SOME BASIC FACTS FROM LINEAR ALGEBRA

$\mathbb{N}$	denotes the set of natural numbers $1, 2, 3, \ldots,$
$\mathbb{R}$	denotes the set of real numbers,
$\mathbb{C}$	denotes the set of complex numbers.

1 Vector Spaces

Definition 1. A *vector space* (or a *linear space*) X over a *field* $\mathfrak{F}$ (the elements of $\mathfrak{F}$ are called *scalars*) is a set of elements called *vectors* equipped with two (binary) operations, namely *vector addition* (the sum of two vectors $\mathbf{x}, \mathbf{y} \in X$ is denoted by $\mathbf{x}+\mathbf{y}$) and scalar multiplication (the scalar product of a scalar $a \in \mathfrak{F}$ and a vector $\mathbf{x} \in X$ is usually denoted by ax; the notation xa is rare) satisfying the following postulates:
1 (Closure)**.** If $\mathbf{x}, \mathbf{y} \in X$ and $a \in \mathfrak{F}$, then $\mathbf{x}+\mathbf{y} \in X$ and $a\mathbf{x} \in X$.
2 (Associativity for $+$)**.** $(\mathbf{x}+\mathbf{y})+\mathbf{z} = \mathbf{x}+(\mathbf{y}+\mathbf{z})$, for all $\mathbf{x}, \mathbf{y}, \mathbf{z}$ in X.
3 (Commutativity for $+$)**.** $\mathbf{x}+\mathbf{y} = \mathbf{y}+\mathbf{x}$, for all $\mathbf{x}, \mathbf{y}$ in X.
4 (Identity for Addition)**.** There is a vector $\mathbf{O} \in X$ such that $\mathbf{x}+\mathbf{O} = \mathbf{x}$, for all $\mathbf{x} \in X$.
5 (Additive Inverse)**.** For any $\mathbf{x} \in X$ there is a vector in X, denoted by $-\mathbf{x}$, such that $\mathbf{x}+(-\mathbf{x}) = \mathbf{O}$.
6 ("Associativity" for Scalar Multiplication)**.** $a(b\mathbf{x}) = (ab)\mathbf{x}$, for all $\mathbf{x} \in X$, $a, b \in \mathfrak{F}$.
7 (Identity for Scalar Multiplication)**.** For any $\mathbf{x} \in X$ we have that $1\mathbf{x} = \mathbf{x}$.
8 (Distributive Laws)**.** $a(\mathbf{x}+\mathbf{y}) = a\mathbf{x}+a\mathbf{y}$ and $(a+b)\mathbf{x} = a\mathbf{x}+b\mathbf{x}$, for all $\mathbf{x}, \mathbf{y} \in X$ and $a, b \in \mathfrak{F}$.

The sum $\mathbf{x}+(-\mathbf{y})$ is denoted by $\mathbf{x}-\mathbf{y}$. For a scalar $a \neq 0$ we can also write $\mathbf{x}/a$ instead of $a^{-1}\mathbf{x}$.

The proof of the following proposition is left as an exercise:

Proposition 1. Suppose X is a vector space. Then
9 (Cancellation Law)**.** Let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in X$. If $\mathbf{x}+\mathbf{y} = \mathbf{x}+\mathbf{z}$, then $\mathbf{y} = \mathbf{z}$.
10 (Uniqueness of $\mathbf{O}$ and $-\mathbf{x}$)**.** If $\mathbf{x}+\mathbf{O}' = \mathbf{x}$, for some $\mathbf{x} \in X$, then $\mathbf{O}' = \mathbf{O}$; if, for some $\mathbf{x} \in X$, $\mathbf{x}+\mathbf{z} = \mathbf{O}$, then $\mathbf{z} = -\mathbf{x}$.
11. $0\mathbf{x} = \mathbf{O}$, for all $\mathbf{x} \in X$.
12. $(-1)\mathbf{x} = -\mathbf{x}$, for all $\mathbf{x} \in X$.

We notice that Postulate 3 (the commutativity for vector addition) is redundant. Actually the distributive laws and Postulate 7 yield

$$(1+1)(\mathbf{x}+\mathbf{y}) = (1+1)\mathbf{x} + (1+1)\mathbf{y} = \mathbf{x}+\mathbf{x}+\mathbf{y}+\mathbf{y}$$

and also

$$(1+1)(\mathbf{x}+\mathbf{y}) = 1(\mathbf{x}+\mathbf{y}) + 1(\mathbf{x}+\mathbf{y}) = \mathbf{x}+\mathbf{y}+\mathbf{x}+\mathbf{y}.$$

Hence, the cancellation law implies $\mathbf{x}+\mathbf{y} = \mathbf{y}+\mathbf{x}$ (!)

In what follows the field of scalars $\mathfrak{F}$ will always be either the field of real numbers $\mathbb{R}$ or the field of complex numbers $\mathbb{C}$.

