# 1 Linear Operators and Matrices

We begin our disussion by presenting several facts about the natural transformations between vector spaces and their representations, i.e., matrices. These form the foundation for our study of *numerical* linear algebra. Herein, we will set in place much of our notation, especially for matrices, that will be used not just for Part I but for the entirety of the book. Every student using this text should master the material from Appendix A and this chapter before moving on. The book by Horn and Johnson [44] is an excellent external reference.

*Why is linear algebra so important to numerical analysis?* That is a fair question. The answer is that many algorithms in numerical analysis — for a broad range of problem types, interpolation, approximation of functions, approximating solutions to differential or integral equations — require, at some stage in the algorithm, the investigation of a system of linear equations:

$$\begin{cases} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n = f_1, \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n = f_2, \\ \qquad\qquad\qquad\vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n = f_m. \end{cases}$$

Many algorithms will require the solution of such systems. Others, by contrast, may need some or all of the eigenvalues or singular values of the associated coefficient matrix for the system.

Before we jump into the topic of how to practically solve such a system of equations, which we will cover in Chapter 3 — or how to compute singular values (Chapter 2) and/or eigenvalues (Chapter 8) of the coefficient matrix — we need to understand the properties of such systems. This will be the topic of this chapter. Let us get started.

## 1.1 Linear Operators and Matrices

We study the *natural* mappings between vector spaces, i.e., those that preserve the vector space structure.

**Definition 1.1** (linear operator)**.** Let $\mathbb{V}$ and $\mathbb{W}$ be complex vector spaces. The mapping $A\colon \mathbb{V} \to \mathbb{W}$ is called a **linear operator** if and only if

$$A(\alpha x + \beta y) = \alpha Ax + \beta Ay, \quad \forall \alpha, \beta \in \mathbb{C}, \ \forall x, y \in \mathbb{V}.$$

https://doi.org/10.1017/9781108942607.004 Published online by Cambridge University Press

The set of all linear operators from $\mathbb{V}$ to $\mathbb{W}$ is denoted by $\mathfrak{L}(\mathbb{V},\mathbb{W})$. For simplicity, we denote by $\mathfrak{L}(\mathbb{V})$ the set of linear operators from $\mathbb{V}$ to itself. Suppose that $A, B \in \mathfrak{L}(\mathbb{V},\mathbb{W})$ and $\alpha, \beta \in \mathbb{C}$ are arbitrary. We define, in a natural way, the object $\alpha A + \beta B$ via

$$(\alpha A + \beta B)x = \alpha A x + \beta B x, \quad \forall x \in \mathbb{V}.$$

It is straightforward to prove that $\alpha A + \beta B$ is a linear operator and we get the following result.

**Proposition 1.2** (properties of $\mathfrak{L}(\mathbb{V},\mathbb{W})$)**.** *Let $\mathbb{V}$ and $\mathbb{W}$ be complex vector spaces. The set $\mathfrak{L}(\mathbb{V},\mathbb{W})$ is a vector space using the natural definitions of addition and scalar multiplication given in the last definition. If* $\dim(\mathbb{V}) = m$ *and* $\dim(\mathbb{W}) = n$, *then* $\dim(\mathfrak{L}(\mathbb{V},\mathbb{W})) = mn$.

*Proof.* See Problem 1.2. □

**Definition 1.3** ($m \times n$ matrices)**.** Let $\mathbb{K}$ be a field. We define, for any $m, n \in \mathbb{N}$,

$$\mathbb{K}^{m\times n} = \{\mathsf{A} = [a_{i,j}] \mid a_{i,j} \in \mathbb{K}, \; i = 1, \ldots, m, \; j = 1, \ldots, n\}.$$

The object $\mathsf{A}$ is called a **matrix** and the elements $a_{i,j} \in \mathbb{K}$ are called its **components** or **entries**. We call $\mathbb{C}^{m\times n}$ the set of **complex** $m \times n$ **matrices** and $\mathbb{R}^{m\times n}$ the set of **real** $m \times n$ **matrices**.

To extract the entry in the $i$th row and $j$th column of the $m \times n$ matrix $\mathsf{A} \in \mathbb{K}^{m\times n}$, we use the notation

$$[\mathsf{A}]_{i,j} = a_{i,j} \in \mathbb{K}.$$

The convention is that the entries of a matrix are denoted by the respective lowercase roman symbol. For example, the matrix $\mathsf{C}$ has entries $c_{i,j}$. We often make this identification explicit, as in writing $\mathsf{A} = [a_{i,j}] \in \mathbb{C}^{m\times n}$. We say that there are $m$ rows and $n$ columns in an $m \times n$ matrix $\mathsf{A}$. We naturally define $m \times n$ matrix addition and scalar multiplication component-wise via

$$[\mathsf{A}+\mathsf{B}]_{i,j} = a_{i,j} + b_{i,j}, \quad [\alpha\mathsf{A}]_{i,j} = \alpha a_{i,j}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n,$$

where $\mathsf{A}, \mathsf{B} \in \mathbb{K}^{m\times n}$ are arbitrary $m \times n$ matrices and $\alpha \in \mathbb{K}$ is an arbitrary scalar.

