{x} - \mathbf{m}_0\| \leq \inf\{\|\mathbf{x} - \mathbf{m}\| : \mathbf{m} \in M\}.$$

A necessary and sufficient condition for $\mathbf{m}_0 \in M$ *to be the projection of* **x** *onto M is that the vector* $\mathbf{x} - \mathbf{m}_0$ *be orthogonal to vectors in M.*

A consequence of this theorem is that the best approximation to **x** in the subspace M generated by the orthonormal vectors $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is given by its Fourier series

$$\sum_{i=1}^{n} \langle \mathbf{x}\cdot\mathbf{e}_i\rangle\,\mathbf{e}_i.$$

This leads naturally to studying the properties of series like this in the case of infinite bases.

Definition B.23 If S is an orthonormal set in a Hilbert space H and no other orthonormal set contains S as a proper subset, that is S is maximal, then S is called an *orthonormal basis* (or: a *complete orthonormal system*) for H.

Theorem B.24 *Every Hilbert space H has an orthonormal basis. Suppose that $S = \{\mathbf{x}_\alpha\}_{\alpha\in A}$ is an orthonormal basis for a Hilbert space H. Then, $\forall \mathbf{y} \in H$,*

$$\mathbf{y} = \sum_{\alpha\in A} \langle \mathbf{y} \cdot \mathbf{x}_\alpha \rangle \, \mathbf{x}_\alpha$$

and $\|\mathbf{y}\|^2 = \sum_{\alpha\in A} |\langle \mathbf{y} \cdot \mathbf{x}_\alpha \rangle|^2$.

This theorem states that, as in the finite dimensional case, every element of a Hilbert space can be expressed as a linear combination of (possibly infinite) basis elements.

The coefficients $\langle \mathbf{y} \cdot \mathbf{x}_\alpha \rangle$ are often called the Fourier coefficients of $\mathbf{y}$ with respect to the basis $S = \{\mathbf{x}_\alpha\}_{\alpha\in A}$.

Example B.25 Consider countable sequences of real numbers. The Hilbert space ℓ_2 is the set of sequences $\mathbf{z} = \{z_1, z_2, \dots, z_i, \dots\}$ such that

$$\|\mathbf{z}\|_2^2 = \sum_{i=1}^{\infty} z_i^2 < \infty,$$

where the inner product of sequences $\mathbf{x}$ and $\mathbf{z}$ is defined by

$$\langle \mathbf{x} \cdot \mathbf{z} \rangle = \sum_{i=1}^{\infty} x_i z_i.$$

If $\boldsymbol{\mu} = \{\mu_1, \mu_2, \dots, \mu_i, \dots\}$ is a countable sequence of positive real numbers, the Hilbert space $\ell_2(\boldsymbol{\mu})$ is the set of sequences $\mathbf{z} = \{z_1, z_2, \dots, z_i, \dots\}$ such that

$$\|\mathbf{z}\|_2^2 = \sum_{i=1}^{\infty} \mu_i z_i^2 < \infty,$$

where the inner product of sequences $\mathbf{x}$ and $\mathbf{z}$ is defined by

$$\langle \mathbf{x} \cdot \mathbf{z} \rangle = \sum_{i=1}^{\infty} \mu_i x_i z_i.$$

The normed space ℓ_1 is the set of sequences $\mathbf{z} = \{z_1, z_2, \dots, z_i, \dots\}$ for which

$$\|\mathbf{z}\|_1 = \sum_{i=1}^{\infty} |z_i| < \infty.$$

Example B.26 Consider the set of continuous real-valued functions on a subset X of $\mathbb{R}^n$. The Hilbert space $L_2(X)$ is the set of functions f for which

$$\|f\|_{L_2} = \int_X f(\mathbf{x})^2 d\mathbf{x} < \infty,$$

where the inner product of functions f and g is defined by

$$\langle f \cdot g \rangle = \int_X f(\mathbf{x}) g(\mathbf{x}) d\mathbf{x}.$$

The normed space $L_\infty(X)$ is the set of functions for which

$$\|f\|_{L_\infty} = \sup_{\mathbf{x} \in X} |f(\mathbf{x})| < \infty.$$

B.4 Operators, Eigenvalues and Eigenvectors

Definition B.27 A linear function from a Hilbert space H to itself is known as a *linear operator*. The linear operator A is *bounded* if there exists a number $\|A\|$ such that $\|Ax\| \leq \|A\| \, \|x\|$, for all $x \in H$.

Definition B.28 Let A be a linear operator on a Hilbert space H. If there is a vector, $0 \neq \mathbf{x} \in H$, such that $A\mathbf{x} = \lambda \mathbf{x}$ for some scalar λ, then λ is an *eigenvalue* of A with corresponding *eigenvector* $\mathbf{x}$.

Definition B.29 A bounded linear operator A on a Hilbert space H is *self-adjoint* if

$$\langle A\mathbf{x} \cdot \mathbf{z} \rangle = \langle \mathbf{x} \cdot A\mathbf{z} \rangle,$$

for all $\mathbf{x}$, $\mathbf{z} \in H$. For the finite dimensional space $\mathbb{R}^n$ this implies that the corresponding $n \times n$ matrix $\mathbf{A}$ satisfies $\mathbf{A} = \mathbf{A}'$, that is $\mathbf{A}_{ij} = \mathbf{A}_{ji}$. Such matrices are known as *symmetric*.

The following theorem involves a property known as compactness. We have omitted the definition of this property as it is quite involved and is not important for an understanding of the material contained in the book.

Theorem B.30 *(Hilbert Schmidt) Let A be a self-adjoint compact linear operator on a Hilbert space H. Then there is a complete orthonormal basis $\{\boldsymbol{\phi}_i\}_{i=1}^{\infty}$ for H such that*

$$A\boldsymbol{\phi}_i = \lambda_i \boldsymbol{\phi}_i,$$

and $\lambda_i \to 0$ as $i \to \infty$.

In the finite case the theorem states that symmetric matrices have an orthonormal set of eigenvectors.

Definition B.31 A square symmetric matrix is said to be *positive (semi-) definite* if its eigenvalues are all positive (non-negative).

Proposition B.32 *Let $\mathbf{A}$ be a symmetric matrix. Then $\mathbf{A}$ is positive (semi-) definite if and only if for any vector $\mathbf{x} \neq 0$*

$$\mathbf{x}'\mathbf{A}\mathbf{x} > 0 \; (\geq 0).$$

Let $\mathbf{M}$ be any (possibly non-square) matrix and set $\mathbf{A} = \mathbf{M}'\mathbf{M}$. Then $\mathbf{A}$ is a positive semi-definite matrix since we can write

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\mathbf{M}'\mathbf{M}\mathbf{x} = (\mathbf{M}\mathbf{x})'\,\mathbf{M}\mathbf{x} = \langle \mathbf{M}\mathbf{x} \cdot \mathbf{M}\mathbf{x} \rangle = \|\mathbf{M}\mathbf{x}\|^2 \geq 0,$$

for any vector $\mathbf{x}$. If we take $\mathbf{M}$ to be the matrix whose columns are the vectors $\mathbf{x}_i$, $i = 1, \ldots, n$, then $\mathbf{A}$ is the Gram matrix of the set $S = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, showing that Gram matrices are always positive semi-definite, and positive definite if the set S is linearly independent.