Example 1 (examples of vector spaces)**.** (i) $\mathbb{R}^n = \{(x_1, \ldots, x_n) : x_j \in \mathbb{R}\}$ over $\mathbb{R}$, and $\mathbb{C}^n = \{(z_1, \ldots, z_n) : z_j \in \mathbb{C}\}$ over $\mathbb{C}$ (conventions: $\mathbb{R}^0 = \mathbb{C}^0 = \{0\}$, $\mathbb{R}^1 = \mathbb{R}$, $\mathbb{C}^1 = \mathbb{C}$). Notice, also, that $\mathbb{C}$ is a vector space over $\mathbb{R}$.
(ii) The set of all polynomials $p(x)$ of degree $\leq n$ with complex coefficients (convention: $\deg 0 = -\infty$). The field of scalars is $\mathbb{C}$.
(iii) The set $C[a,b]$ of all continuous complex-valued functions $f(x)$, $x \in [a,b]$. The field of scalars is $\mathbb{C}$.

Definition 2. A *subspace* S of a vector space X is a set such that:
(i) $S \subset X$;
(ii) if $\mathbf{x}, \mathbf{y} \in S$, then $a\mathbf{x} + b\mathbf{y} \in S$, for all scalars a, b.

Notice that a subspace of a vector space is itself a vector space over the same field of scalars. Given a vector space X, the spaces X and $\{\mathbf{O}\}$ are the *trivial subspaces* of X. Any other subspace of X is a *proper subspace*.
If $\{S_\alpha\}_{\alpha \in J}$ is any family of subspaces of a vector space X, then the intersection $\bigcap_{\alpha \in J} S_\alpha$ is also a subspace of X (convention: an empty intersection equals X).

Definition 3. Suppose $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are vector of X. The set of all linear combinations $c_1\mathbf{x}_1 + \cdots + c_k\mathbf{x}_k$ is called the *linear span of* $\mathbf{x}_1, \ldots, \mathbf{x}_k$ and it is denoted by $\langle \mathbf{x}_1, \ldots, \mathbf{x}_k \rangle$ (convention: $\langle \varnothing \rangle = \{\mathbf{O}\}$).

Notice that $\langle \mathbf{x}_1, \ldots, \mathbf{x}_k \rangle$ is a subspace of X; in fact, this is a way of constructing subspaces.

Definition 4. A family of vectors $\{\mathbf{x}_\alpha\}_{\alpha \in J}$ in a vector space X is said to be *linearly independent* when each relation of the form

$$c_1\mathbf{x}_{\alpha_1} + \cdots + c_n\mathbf{x}_{\alpha_n} = \mathbf{O}, \qquad \mathbf{x}_{\alpha_j} \in \{\mathbf{x}_\alpha\}_{\alpha \in J}$$

(where the c_j's are scalars), implies

$$c_1 = \cdots = c_n = 0.$$

If the family $\{\mathbf{x}_\alpha\}_{\alpha\in J}$ is not linearly independent, then it is called *linearly dependent*.
If the family $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is linearly (in)dependent, we say that the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are linearly (in)dependent.

Notice that if $\mathbf{O} \in \{\mathbf{x}_\alpha\}_{\alpha\in J}$, then $\{\mathbf{x}_\alpha\}_{\alpha\in J}$ is linearly dependent. The family $\{\mathbf{x}\}$ consisting of just one vector $\mathbf{x} \neq \mathbf{O}$ is linearly independent.
If $\mathbf{x}, \mathbf{y}$ are linearly dependent, then $\mathbf{y} = c\mathbf{x}$, for some scalar c, or $\mathbf{x} = c'\mathbf{y}$, for some scalar c'.

Example 2. In the vector space of continuous functions $C[0,1]$ the set $\{e^x, 1, x, x^2, \ldots, x^n, \ldots\}$, where

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!},$$

is linearly independent.

Definition 5. The *dimension* of a vector space X, written $\dim X$, is the largest number of linearly independent vectors in X, if that number is finite (hence, in particular we have $\dim\{\mathbf{O}\} = 0$). The dimension of X is said to be infinite ($\dim X = \infty$) if there exist linearly independent vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ in X, for arbitrarily large n.

$$\begin{aligned} \dim < \infty &\quad\longleftrightarrow\quad \text{Linear Algebra}, \\ \dim = \infty &\quad\longleftrightarrow\quad \text{Functional Analysis}. \end{aligned}$$

Example 3. $\dim \mathbb{R}^n = n$, $\dim \mathbb{C}^n = n$, $\dim C[0,1] = \infty$. Notice that $\mathbb{C}^n$ viewed as a vector space over $\mathbb{R}$ has dimension $2n$. Unless otherwise stated $\mathbb{C}^n$ is considered a vector space over $\mathbb{C}$.

Definition 6. A set of vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is said to be a *basis* for X, if every vector $\mathbf{x}$ in X can be written uniquely as

$$\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n, \qquad x_j \in \mathfrak{F}.$$

If the above equation holds, then we say that $\mathbf{x}$ is represented by the column (vector)

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

with respect to the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$.

Remark 1. If X has a basis of n vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$, then these basis vectors are linearly independent. This is because $\mathbf{O} = 0\mathbf{e}_1 + \cdots + 0\mathbf{e}_n$ uniquely.

The next two theorems follow immediately from the previous discussion.

Theorem 1. If $\dim X = n$ and $\mathbf{e}_1, \ldots, \mathbf{e}_n$ are linearly independent vectors, then $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is a basis.