**Proposition 1.4** ($\mathbb{K}^{m\times n}$ is a vector space)**.** *With addition and scalar multiplication defined as above, $\mathbb{K}^{m\times n}$ is a vector space over $\mathbb{K}$ and* $\dim(\mathbb{K}^{m\times n}) = m \cdot n$.

*Proof.* See Problem 1.3. □

Of course, the reader will remember that matrices can be combined in more exotic ways.

**Definition 1.5** (matrix product)**.** Let $\mathsf{A} = [a_{i,k}] \in \mathbb{K}^{m\times p}$ and $\mathsf{B} = [b_{k,j}] \in \mathbb{K}^{p\times n}$. The **matrix product** $\mathsf{C} = \mathsf{AB}$ is a matrix in $\mathbb{K}^{m\times n}$ whose entries are computed according to the formula

$$[\mathsf{C}]_{i,j} = c_{i,j} = \sum_{k=1}^{p} a_{i,k} b_{k,j}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n.$$

https://doi.org/10.1017/9781108942607.004 Published online by Cambridge University Press

Next, we define a matrix–vector product, which, the reader will see, is similar to the last definition.

**Definition 1.6** (matrix–vector product)**.** Suppose that $\boldsymbol{x} = [x_s] \in \mathbb{K}^n$ and $\mathsf{A} = [a_{k,s}] \in \mathbb{K}^{m\times n}$. Then the **matrix–vector product** $\boldsymbol{y} = \mathsf{A}\boldsymbol{x}$ is a vector in $\mathbb{K}^m$ whose components are computed via the formula

$$[\boldsymbol{y}]_k = y_k = \sum_{s=1}^{n} a_{k,s}x_s, \quad k = 1, \dots, m.$$

**Remark 1.7** (identification)**.** Suppose that $\mathsf{A} \in \mathbb{C}^{m\times n}$. Then the (canonical) mapping $\mathsf{A}\colon \mathbb{C}^n \to \mathbb{C}^m$ defined by $\boldsymbol{y} = \mathsf{A}\boldsymbol{x}$ — where $\boldsymbol{x} \in \mathbb{C}^n$, so that the matrix–vector product $\boldsymbol{y}$ is in $\mathbb{C}^m$ — is linear. Mimicking the identification process outlined in Theorem A.24, we can also identify $\mathfrak{L}(\mathbb{C}^n, \mathbb{C}^m)$ with the space $\mathbb{C}^{m\times n}$ of matrices having $m$ rows and $n$ columns of complex entries. This says that all linear mappings from $\mathbb{C}^n$ to $\mathbb{C}^m$ are, essentially, matrices. This result can be generalized to identify $\mathfrak{L}(\mathbb{K}^n, \mathbb{K}^m)$ with $\mathbb{K}^{m\times n}$ for a generic field $\mathbb{K}$.

**Remark 1.8** (notation)**.** It will be helpful from this point on to always view $\mathbb{C}^k$ as a vector space of column $k$-vectors, i.e., $\mathbb{C}^{k\times 1}$. When we consider $\boldsymbol{x} \in \mathbb{C}^k$, we think

$$\boldsymbol{x} = \begin{bmatrix} | \\ \boldsymbol{x} \\ | \end{bmatrix} = \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix}.$$

Upon introducing the *transpose* operation $\cdot^{\intercal}\colon \mathbb{C}^{k\times 1} \to \mathbb{C}^{1\times k}$ as mapping column $k$-vectors to row $k$-vectors, we will often express $\boldsymbol{x} \in \mathbb{C}^k$ inline as $\boldsymbol{x} = [x_1, \dots, x_k]^{\intercal}$, i.e., as the transpose of a row vector. In a related way, given a matrix $\mathsf{A} \in \mathbb{C}^{m\times n}$ we commonly wish to represent it in a column-wise format (as a collection of column vectors) via

$$\mathsf{A} = \begin{bmatrix} | & & | \\ \boldsymbol{c}_1 & \cdots & \boldsymbol{c}_n \\ | & & | \end{bmatrix}, \quad \boldsymbol{c}_j \in \mathbb{C}^m,\ j = 1, \dots, n,$$

or in a row-wise format (as a collection of row vectors) via

$$\mathsf{A} = \begin{bmatrix} - & \boldsymbol{r}_1^{\intercal} & - \\ & \vdots & \\ - & \boldsymbol{r}_m^{\intercal} & - \end{bmatrix}, \quad \boldsymbol{r}_i \in \mathbb{C}^n,\ i = 1, \dots, m.$$

