Appendix A

Finite-dimensional vector space

This appendix provides a catalog of many of the definitions and properties of vectors and finite-dimensional vector spaces. Our aim is to provide a ready reference to concepts and tools. Readers interested in the details of the proof of these results are urged to consult any of the classic references cited at the end of this appendix.

A.1 Definition and operations

Vectors Let $\mathbb{R}$ denote the set of all *real* numbers. By definition, a real vector $\mathbf{x}$ is an ordered m-tuple of real numbers $(x_1, x_2, \ldots, x_m)$ arranged vertically in a *column* as

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}.$$

The term x_i is called the ith *component* or the *coordinate* (with respect to the standard coordinate system) of the vector $\mathbf{x}$ of *size* or *order* or *dimension* m. Often to save space, $\mathbf{x}$ is written as the *transpose* of a *row*-vector $\mathbf{x} = (x_1, x_2, \ldots, x_m)^{\mathrm{T}}$, where T denotes the transpose operation. By convention, $\mathbb{R}^m$ denotes *the set of all m-dimensional real (column) vectors*. A vector all of whose components are zero is called a *zero* or *null* vector, and is denoted by $\mathbf{0}$. Any real vector of size one is called a real *scalar*. Likewise, if $\mathbb{C}$ denotes the set of all *complex* numbers, then $\mathbb{C}^m$ denotes the set of all complex vectors of size m.

Let $\mathbf{x} = (x_1, x_2, \ldots, x_m)^{\mathrm{T}}$, $\mathbf{y} = (y_1, y_2, \ldots, y_m)^{\mathrm{T}}$ and $\mathbf{z} = (z_1, z_2, \ldots, z_m)^{\mathrm{T}}$, and let a, b, c be real scalars.

Operations on vectors $\mathbf{z} = \mathbf{x} + \mathbf{y}$, where $z_i = x_i + y_i$ for $i = 1, \ldots, m$ is called the component-wise *sum* or *addition* of vectors. The vector *difference* is likewise defined as the component-wise difference. $\mathbf{z} = a\mathbf{x}$, where $z_i = ax_i$, for $i = 1, \ldots, m$, is called the *scalar multiplication* of a vector $\mathbf{x}$ by a real number a.

Linear vector space Let $\mathcal{V}$ denote a *set* or a *collection* of real vectors of size m. Then $\mathcal{V}$ is called a *(linear) vector space* if the following conditions are true:

(V1)

(a) $\mathbf{x} + \mathbf{y} \in \mathcal{V}$ whenever $\mathbf{x}, \mathbf{y} \in \mathcal{V}$ – ($\mathcal{V}$ is *closed* under addition).
(b) $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$ – (addition is *associative*).
(c) $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ – (addition is *commutative*).
(d) $\mathcal{V}$ contains the null vector $\mathbf{0}$ and $\mathbf{x} + \mathbf{0} = \mathbf{0} + \mathbf{x} = \mathbf{x}$.
(e) For every $\mathbf{x}$, there is a *unique* $\mathbf{y}$ in $\mathcal{V}$ such that $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x} = \mathbf{0}$. $\mathbf{y}$ is called the (additive) *inverse* of $\mathbf{x}$ and is denoted by $-\mathbf{x}$.

(V2)

(a) $a\mathbf{x}$ is in $\mathcal{V}$ if $\mathbf{x}$ is in $\mathcal{V}$ – ($\mathcal{V}$ is closed under *scalar* multiplication).
(b) $a(b\mathbf{x}) = (ab)\mathbf{x}$, for all $\mathbf{x} \in \mathcal{V}$, and $a, b \in \mathbb{R}$.
(c) $1(\mathbf{x}) = \mathbf{x}$, where 1 is the real number 1, for all $\mathbf{x} \in \mathcal{V}$.