Theorem 2. If $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is a basis for X, then $\dim X = n$.

CAUTION. If $\dim X = \infty$, then the notion of a basis of X becomes tricky!

Example 4. (i) The vectors

$$\mathbf{e}_1 = (1, 0, ..., 0), \mathbf{e}_2 = (0, 1, ..., 0), \ldots, \mathbf{e}_n = (0, 0, ..., 1)$$

form a basis for both $\mathbb{R}^n$ and $\mathbb{C}^n$.
(ii) Let X be the vector space of polynomials of degree $\leq n$ (with real or complex coefficients). Then the polynomials

$$1, x, x^2, \ldots, x^n$$

form a basis for X and hence $\dim X = n + 1$.

2 Linear Operators

Definition 7. Let X and Y be vector spaces over the same field $\mathfrak{F}$. A *linear operator* $\mathcal{L}$ from X to Y is a function $\mathcal{L} : X \to Y$ such that
(i) $\mathcal{L}(\mathbf{u} + \mathbf{v}) = \mathcal{L}\mathbf{u} + \mathcal{L}\mathbf{v}$, for all $\mathbf{u}, \mathbf{v} \in X$ and
(ii) $\mathcal{L}(a\mathbf{u}) = a\mathcal{L}\mathbf{u}$, for all $\mathbf{u} \in X$, $a \in \mathfrak{F}$.
In many cases we have $Y = X$; then we say that $\mathcal{L}$ is a linear operator (acting) on X. An important example is the *identity operator* on X, namely the operator $\mathcal{I}$ which satisfies $\mathcal{I}\mathbf{x} = \mathbf{x}$ for all $\mathbf{x} \in X$.

Assume $n = \dim X$ and $m = \dim Y$. Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be a basis for X and $\mathbf{f}_1, \ldots, \mathbf{f}_m$ be a basis for Y. Suppose

$$\begin{aligned}
\mathcal{L}\mathbf{e}_1 &= a_{11}\mathbf{f}_1 + a_{21}\mathbf{f}_2 + \cdots + a_{m1}\mathbf{f}_m \\
\mathcal{L}\mathbf{e}_2 &= a_{12}\mathbf{f}_1 + a_{22}\mathbf{f}_2 + \cdots + a_{m2}\mathbf{f}_m \\
&\vdots \\
\mathcal{L}\mathbf{e}_1 &= a_{1n}\mathbf{f}_1 + a_{2n}\mathbf{f}_2 + \cdots + a_{mn}\mathbf{f}_m,
\end{aligned}$$

then we say that the *matrix* of $\mathcal{L}$ with respect to the bases $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and $\mathbf{f}_1, \ldots, \mathbf{f}_m$ is the $m \times n$ array

$$[\mathcal{L}]_{\mathbf{e},\mathbf{f}} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

Thus, if $\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$, then the action of $\mathcal{L}$ on $\mathbf{x}$ is described by the matrix multiplication of $[\mathcal{L}]_{\mathbf{e},\mathbf{f}}$ by the column

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

Furthermore, if $\mathcal{M} : Y \to Z$ is another linear operator, and $\mathbf{g}_1, \ldots, \mathbf{g}_p$ is a basis for the vector space Z (hence $\dim Z = p$), then the $p \times n$ matrix of the composition $\mathcal{ML}$ ($:= \mathcal{M} \circ \mathcal{L}$) with respect to the bases $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and $\mathbf{g}_1, \ldots, \mathbf{g}_p$ satisfies

$$[\mathcal{ML}]_{\mathbf{e},\mathbf{g}} = [\mathcal{M}]_{\mathbf{f},\mathbf{g}} [\mathcal{L}]_{\mathbf{e},\mathbf{f}},$$

where the right-hand side is the standard matrix product of the matrices $[\mathcal{M}]_{\mathbf{f},\mathbf{g}}$ and $[\mathcal{L}]_{\mathbf{e},\mathbf{f}}$.

3 Eigenvalues and Eigenvectors

Definition 8. Let $\mathcal{L}$ be a linear operator on a vector space X over $\mathbb{C}$ and $\mathbf{v} \neq \mathbf{O}$ a vector such that

$$\mathcal{L}\mathbf{v} = \lambda\mathbf{v}, \qquad \text{where } \lambda \in \mathbb{C}.$$

Then, we say that $\mathbf{v}$ is an *eigenvector of $\mathcal{L}$ with eigenvalue* λ. In the case of a finite dimensional vector space X, the *spectrum* $\sigma(\mathcal{L})$ of $\mathcal{L}$ is the set of eigenvalues of $\mathcal{L}$. If $\mathcal{L}$ possesses n linearly independent eigenvectors, where $n = \dim X$, we say that $\mathcal{L}$ is *diagonalizable.*

Suppose $\dim X = n$ and $\mathcal{L}$ has n linearly independent eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ with corresponding eigenvalues $\lambda_1, \dots, \lambda_n$ respectively (these eigenvalues need not be distinct). Then, if we use the eigenvectors of $\mathcal{L}$ as a basis for X, it is very easy to express the action of $\mathcal{L}$ on any vector $\mathbf{x} \in X$:

$$\text{If } \ \mathbf{x} = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n, \quad \text{then } \ \mathcal{L}\mathbf{x} = \lambda_1 x_1 \mathbf{v}_1 + \cdots + \lambda_n x_n \mathbf{v}_n.$$