As a further shorthand, we will often write (inline) $\mathsf{A} = [\boldsymbol{c}_1, \dots, \boldsymbol{c}_n]$ and $\mathsf{A} = [\boldsymbol{r}_1, \dots, \boldsymbol{r}_m]^{\intercal}$. It is important to notice that if we view the matrix $\mathsf{A}$ in column-wise format, then the matrix–vector product $\boldsymbol{y} = \mathsf{A}\boldsymbol{x} \in \mathbb{C}^m$ is precisely

$$\boldsymbol{y} = \sum_{k=1}^{n} x_k \boldsymbol{c}_k.$$

In other words, the column vector $\boldsymbol{y}$ is a linear combination of the columns of $\mathsf{A}$.

https://doi.org/10.1017/9781108942607.004 Published online by Cambridge University Press

Thinking about $\mathsf{A} \in \mathbb{C}^{m\times n}$ as a mapping from $\mathbb{C}^n$ to $\mathbb{C}^m$, the following definitions are natural.

**Definition 1.9** (range and kernel)**.** Let $\mathsf{A} \in \mathbb{C}^{m\times n}$. The **image** (or **range**) of A is defined as

$$\operatorname{im}(\mathsf{A}) = \mathcal{R}(\mathsf{A}) = \{\boldsymbol{y} \in \mathbb{C}^m \mid \exists\, \boldsymbol{x} \in \mathbb{C}^n,\ \boldsymbol{y} = \mathsf{A}\boldsymbol{x}\} \subseteq \mathbb{C}^m.$$

The **kernel** (or **null space**) of A is

$$\ker(\mathsf{A}) = \mathcal{N}(\mathsf{A}) = \{\boldsymbol{x} \in \mathbb{C}^n \mid \mathsf{A}\boldsymbol{x} = \boldsymbol{0}\} \subseteq \mathbb{C}^n.$$

**Definition 1.10** (row and column space)**.** Suppose that the matrix $\mathsf{A} \in \mathbb{C}^{m\times n}$ is expressed column-wise as $\mathsf{A} = [\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n]$ and row-wise as $\mathsf{A} = [\boldsymbol{r}_1, \ldots, \boldsymbol{r}_m]^\intercal$. The **row space** of A is

$$\operatorname{row}(\mathsf{A}) = \operatorname{span}(\{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_m\}) \leq \mathbb{C}^n$$

and the **column space** of A is

$$\operatorname{col}(\mathsf{A}) = \operatorname{span}(\{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n\}) \leq \mathbb{C}^m.$$

The **row rank** of A is the dimension of row(A); similarly, the **column rank** is the dimension of col(A).

A very important result in linear algebra states that the row and column ranks coincide. For a proof, see, for example, [44].

**Theorem 1.11** (row and column rank)**.** *Suppose that* $\mathsf{A} \in \mathbb{C}^{m\times n}$. *The row and column ranks of* A *are equal.*

Since this is an important invariant between the domain and range of an operator, we give it a name.

**Definition 1.12** (rank)**.** The **rank** of a matrix $\mathsf{A} \in \mathbb{C}^{m\times n}$ is the dimension of its row/column space. We denote it by the symbol rank(A).

**Theorem 1.13** (range and column space)**.** *Let* $\mathsf{A} \in \mathbb{C}^{m\times n}$ *be represented column-wise as* $\mathsf{A} = [\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n]$. *Then*

$$\operatorname{im}(\mathsf{A}) = \operatorname{span}(\{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n\}) = \operatorname{col}(\mathsf{A}).$$

*In other words, the range of* A *coincides with its column space.*

*Proof.* ($\subseteq$) Let $\boldsymbol{y} \in \operatorname{im}(\mathsf{A}) \subseteq \mathbb{C}^m$. Then, by definition, there is an $\boldsymbol{x} \in \mathbb{C}^n$ for which $\boldsymbol{y} = \mathsf{A}\boldsymbol{x}$, or

$$\boldsymbol{y} = \sum_{k=1}^{n} x_k \boldsymbol{c}_k,$$

which implies that $\boldsymbol{y} \in \operatorname{col}(\mathsf{A})$.