(V3)

(a) $a(\mathbf{x} + \mathbf{y}) = a\mathbf{x} + a\mathbf{y}$ – *Distributivity*.
(b) $(a + b)\mathbf{x} = a\mathbf{x} + b\mathbf{x}$ – *Distributivity*.

Typical examples of (linear) vector spaces include $\mathbb{R}^n$ for every $n \geq 1$. Thus, for any n, $\mathbb{R}^n$ is called the *finite-dimensional vector space of dimension* n.

Another example of a finite-dimensional vector space include P_n = set of all polynomials with real coefficients and of degree less than n.

Remark A.1.1 The concept of vector spaces readily carries over to infinite dimensions. Let $x = \{x_0, x_1, x_2, \ldots\}$ be an infinite sequence. We say that x is square summable if

$$\sum_{i=0}^{\infty} x_i^2 < \infty.$$

The set of all square summable sequences is denoted by l_2. It can be verified that under component-wise addition and scalar multiplication, l_2 is a vector space. For other examples of infinite dimensional vector spaces and their properties refer to Kolmogorov and Fomin (1975).

Inner Product The *inner* or *scalar* product of two vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^m$, denoted by $< \mathbf{x}, \mathbf{y} >$ and $\mathbf{x}^{\mathrm{T}}\mathbf{y}$ is defined as

$$\begin{aligned} < \mathbf{x}, \mathbf{y} >= \mathbf{x}^{\mathrm{T}}\mathbf{y} &= \sum_{i=1}^{m} x_i y_i \\ &= \sum_{i=1}^{m} y_i x_i = \mathbf{y}^{\mathrm{T}}\mathbf{x} =< \mathbf{y}, \mathbf{x} > \end{aligned} \tag{A.1.1}$$

and satisfies the following properties:

(a) $< \mathbf{x}, \mathbf{x} > \begin{cases} > 0 & \text{if } \mathbf{x} \neq \mathbf{0} \\ = 0 & \text{if } \mathbf{x} = \mathbf{0} \end{cases}$ – (*Positive definite*).
(b) $< \mathbf{x}, \mathbf{y} >=< \mathbf{y}, \mathbf{x} >$ – (*Commutative*).
(c) $< \mathbf{x} + \mathbf{y}, \mathbf{z} >=< \mathbf{x}, \mathbf{z} > + < \mathbf{y}, \mathbf{z} >$ – (*Additivity*).
(d) $< a\mathbf{x}, \mathbf{y} >= a < \mathbf{x}, \mathbf{y} >=< \mathbf{x}, a\mathbf{y} >$ – (*Homogeneity*).

Remark A.1.1 When $\mathbf{x}$ and $\mathbf{y}$ are *complex vectors*, then the inner product is defined as $< \mathbf{x}, \mathbf{y} >= \sum_{i=1}^{m} x_i \bar{y}_i$, where $\bar{y}_i$ is the complex conjugate of y_i.

Let $\mathbf{x}, \mathbf{y}$ be two vectors in $\mathbb{R}^m$. Then, if $< \mathbf{x}, \mathbf{z} >=< \mathbf{y}, \mathbf{z} >$ for all $\mathbf{z} \in \mathbb{R}^m$, then $\mathbf{x} = \mathbf{y}$.

Outer Product The *outer product* of $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^m$, is an $m \times m$ matrix denoted by $\mathbf{x}\mathbf{y}^{\mathrm{T}}$ and is defined as:

$$\mathbf{x}\mathbf{y}^{\mathrm{T}} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} (y_1, y_2, \ldots, y_m) = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \ldots & x_1 y_m \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_m \\ \vdots & \vdots & \vdots & \vdots \\ x_m y_1 & x_m y_2 & \cdots & x_m y_m \end{bmatrix}$$

$$[\mathbf{x}y_1, \mathbf{x}y_2, \ldots, \mathbf{x}y_m] = \begin{bmatrix} x_1 \mathbf{y}^{\mathrm{T}} \\ x_2 \mathbf{y}^{\mathrm{T}} \\ \vdots \\ x_m \mathbf{y}^{\mathrm{T}} \end{bmatrix}.$$