Essentially $\mathcal{L}$ becomes a multiplication operator, which is a huge simplification of the action of $\mathcal{L}$. Furthermore, if $f(z)$ is any polynomial in z (or even a much more general function, defined on the spectrum of $\mathcal{L}$), then $f(\mathcal{L})$ is an operator on X and its action is described as

$$f(\mathcal{L})\mathbf{x} = f(\lambda_1)x_1\mathbf{v}_1 + \cdots + f(\lambda_n)x_n\mathbf{v}_n.$$

It also follows that the matrix of $\mathcal{L}$ with respect to the basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ is diagonal:

$$[\mathcal{L}]_{\mathbf{v}} = \operatorname{diag}[\lambda_1, \dots, \lambda_n] := \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

More generally,

$$[f(\mathcal{L})]_{\mathbf{v}} = \operatorname{diag}[f(\lambda_1), \dots, f(\lambda_n)] = \begin{bmatrix} f(\lambda_1) & 0 & \cdots & 0 \\ 0 & f(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & f(\lambda_n) \end{bmatrix}.$$

Unfortunately, not all linear operators on an n-th dimensional vector space are diagonalizable. In other words there are operators which possess less that n linear independent eigenvectors. In such anomalous cases, we add to set of the eigenvectors some other vectors, called *generalized eigenvectors of level* m ($m = 2, 3, \dots$), satisfying

$$(\mathcal{L} - \lambda\mathcal{I})^m \, \mathbf{g} = \mathbf{O}, \quad (\mathcal{L} - \lambda\mathcal{I})^{m-1} \, \mathbf{g} \neq \mathbf{O}$$

(λ is necessarily an eigenvalue of $\mathcal{L}$), so that the eigenvectors of $\mathcal{L}$ together with these generalized eigenvectors form a basis for X and the matrix of $\mathcal{L}$ with respect to this basis is in *Jordan canonical form.* Notice that if $\mathcal{L}$ possesses generalized eigenvectors of level $m \geq 2$, associated to some eigenvalue

λ, then it also possesses generalized eigenvectors of levels $m-1, m-2, \dots, 2, 1$, associated to the same eigenvalue λ, where "generalized eigenvectors of level 1" means pure eigenvector. Needless to say that diagonalizable operators do not possess generalized eigenvectors.

4 Inner Product Spaces

Definition 9. An *inner product* (*vector*) *space* (or *pre-Hilbert space*) is a vector space X equipped with an *inner product* $(\cdot\,,\cdot)$, namely a (binary) operation from $X \times X$ to $\mathfrak{F}$ (where, as usual, the field $\mathfrak{F}$ of scalars is either $\mathbb{R}$ or $\mathbb{C}$) such that
(i) $(\mathbf{x}, \mathbf{x}) \geq 0$ for every $\mathbf{x} \in X$ and $(\mathbf{x}, \mathbf{x}) = 0$ if and only if $\mathbf{x} = \mathbf{O}$.
(ii) $(\mathbf{x} + \mathbf{y}, \mathbf{z}) = (\mathbf{x}, \mathbf{z}) + (\mathbf{y}, \mathbf{z})$ for every $\mathbf{x}, \mathbf{y}, \mathbf{z} \in X$.
(iii) $(a\mathbf{x}, \mathbf{y}) = a(\mathbf{x}, \mathbf{y})$ for every $\mathbf{x}, \mathbf{y} \in X$ and $a \in \mathfrak{F}$.
(iv) $(\mathbf{y}, \mathbf{x}) = \overline{(\mathbf{x}, \mathbf{y})}$ for every $\mathbf{x}, \mathbf{y} \in X$, where $\overline{(\mathbf{x}, \mathbf{y})}$ denotes the complex conjugate of $(\mathbf{x}, \mathbf{y})$ (thus, if $\mathfrak{F} = \mathbb{R}$, then $(\mathbf{y}, \mathbf{x}) = (\mathbf{x}, \mathbf{y})$).

Exercise 1. Show that $(\mathbf{x}, \mathbf{O}) = 0$ for all $\mathbf{x} \in X$.

Theorem 3 (the Schwarz inequality)**.** If $\mathbf{x}$ and $\mathbf{y}$ are vectors in an inner product space X, then

$$|(\mathbf{x}, \mathbf{y})| \leq \sqrt{(\mathbf{x}, \mathbf{x})}\,\sqrt{(\mathbf{y}, \mathbf{y})}.$$

For a proof see Remark 3 below.