Thus, $\mathbf{x}\mathbf{y}^{\mathrm{T}}$ is a matrix (Appendix B) whose jth column is obtained by multiplying the column vector $\mathbf{x}$ by the scalar y_j for $j = 1, \ldots, m$. Similarly, the ith row if $\mathbf{x}\mathbf{y}^{\mathrm{T}}$ is obtained by multiplying the row vector $\mathbf{y}^{\mathrm{T}}$ by the scalars x_i, for $i = 1, \ldots, m$.

A.2 Norm and distance

The *norm* of a vector $\mathbf{x} \in \mathbb{R}^m$, denoted by $\|\mathbf{x}\|$ is a real number that denotes the *size* or the length of the vector $\mathbf{x}$. There are several ways to specify this norm.

(a) *Eucledian or 2-norm* $\|\mathbf{x}\|_2 = [\sum_{i=1}^{m} x_i^2]^{1/2}$.
(b) *Manhattan or 1-norm* $\|\mathbf{x}\|_1 = \sum_{i=1}^{m} |x_i|$.
(c) *Chebyshev or* ∞*-norm* $\|\mathbf{x}\|_\infty = \max_{1 \leq i \leq n} |x_i|$.
(d) *Minkowski or p-norm* $\|\mathbf{x}\|_p = [\sum_{i=1}^{m} |x_i|^p]^{1/p}$ for any $p > 0$
(e) *Energy norm* $\|\mathbf{x}\|_{\mathbf{A}} =< \mathbf{x}, \mathbf{A}\mathbf{x} >^{1/2}= (\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x})^{1/2}$, where $\mathbf{A} \in \mathcal{R}^{m \times m}$ is a *real, symmetric* and *positive definite* matrix.

The *distance* between two vectors $\mathbf{x}$ and $\mathbf{y}$, denoted by $d(\mathbf{x}, \mathbf{y})$ is defined to be the *norm of the difference*, $(\mathbf{x} - \mathbf{y})$, that is,

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|. \tag{A.2.1}$$

Table A.2.1 *FLOP ratings of basic operations on vectors*

Operation	Expression	Number of Operations (FLOPs)
Vector sum/difference	$\mathbf{z} = \mathbf{x} \pm \mathbf{y}$	m
Scalar times a vector	$\mathbf{z} = a\mathbf{x}$	m
Inner product	$< \mathbf{x}, \mathbf{y} >= \mathbf{x}^{\mathrm{T}}\mathbf{y}$	$2m - 1$
Outer product	$\mathbf{x}\mathbf{y}^{\mathrm{T}}$	m^2
2-norm squared	$\|\mathbf{x}\|_2^2 =< \mathbf{x}, \mathbf{x} >= \mathbf{x}^{\mathrm{T}}\mathbf{x}$	$2m - 1$

Thus, we can define the Euclidean, Manhattan, Chebyshev, Minkowski, or Energy distance by using the respective norm in (A.2.1).

Every vector norm must satisfy the following conditions:

(a) $\|\mathbf{x}\| \begin{cases} > 0 & \text{when } \mathbf{x} \neq \mathbf{0} \\ = 0 & \text{when } \mathbf{x} = \mathbf{0} \end{cases}$ – (*Positive definite*).
(b) $\|a\mathbf{x}\| = |a|\,\|x\|$ – (*Homogeneity*).
(c) $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ – (*Triangle inequality*).

It can be verified that 2-norm is *derivable* from the inner product in the sense

$$\|\mathbf{x}\|_2^2 =< \mathbf{x}, \mathbf{x} >= \mathbf{x}^{\mathrm{T}}\mathbf{x},$$

and the 2-norm satisfies the *parallelogram identity*

$$\|\mathbf{x} + \mathbf{y}\|_2^2 + \|\mathbf{x} - \mathbf{y}\|_2^2 = 2(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2).$$

Remark A.2.1 The amount of *work* required, measured in terms of the number of basic (floating point) operations (FLOP) in various vector operations are summarized in Table A.2, where it is assumed that the basic operations – *add, subtract, multiply, divide* and *compare*, take an *equal* amount of time, which is taken as the unit of time. This unit cost model for computation time is a good approximation to reality and simplifies the quantification of work required by various algorithms.

Unit Sphere The set S of all vectors of length one, that is,

$$S = \{\mathbf{x} \in \mathbb{R}^m \,|\, \|\mathbf{x}\| = 1\}$$

is called a *unit sphere*. The distinction between the various vector norms can be understood by sketching the surface of the unit sphere in $\mathbb{R}^2$. Refer to Figure A.2.1. We invite the reader to draw the unit sphere in $\mathbb{R}^2$ using $p = 4$ for $\|\mathbf{x}\|_p$.

Remark A.2.2 Any vector space endowed with a norm is called a *normed vector space*. If the norm is derivable from the inner product, then it is called the *Euclidean space*. Hence, $\mathbb{R}^n$ with 2-norm has often come to be known as the *finite-dimensional Euclidean space*. It can be verified that the p-norm, for $p \neq 2$ are *not* derivable or related to the inner product.

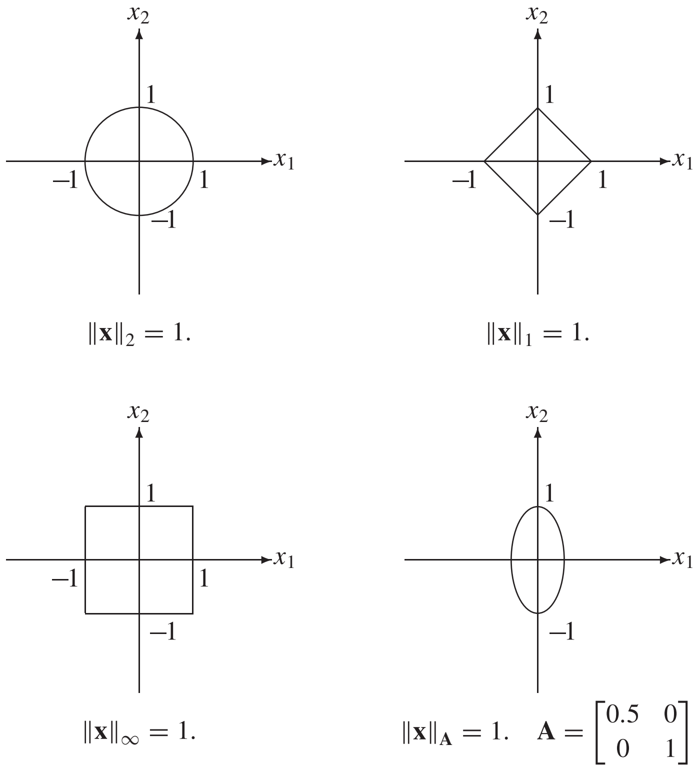

Fig. A.2.1 Unit spheres in $\mathbb{R}^2$ under various norms.

Unit vector/direction cosines If $\mathbf{x} \in \mathbb{R}^m$, then the *unit vector* in the direction of $\mathbf{x}$ is given by

$$\hat{\mathbf{x}} = \frac{\mathbf{x}}{\|\mathbf{x}\|_2} = (\hat{x_1}, \hat{x_2}, \ldots, \hat{x_m})^{\mathrm{T}}$$

and the components $\hat{x_i}$ are called the *direction cosines* – the cosine of the angle made by the vector $\mathbf{x}$ with respect to the standard ith coordinate axis, for $i = 1, \ldots, m$. It can be verified that $\|\hat{\mathbf{x}}\|_2 = 1$.