A consequence of Definition 1 and the Schwarz inequality is that the inner product induces a *norm* on X:

$$\|\mathbf{x}\| := \sqrt{(\mathbf{x}, \mathbf{x})}, \qquad \mathbf{x} \in X. \tag{4.1}$$

Reminder. A *norm* on a vector space X is a function $\|\cdot\|$ from X to $\mathbb{R}$ which satisfies:
(i) (nonnegativity) $\|\mathbf{x}\| \geq 0$ for every $\mathbf{x} \in X$ and $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{O}$.
(ii) (positive homogeneity) $\|a\mathbf{x}\| = |a|\|\mathbf{x}\|$ for every $\mathbf{x} \in X$ and $a \in \mathfrak{F}$.
(iii) (the triangle inequality) $\|\mathbf{x} + \mathbf{y}\| \geq \|\mathbf{x}\| + \|\mathbf{y}\|$ for every $\mathbf{x}, \mathbf{y} \in X$.
A vector space equipped with a norm is called *normed linear space.* If X is a normed linear space, its norm may <u>not</u> come from an inner product (i.e. there may not exist an inner product for which (4.1) is satisfied by the norm

of X). In fact, a norm $\|\cdot\|$ on a vector space X is induced by an inner product if and only if it satisfies the *parallelogram law*:

$$\|\mathbf{x}+\mathbf{y}\|^2 + \|\mathbf{x}-\mathbf{y}\|^2 = 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2 \qquad \text{for all } \mathbf{x},\mathbf{y} \in X.$$

In this case, if $\mathfrak{F} = \mathbb{R}$, the inner product is given by the formula

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

while if $\mathfrak{F} = \mathbb{C}$, the inner product is given by the so-called *polarization identity*

$$4(\mathbf{x},\mathbf{y}) = \|\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}-\mathbf{y}\|^2 - i\left(\|\mathbf{x}+i\mathbf{y}\|^2 - \|\mathbf{x}-i\mathbf{y}\|^2\right).$$

Example 5 (examples of inner products and norms)**.** (i) Let $\mathbf{z} = (z_1,\dots,z_n)$ and $\mathbf{w} = (w_1,\dots,w_n)$ be two arbitrary vectors of $\mathbb{C}^n$. Then, the standard *dot product* $\mathbf{z}\cdot\mathbf{w}$ of $\mathbf{z}$ and $\mathbf{w}$ given by

$$\mathbf{z}\cdot\mathbf{w} := \sum_{j=1}^{n} z_j \overline{w_j}$$

is an example of an inner product. The induced norm is

$$\|\mathbf{z}\|_2 := \sqrt{\sum_{j=1}^{n} |z_j|^2}$$

(if $\mathbf{z} \in \mathbb{R}^n$, then the above norm is the length of $\mathbf{z}$). Other typical examples of norms of $\mathbb{C}^n$ are

$$\|\mathbf{z}\|_p := \left(\sum_{j=1}^{n} |z_j|^p\right)^{1/p}, \qquad 1 \le p < \infty,$$

and

$$\|\mathbf{z}\|_\infty := \max_{1\le j\le n} |z_j|.$$

Among the above norms, only $\|\cdot\|_2$ is induced by an inner product.
(ii) A typical inner product on the space $C[a,b]$ of the continuous complex-valued functions defined on $[a,b]$ is

$$(f,g) = \int_a^b f(x)\overline{g(x)}dx.$$

The induced norm is

$$\|f\|_2 := \sqrt{\int_a^b |f(x)|^2 dx}.$$

Other typical examples of norms of $C[a, b]$ are

$$\|f\|_p := \left[\int_a^b |f(x)|^p dx\right]^{1/p}, \qquad 1 \le p < \infty,$$

and

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

Again, among the above norms, only $\|\cdot\|_2$ is induced by an inner product.

Let us point out that if X is an inner product space, then is (natural) norm is the norm given by (4.1).

In an inner product space over $\mathbb{R}$, thanks to Schwarz inequality, we can define an *angle* between two (nonzero) vectors: Let $\mathbf{x} \neq \mathbf{O}$ and $\mathbf{y} \neq \mathbf{O}$. Their angle $\theta \in [0, \pi]$ is defined by

$$\cos\theta := \frac{(\mathbf{x}, \mathbf{y})}{\|\mathbf{x}\| \|\mathbf{y}\|}.$$

In particular, if $(\mathbf{x}, \mathbf{y}) = 0$, we say that the vectors $\mathbf{x}$ and $\mathbf{y}$ are *orthogonal* (or *perpendicular*; sometimes, the notation $\mathbf{x} \perp \mathbf{y}$ is used).
$\star$ Orthogonality can be also defined in the same way for inner product spaces over $\mathbb{C}$. Furthermore, the standard convention is that the vector $\mathbf{O}$ is orthogonal to all vectors.

Definition 10. A collection of vectors $\{\mathbf{e}_\alpha\}_{\alpha \in J}$ in an inner product space X is called an *orthonormal* (*O-N*) *set* if $(\mathbf{e}_\alpha, \mathbf{e}_\alpha) = 1$ for all $\alpha \in J$, and $(\mathbf{e}_\alpha, \mathbf{e}_\beta) = 0$ if $\alpha \neq \beta$. In abbreviated form:

$$(\mathbf{e}_\alpha, \mathbf{e}_\beta) = \delta_{\alpha\beta}, \qquad \text{for all } \alpha, \beta \in J,$$

where $\delta_{\alpha\beta}$ is the *Kronecker delta* which is equal to 1 if $\alpha = \beta$ and 0 if $\alpha \neq \beta$.