Fundamental inequalities Let θ be the angle between two vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^m$. Then

$$< \mathbf{x}, \mathbf{y} >= \mathbf{x}^{\mathrm{T}}\mathbf{y} = \|\mathbf{x}\|_2 \, \|\mathbf{y}\|_2 \cos\theta \leq \|\mathbf{x}\|_2 \, \|\mathbf{y}\|_2$$

is a very useful fact, called the *Bunyakowski–Cauchy–Schwartz* (BCS) inequality.

If p and q are such that $1/p + 1/q = 1$, then

$$\mathbf{x}^{\mathrm{T}}\mathbf{y} \leq \|\mathbf{x}\|_p \, \|\mathbf{y}\|_q$$

is called the *Minkowski inequality*. Clearly, BCS inequality is the special case when $p = q = 2$.

Equivalent norms Two norms are said to be *equivalent* if the length of a vector $\mathbf{x}$ is finite in one norm then it is also finite in the other norm. The 2-norm, 1-norm, and the ∞-norm are equivalent, and they are related as follows.

$$\begin{aligned} \|\mathbf{x}\|_2 &\leq \|\mathbf{x}\|_1 \leq \sqrt{n}\,\|\mathbf{x}\|_2 \\ \|\mathbf{x}\|_\infty &\leq \|\mathbf{x}\|_2 \leq \sqrt{n}\,\|\mathbf{x}\|_\infty \\ \|\mathbf{x}\|_\infty &\leq \|\mathbf{x}\|_1 \leq n\,\|\mathbf{x}\|_\infty . \end{aligned}$$

Functionals Let $\mathcal{V}$ be a vector space. Any scalar-valued function f that maps $\mathcal{V}$ into the real line, that is, $f : \mathcal{V} \to \mathbb{R}$ is called a **functional**. In addition, if it satisfies two conditions:

$$f(\mathbf{x}_1 + \mathbf{x}_2) = f(\mathbf{x}_1) + f(\mathbf{x}_2) \quad \text{– additivity}$$

and

$$f(a\mathbf{x}) = af(\mathbf{x}),\ a \in \mathbb{R} \quad \text{– homogeneity}$$

then f is called a **linear functional**. Otherwise, it is a **nonlinear** functional.

For any $\mathbf{a} \in \mathbb{R}^n$ fixed, $f_1(\mathbf{x}) =< \mathbf{a}, \mathbf{x} >$ denoting the inner product is a linear functional. $f_2(\mathbf{x}) = \frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$ with $\mathbf{A}$ being $n \times n$ symmetric matrix which is a quadratic form is an example of a nonlinear functional.

A.3 Orthogonality

Two vectors $\mathbf{x}$ and $\mathbf{y}$ are said to be *orthogonal*, denoted by $\mathbf{x}\perp\mathbf{y}$, if their inner product is zero,

$$\mathbf{x}\perp\mathbf{y} \Leftrightarrow < \mathbf{x}, \mathbf{y} >= \mathbf{x}^{\mathrm{T}}\mathbf{y} = 0.$$

A set $S = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$ on n vectors in $\mathbb{R}^m$ are said to be *mutually orthogonal* if they are pairwise orthogonal.

$$< \mathbf{x}_i, \mathbf{x}_j >= \mathbf{x}_i^T\mathbf{x}_j = \begin{cases} 0 & \text{for } i \neq j \\ \|\mathbf{x}_i\|_2\,, & \text{for } i = j \end{cases}$$

In addition, if the vectors in S are also normalized to have unit length, then S is called the *orthonormal* set.