Theorem 4. If $\{\mathbf{e}_\alpha\}_{\alpha \in J}$ is an orthonormal set, then $\{\mathbf{e}_\alpha\}_{\alpha \in J}$ is linear independent.

Proof. Assume $c_1\mathbf{e}_{\alpha_1} + \cdots + c_n\mathbf{e}_{\alpha_n} = \mathbf{O}$ and take the inner product of both sides with $\mathbf{e}_{\alpha_j}$. Then, the orthonormality of $\{\mathbf{e}_\alpha\}_{\alpha \in J}$ implies immediately that $c_j = 0$. ■

⋆**Remark 2.** Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be an orthonormal basis of the space X (hence $\dim X = n$). If

$$\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n,$$

then $x_j = (\mathbf{x}, \mathbf{e}_j)$, $j = 1, 2, \dots, n$. Furthermore, if $\mathcal{L}$ is a linear operator on X and $[\mathcal{L}]_{\mathbf{e}} = [a_{jk}]_{1 \le j,k \le n}$ is its $n \times n$ matrix with respect to the O-N basis $\mathbf{e}_1, \dots, \mathbf{e}_n$, then

$$a_{jk} = (\mathcal{L}\mathbf{e}_k, \mathbf{e}_j), \qquad 1 \le j, k \le n.$$

Given the linearly independent vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ in an inner product space X, there is a useful procedure, called *Gram-Schmidt orthogonalization*, for constructing an orthonormal set $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$, such that $\langle \mathbf{x}_1, \dots, \mathbf{x}_k \rangle = \langle \mathbf{e}_1, \dots, \mathbf{e}_k \rangle$ for $k = 1, 2, \dots, n$, namely the vectors $\mathbf{x}_1, \dots, \mathbf{x}_k$ and $\mathbf{e}_1, \dots, \mathbf{e}_k$ span the same subspace of X, for all $k = 1, 2, \dots, n$ (recall Definition 3):

$$\begin{aligned}
\mathbf{w}_1 &:= \mathbf{x}_1, & \mathbf{e}_1 &:= \mathbf{w}_1/\|\mathbf{w}_1\|,\\
\mathbf{w}_2 &:= \mathbf{x}_2 - (\mathbf{x}_2, \mathbf{e}_1)\mathbf{e}_1, & \mathbf{e}_2 &:= \mathbf{w}_2/\|\mathbf{w}_2\|,\\
&\ \vdots & &\ \vdots\\
\mathbf{w}_n &:= \mathbf{x}_n - \sum_{j=1}^{n-1}(\mathbf{x}_n, \mathbf{e}_j)\mathbf{e}_j, & \mathbf{e}_n &:= \mathbf{w}_n/\|\mathbf{w}_n\|.
\end{aligned}$$

The same procedure works for a countably infinite family of linearly independent vectors $\{\mathbf{x}_j\}_{j\in\mathbb{N}}$, in which case it produces a (countably infinite) orthonormal set $\{\mathbf{e}_j\}_{j\in\mathbb{N}}$, such that $\langle \mathbf{x}_1, \dots, \mathbf{x}_k \rangle = \langle \mathbf{e}_1, \dots, \mathbf{e}_k \rangle$ for all $k \in \mathbb{N}$.

Theorem 5 (Pythagorean theorem)**.** Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be orthonormal vectors in an inner product space X. Then, for all $\mathbf{x} \in X$,

$$\|\mathbf{x}\|^2 = \sum_{j=1}^{n} |(\mathbf{x}, \mathbf{e}_j)|^2 + \left\| \mathbf{x} - \sum_{j=1}^{n}(\mathbf{x}, \mathbf{e}_j)\mathbf{e}_j \right\|^2.$$

In particular we have the inequality (called *Bessel's inequality*)

$$\|\mathbf{x}\|^2 \ge \sum_{j=1}^{n} |(\mathbf{x}, \mathbf{e}_j)|^2,$$

which becomes equality if and only if $\mathbf{x}$ lies in the linear span of $\mathbf{e}_1, \dots, \mathbf{e}_n$.

The idea of the proof is very simple: Notice that, since each $\mathbf{e}_j$, $j = 1, \dots, n$, is orthogonal to $\mathbf{x} - \sum_{j=1}^{n}(\mathbf{x}, \mathbf{e}_j)\mathbf{e}_j$, we have that the vectors

$$\mathbf{x}_1 := \sum_{j=1}^{n}(\mathbf{x}, \mathbf{e}_j)\mathbf{e}_j \qquad \text{and} \qquad \mathbf{x}_2 := \mathbf{x} - \sum_{j=1}^{n}(\mathbf{x}, \mathbf{e}_j)\mathbf{e}_j \qquad \text{are orthogonal.}$$

Thus, $\|\mathbf{x}\|^2 = (\mathbf{x}, \mathbf{x}) = (\mathbf{x}_1 + \mathbf{x}_2, \mathbf{x}_1 + \mathbf{x}_2) = (\mathbf{x}_1, \mathbf{x}_1) + (\mathbf{x}_2, \mathbf{x}_2)$.

Remark 3. We can use Theorem 5 to prove the Schwarz inequality (see Theorem 3): Assume $\mathbf{y} \neq \mathbf{O}$ (the case $\mathbf{y} = \mathbf{O}$ is trivial). Set $\mathbf{e}_1 := \mathbf{y}/\|\mathbf{y}\|$, so that $\{\mathbf{e}_1\}$ is an orthonormal set, and apply Bessel's inequality to any $\mathbf{x} \in X$ using $\{\mathbf{e}_1\}$ ($n = 1$):

$$\|\mathbf{x}\|^2 \geq |(\mathbf{x}, \mathbf{e}_1)|^2 = |(\mathbf{x}, \mathbf{y}/\|\mathbf{y}\|)|^2 = \frac{|(\mathbf{x}, \mathbf{y})|^2}{\|\mathbf{y}\|^2},$$

from which $|(\mathbf{x}, \mathbf{y})| \leq \|\mathbf{x}\|\|\mathbf{y}\|$ follows.

5 Adjoints

Definition 11. Let $\mathcal{L}$ be a linear operator on an inner product space X over $\mathbb{C}$. The *adjoint operator* $\mathcal{L}^*$ *of* $\mathcal{L}$ is the operator satisfying

$$(\mathcal{L}\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathcal{L}^*\mathbf{y}) \qquad \text{for all } \ \mathbf{x}, \mathbf{y} \in X.$$

Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be an orthonormal basis of X. If $A = [\mathcal{L}]_{\mathbf{e}} = [a_{jk}]_{1 \leq j,k \leq n}$ is the matrix of $\mathcal{L}$ with respect to the basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$, then by the Remark 2 we have that the matrix $[\mathcal{L}^*]_{\mathbf{e}}$ of $\mathcal{L}^*$ is A^H, namely

$$[\mathcal{L}^*]_{\mathbf{e}} = A^H = [\overline{a_{kj}}]_{1 \leq j,k \leq n}$$

(recall that $A^H := \bar{A}^\top$, where $A^\top$ is the *transpose* of A).

Proposition 2. The adjoint operator satisfies:
(i) $(\mathcal{L}^*)^* = \mathcal{L}$;
(ii) $(\mathcal{L} + \mathcal{M})^* = \mathcal{L}^* + \mathcal{M}^*$;
(iii) $(\mathcal{L}\mathcal{M})^* = \mathcal{M}^*\mathcal{L}^*$.

Example 6. If $\mathcal{M} = \lambda\mathcal{I}$, where λ is some complex number and $\mathcal{I}$ is the identity operator, then $\mathcal{M}^* = \bar{\lambda}\mathcal{I}$.

$\star$**Definition 12.** (i) Let $\mathcal{L}$ be a linear operator on an inner product space X over $\mathbb{C}$. If

$$\mathcal{L}\mathcal{L}^* = \mathcal{L}^*\mathcal{L},$$

i.e. if $\mathcal{L}$ commutes with its adjoint, then $\mathcal{L}$ is called *normal operator*. Likewise, if a square matrix A satisfies

$$AA^H = A^H A,$$

then A is called *normal matrix.*
(ii) If $\mathcal{L}\mathcal{L}^* = \mathcal{L}^*\mathcal{L} = \mathcal{I}$, then $\mathcal{L}$ is called *unitary operator.* Likewise if a square matrix A satisfies $AA^H = I_n$, where $I_n = \text{diag}\,[1, 1, \ldots, 1]$ is the *identity matrix*, (I_n is the matrix of $\mathcal{I}$ acting on an n-dimensional space X), then it is called *unitary matrix* (notice that for matrices the equation $AA^H = I_n$ implies also $A^H A = I_n$). A unitary matrix with real elements is called *orthogonal.*
(iii) If $\mathcal{L}^* = \mathcal{L}$, then the operator $\mathcal{L}$ is called *self-adjoint.* Likewise if a square matrix A satisfies $A^H = A$, then A is called *Hermitian matrix.* A Hermitian matrix with real elements is *symmetric*, i.e. it satisfies $A = A^\top$.

Of course, unitary operators and self-adjoint operators are special cases of normal operators.

$\star$**Theorem 6.** (i) Let $\mathcal{L}$ be a normal operator on a space X. Then

$$(\mathcal{L}^*\mathbf{x}, \mathcal{L}^*\mathbf{y}) = (\mathcal{L}\mathbf{x}, \mathcal{L}\mathbf{y}), \qquad \text{hence} \qquad \|\mathcal{L}^*\mathbf{x}\| = \|\mathcal{L}\mathbf{x}\|\,.$$

(ii) Let $\mathcal{U}$ be a unitary operator on a space X. Then

$$(\mathcal{U}\mathbf{x}, \mathcal{U}\mathbf{y}) = (\mathbf{x}, \mathbf{y}), \qquad \text{hence} \qquad \|\mathcal{U}\mathbf{x}\| = \|\mathbf{x}\|\,.$$

Proof. (i) Since $(\mathcal{L}^*)^* = \mathcal{L}$ and $\mathcal{L}\mathcal{L}^* = \mathcal{L}^*\mathcal{L}$ we have

$$(\mathcal{L}^*\mathbf{x}, \mathcal{L}^*\mathbf{y}) = (\mathbf{x}, \mathcal{L}\mathcal{L}^*\mathbf{y}) = (\mathbf{x}, \mathcal{L}^*\mathcal{L}\mathbf{y}) = (\mathcal{L}\mathbf{x}, \mathcal{L}\mathbf{y}).$$