Conjugacy Two vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^m$ are said to be $\mathbf{A}$-*conjugate* if

$$< \mathbf{x}, \mathbf{A}\mathbf{y} >= \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{y} = 0$$

where $\mathbf{A} \in \mathbb{R}^{m\times m}$ is a *symmetric positive definite matrix*. Similarly, a set of n vectors, $S = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$, $\mathbf{x}_i \in \mathbb{R}^m$, is said to be $\mathbf{A}$-conjugate if

$$< \mathbf{x}_i, \mathbf{A}\mathbf{x}_j >= \begin{cases} 0 & \text{for } i \neq j \\ \mathbf{x}_i^{\mathrm{T}}\mathbf{A}\mathbf{x}_i = \|\mathbf{x}_i\|_{\mathbf{A}}^2, & \text{for } i = j \end{cases}$$

A.4 Linear combination and subspace

Let $S = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\}$ be a set of n vectors in $\mathbb{R}^m$, where $\mathbf{x}_i = (x_{i1}, x_{i2}, \cdots, x_{im})^{\mathrm{T}}$, and let $a_1, a_2, \ldots, a_n$ be scalars. The vector $\mathbf{y}$ defined as the sum of the scalar multiples of the vectors in S

$$\mathbf{y} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_n\mathbf{x}_n \tag{A.4.1}$$

is called a *linear combination* of vectors in S. When $a_i = 1/n$, then the linear combination

$$\mathbf{y} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$$

is called the *centroid* of S.

Linear dependence/independence If the linear combination in A.4.1 is zero when *not all* the scalars a_i's are zero, then the vectors in S are said to be *linearly dependent*. On the other hand, if $\mathbf{y}$ is zero only when all the scalars a_i's in A.4.1 are zero, then the vectors in S are said to be *linearly independent*.

It can be verified that if the vectors in the set S are mutually orthogonal, then they are independent. The converse is not true.

Span and Subspace Given S,

$$\mathrm{Span}(S) = \left\{ \mathbf{y} | \mathbf{y} = \sum_{i=1}^{n} a_i\mathbf{x}_i, a_i \in \mathbb{R}^m \right\}$$

denotes the set of *all* linear combination of vectors in S. It can be verified that the null vector $\mathbf{0}$ is in Span(S). Span(S) is a (linear) vector space generated by vectors in S and is called the *subspace* of $\mathbb{R}^m$, that is, $\mathrm{Span}(S) \subseteq \mathbb{R}^m$. The set $\{\mathbf{0}\}$ consisting of the null vector is a subspace, called the *trivial subspace*.

Basis and Dimension Let $\mathcal{V}$ be a vector space and S be a subset of linearly independent vectors in $\mathcal{V}$. If every vector in $\mathcal{V}$ can be (uniquely) expressed as a linear combination of those in S, then this set S is called a *basis* for $\mathcal{V}$. The number of linearly independent vectors in S is called the *dimension* of $\mathcal{V}$, and is denoted by Dim($\mathcal{V}$). If the members of S are orthogonal (orthonormal), then it is called an orthogonal (orthonormal) basis for $\mathcal{V}$. It can be verified that for given $\mathcal{V}$, the basis is *not* unique. $S_1 = \{(1, 0)^{\mathrm{T}}, (0, 1)^{\mathrm{T}}\}$ and $S_2 = \{(1, 0.5)^{\mathrm{T}}, (1, 1)^{\mathrm{T}}\}$ are two distinct basis for $\mathbb{R}^2$, since $\mathbb{R}^2 = \text{Span}(S_1) = \text{Span}(S_2)$. S_1 is also an orthonormal basis for $\mathbb{R}^2$.

Standard Unit Vectors Let $\mathbf{e}_i = (0, 0, \ldots, 1, 0, \ldots, 0)^{\mathrm{T}} \in \mathbb{R}^m$ be a vector with 1 as its ith element. The set

$$S = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_m\}$$

is the standard basis for $\mathbb{R}^m$. Any vector $\mathbf{x} = (x_1, x_2, \ldots, x_m)^{\mathrm{T}}$ is given by

$$\mathbf{x} = \{x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_m\mathbf{e}_m\}$$

and x_i is called the ith coordinate of $\mathbf{x}$.

Direct sum and orthogonal complement Let $\mathcal{V}_1$ and $\mathcal{V}_2$ be two vector spaces. If every vector in $\mathcal{V}_1$ is also a vector in $\mathcal{V}_2$, but *not vice versa*, then $\mathcal{V}_1$ is called the *proper subspace* of $\mathcal{V}_2$, denoted by $\mathcal{V}_1 \subset \mathcal{V}_2$.
Let $\mathcal{V}_1 \cup \mathcal{V}_2$ and $\mathcal{V}_1 \cap \mathcal{V}_2$ denote the set *union* and set *intersection* of $\mathcal{V}_1$ and $\mathcal{V}_2$. If $\text{Dim}(\mathcal{V}_i) = k_i$, $i = 1, 2$, then

$$\text{Dim}(\mathcal{V}_1 \cup \mathcal{V}_2) = \text{Dim}(\mathcal{V}_1) + \text{Dim}(\mathcal{V}_2) - \text{Dim}(\mathcal{V}_1 \cap \mathcal{V}_2). \tag{A.4.2}$$

Let $\mathcal{V}$ be a vector space and $\mathcal{S} \subset \mathcal{V}$ be a subspace. Let $\mathcal{S}^{\perp}$ denote the set of all vectors in $\mathcal{V}$ that are orthogonal to those in $\mathcal{S}$. The set $\mathcal{S}^{\perp}$ is called the *orthogonal complement* of $\mathcal{S}$ in $\mathcal{V}$ and

$$\mathcal{V} = \mathcal{S} \oplus \mathcal{S}^{\perp}$$

called the *direct sum* of $\mathcal{S}$ and $\mathcal{S}^{\perp}$. It can be verified that $\mathcal{S} \cap \mathcal{S}^{\perp} = \{\mathbf{0}\}$ and $\text{Dim}(\mathcal{V}) = \text{Dim}(\mathcal{S}) + \text{Dim}(\mathcal{S}^{\perp})$.

Example A.4.1 Let $\mathcal{V} = \mathbb{R}^3$, and $\mathbf{x} = (1, 0, 0)^{\mathrm{T}}$, and $\mathbf{y} = (1, -1, 0)^{\mathrm{T}}$. Let $\mathcal{S} = \text{Span}(\{\mathbf{x}, \mathbf{y}\}) = \mathbb{R}^2$. Then $\mathcal{S}^{\perp} = \text{Span}(\mathbf{z})$, where $\mathbf{z} = (0, 0, 1)^{\mathrm{T}}$, is the orthogonal complement of $\text{Span}(\{\mathbf{x}, \mathbf{y}\})$, and $\text{Dim}(\mathcal{S}) = 2$, and $\text{Dim}(\mathcal{S}^{\perp}) = 1$, with $\mathcal{S} \cap \mathcal{S}^{\perp} = \{\mathbf{0}\}$.

Completion Theorem Given a vector space $\mathcal{V}$ of dimension m. Let $\mathcal{S}_1 = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ be a set of n linearly independent vectors with $n < m$. Then there exists a set $\mathcal{S}_2 = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{m-n}\}$ of linearly independent vectors (distinct from $\mathcal{S}_1$), such that $\mathcal{S} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n, \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{m-n}\}$ is linearly independent and constitutes a basis for $\mathcal{V}$.

A.5 Projection of a vector

Let $\mathbf{h} \in \mathbb{R}^m$ be a *unit vector*. Given any vector $\mathbf{z} \in \mathbb{R}^m$, the inner product of $\mathbf{z}$ with $\mathbf{h}$, that is $\mathbf{z}^{\mathrm{T}}\mathbf{h}$, is called the magnitude of the *projection* of $\mathbf{z}$ onto the direction of

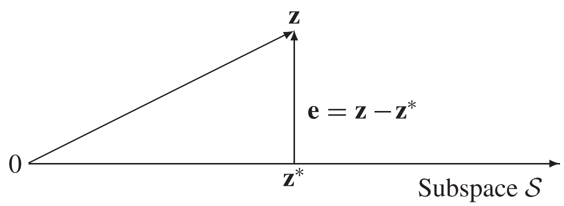

Fig. A.5.1 An illustration of projection theorem: $\mathbf{z} = \mathbf{z}^* + \mathbf{e}$ and $\mathbf{z}^* \perp \mathbf{e}$.