(ii) Since $\mathcal{U}^*\mathcal{U} = \mathcal{I}$ we have

$$(\mathcal{U}\mathbf{x}, \mathcal{U}\mathbf{y}) = (\mathbf{x}, \mathcal{U}^*\mathcal{U}\mathbf{y}) = (\mathbf{x}, \mathbf{y}).$$

■

$\star$**Theorem 7.** Let $\lambda \in \mathbb{C}$ be an eigenvalue of a <u>normal</u> operator $\mathcal{L}$ with corresponding eigenvector $\mathbf{v}$, namely $\mathcal{L}\mathbf{v} = \lambda\mathbf{v}$. Then

$$\mathcal{L}^*\mathbf{v} = \bar{\lambda}\mathbf{v}.$$

Proof. Given that $\mathcal{L}$ is normal, Proposition 2(ii) and Example 6 imply that so is $(\mathcal{L} - \lambda\mathcal{I})$. Thus, by Theorem 6(i) we have

$$\|(\mathcal{L} - \lambda\mathcal{I})^*\mathbf{v}\| = \|(\mathcal{L} - \lambda\mathcal{I})\mathbf{v}\| = \|\mathcal{L}\mathbf{v} - \lambda\mathbf{v}\| = \|\mathbf{O}\| = 0,$$

Hence, $(\mathcal{L} - \lambda\mathcal{I})^*\mathbf{v} = \mathbf{O}$. The rest follows by invoking again Proposition 2(ii) and Example 6. ■

⋆⋆**Theorem 8.** Let $\mathcal{L}$ be a normal operator and assume

$$\mathcal{L}\mathbf{v} = \lambda\mathbf{v}, \qquad \mathcal{L}\mathbf{w} = \mu\mathbf{w}, \qquad \lambda \neq \mu.$$

Then, $(\mathbf{v}, \mathbf{w}) = 0$.

Proof. By Theorem 7 we have $\mathcal{L}^*\mathbf{w} = \bar{\mu}\mathbf{w}$. Hence,

$$\lambda(\mathbf{v}, \mathbf{w}) = (\lambda\mathbf{v}, \mathbf{w}) = (\mathcal{L}\mathbf{v}, \mathbf{w}) = (\mathbf{v}, \mathcal{L}^*\mathbf{w}) = (\mathbf{v}, \bar{\mu}\mathbf{w}) = \mu(\mathbf{v}, \mathbf{w}).$$

Therefore, $(\mathbf{v}, \mathbf{w}) = 0$. ■

⋆⋆**Theorem 9.** A normal operator $\mathcal{L}$ on a finite dimensional space X is diagonalizable.

Proof. Suppose $\mathcal{L}$ is not diagonalizable. Then, it possess generalized eigenvectors. In particular, there is a $\mathbf{v} \in X$ such that

$$(\mathcal{L} - \lambda\mathcal{I})^2 \mathbf{v} = \mathbf{O} \qquad \text{and} \qquad (\mathcal{L} - \lambda\mathcal{I})\,\mathbf{v} \neq \mathbf{O}.$$

If this is the case, since by Proposition 2(ii) and Example 6 the operator $(\mathcal{L} - \lambda\mathcal{I})$ is normal, Theorem 6(i) implies

$$0 = \left\|(\mathcal{L} - \lambda\mathcal{I})^2 \mathbf{v}\right\| = \|(\mathcal{L} - \lambda\mathcal{I})(\mathcal{L} - \lambda\mathcal{I})\,\mathbf{v}\| = \|(\mathcal{L} - \lambda\mathcal{I})^* (\mathcal{L} - \lambda\mathcal{I})\,\mathbf{v}\|.$$

Thus, $(\mathcal{L} - \lambda\mathcal{I})^* (\mathcal{L} - \lambda\mathcal{I})\,\mathbf{v} = \mathbf{O}$ and hence

$$(\mathbf{v}, (\mathcal{L} - \lambda\mathcal{I})^* (\mathcal{L} - \lambda\mathcal{I})\,\mathbf{v}) = 0$$

or

$$((\mathcal{L} - \lambda\mathcal{I})\,\mathbf{v}, (\mathcal{L} - \lambda\mathcal{I})\,\mathbf{v}) = 0,$$

i.e. $(\mathcal{L} - \lambda\mathcal{I})\,\mathbf{v} = \mathbf{O}$, which contradicts our assumption that $\mathbf{v}$ is a generalized eigenvector of level 2. Therefore, $\mathcal{L}$ does not possess any generalized eigenvectors. ■

Theorems 7, 8, and 9 applied to a self-adjoint operator yield the following very important corollary:

⋆⋆⋆**Corollary 1.** Let $\mathcal{L}$ be a self-adjoint operator on a space X. Then
(i) All eigenvalues of $\mathcal{L}$ are real.
(ii) Eigenvectors of $\mathcal{L}$ corresponding to different eigenvalues are orthogonal.
(iii) If $\dim X < \infty$, then the eigenvectors of $\mathcal{L}$ form a basis for X.