$\mathbf{h}$. The vector $(\mathbf{z}^{\mathrm{T}}\mathbf{h})\mathbf{h}$ is the representation of $\mathbf{z}$ in the direction $\mathbf{h}$ and is called the *projection of* $\mathbf{z}$ *onto* $\mathbf{h}$.

Bessel's Inequality Let $\mathcal{S} = \{\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_n\}$ be an orthonormal set of vectors in $\mathbb{R}^m$. Let $\mathbf{z} \in \mathbb{R}^m$. Then $\mathbf{z}^{\mathrm{T}}\mathbf{h}_i$ denotes the magnitude of the projection of $\mathbf{z}$ onto the direction $\mathbf{h}_i$. Then,

$$\mathbf{z}^* = (\mathbf{z}^{\mathrm{T}}\mathbf{h}_1)\mathbf{h}_1 + (\mathbf{z}^{\mathrm{T}}\mathbf{h}_2)\mathbf{h}_2 + \cdots + (\mathbf{z}^{\mathrm{T}}\mathbf{h}_m)\mathbf{h}_n$$

is a representation of $\mathbf{z}$ in S, called the projection of $\mathbf{z}$ onto $\mathcal{S}$. It can be verified that $(\mathbf{z} - \mathbf{z}^*)$ is *orthogonal* to S, and

$$\sum_{i=1}^{n} (\mathbf{z}^{\mathrm{T}}\mathbf{h}_i)^2 \leq \|\mathbf{z}\|^2 \tag{A.5.1}$$

is called *Bessel's inequality*.

Parseval's Identity If the orthonormal set $\mathcal{S}$ is such that A.5.1 holds with equality

$$\sum_{i=1}^{n} (\mathbf{z}^{\mathrm{T}}\mathbf{h}_i)^2 = \|\mathbf{z}\|^2 \tag{A.5.2}$$

then $\mathcal{S}$ is called the *complete* orthonormal set and A.5.2 is called *Parseval's identity*.

Remark A.5.1 The representation of $\mathbf{z}$ in any complete orthonormal basis $\mathcal{S} = \{\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_m\}$, as

$$\mathbf{z} = \sum_{i=1}^{m} (\mathbf{z}^{\mathrm{T}}\mathbf{h}_i)\mathbf{h}_i \tag{A.5.3}$$

is called the *Fourier* expansion of $\mathbf{z}$ and the inner products $(\mathbf{z}^{\mathrm{T}}\mathbf{h}_i)$ are called the *Fourier coefficients*. The right-hand side in (A.5.3) is called the *spectral* expansion of $\mathbf{z}$ and Parseval's identity states that the total energy in either representation is the same when $\mathcal{S}$ is a complete orthonormal basis.

Orthogonal Projection Theorem Let $\mathcal{V}$ be a vector space and let $\mathcal{S}$ be a subspace of $\mathcal{V}$. For any vector $\mathbf{z} \in \mathcal{V}$, a vector $\mathbf{z}^* \in \mathcal{S}$ minimizes $\|\mathbf{z} - \mathbf{z}^*\|_2$ exactly when

$\mathbf{z} - \mathbf{z}^*$ is orthogonal to $\mathcal{S}$. Then, $\mathbf{z}^*$ is called the *orthogonal projection* of $\mathbf{z}$ onto $\mathcal{S}$. See Figure A.5.1.

Notes and references

There are several excellent textbooks on the topic of this appendix. Halmos (1958) is undoubtedly a classic in this area. Stewart (1973) provides an elementary treatment of these topics. Basilevsky (1983) and Meyer (2000) are suitable for first year graduate level.