($\supseteq$) On the other hand, if $\boldsymbol{y} \in \operatorname{col}(\mathsf{A})$, this implies that there are $\alpha_i \in \mathbb{C}$, $i = 1, \ldots, n$ such that

$$\boldsymbol{y} = \sum_{i=1}^{n} \alpha_i \boldsymbol{c}_i.$$

https://doi.org/10.1017/9781108942607.004 Published online by Cambridge University Press

Define $\boldsymbol{x} = [\alpha_1, \ldots, \alpha_n]^\intercal \in \mathbb{C}^n$. The previous identity shows that $\boldsymbol{y} = \mathsf{A}\boldsymbol{x}$, so that $\boldsymbol{y} \in \operatorname{im}(\mathsf{A})$. □

**Corollary 1.14** (range and rank)**.** *For any* $\mathsf{A} \in \mathbb{C}^{m\times n}$,

$$\dim(\operatorname{im}(\mathsf{A})) = \operatorname{rank}(\mathsf{A}).$$

**Definition 1.15** (nullity)**.** Suppose that $\mathsf{A} \in \mathbb{C}^{m\times n}$. The **nullity** of A is the dimension of $\ker(\mathsf{A})$:

$$\operatorname{nullity}(\mathsf{A}) = \dim(\ker(\mathsf{A})).$$

**Theorem 1.16** (properties of the rank)**.** *Let* $\mathsf{A} \in \mathbb{C}^{m\times n}$. *Then*

1. $\operatorname{rank}(\mathsf{A}) \le \min\{m, n\}$.
2. $\operatorname{rank}(\mathsf{A}) + \operatorname{nullity}(\mathsf{A}) = n$.
3. *For any* $\mathsf{B} \in \mathbb{C}^{n\times p}$, *we have* $\operatorname{rank}(\mathsf{AB}) \ge \operatorname{rank}(\mathsf{A}) + \operatorname{rank}(\mathsf{B}) - n$.
4. *For any* $\mathsf{C} \in \mathbb{C}^{m\times m}$ *with* $\operatorname{rank}(\mathsf{C}) = m$ *and any* $\mathsf{B} \in \mathbb{C}^{n\times n}$ *with* $\operatorname{rank}(\mathsf{B}) = n$, *it holds that*
$$\operatorname{rank}(\mathsf{CA}) = \operatorname{rank}(\mathsf{A}) = \operatorname{rank}(\mathsf{AB}).$$
5. $\operatorname{rank}(\mathsf{AB}) \le \min\{\operatorname{rank}(\mathsf{A}), \operatorname{rank}(\mathsf{B})\}$.
6. $\operatorname{rank}(\mathsf{A} + \mathsf{B}) \le \operatorname{rank}(\mathsf{A}) + \operatorname{rank}(\mathsf{B})$.

*Proof.* Some of these are given as exercises. Otherwise, see, for example, [44]. □

**Definition 1.17** (adjoint)**.** Suppose that $(\mathbb{V}, (\cdot\,,\cdot)_{\mathbb{V}})$ and $(\mathbb{W}, (\cdot\,,\cdot)_{\mathbb{W}})$ are inner product spaces over $\mathbb{C}$. Let $A \in \mathfrak{L}(\mathbb{V}, \mathbb{W})$. The **adjoint** of $A$ is a linear operator $A^* \in \mathfrak{L}(\mathbb{W}, \mathbb{V})$ that satisfies

$$(Ax, y)_{\mathbb{W}} = (x, A^*y)_{\mathbb{V}}, \quad \forall x \in \mathbb{V},\ y \in \mathbb{W}.$$

A linear operator $A \in \mathfrak{L}(\mathbb{V}) = \mathfrak{L}(\mathbb{V}, \mathbb{V})$ is called **self-adjoint** if and only if $A = A^*$.

For matrices, the adjoint has a familiar definition.

**Definition 1.18** (matrix adjoint, conjugate transpose)**.** Let $\mathsf{A} = [a_{i,j}] \in \mathbb{C}^{m\times n}$. The **matrix adjoint** (or **conjugate transpose**) of A is the matrix $\mathsf{A}^\mathsf{H} \in \mathbb{C}^{n\times m}$ with entries

$$\left[\mathsf{A}^\mathsf{H}\right]_{i,j} = \bar{a}_{j,i}.$$

The **transpose** of A is the matrix $\mathsf{A}^\intercal \in \mathbb{C}^{n\times m}$ with entries

$$[\mathsf{A}^\intercal]_{i,j} = a_{j,i}.$$

A matrix $\mathsf{A} \in \mathbb{C}^{n\times n}$ is called **Hermitian**[1] if and only if $\mathsf{A} = \mathsf{A}^\mathsf{H}$. A is called **skew-Hermitian** if and only if $\mathsf{A} = -\mathsf{A}^\mathsf{H}$. A matrix $\mathsf{A} \in \mathbb{R}^{n\times n}$ is called **symmetric** if and only if $\mathsf{A} = \mathsf{A}^\intercal$ and **skew-symmetric** if $\mathsf{A} = -\mathsf{A}^\intercal$.

Simple calculations yield the following results.

[1] Named in honor of the French mathematician Charles Hermite (1822–1901).

https://doi.org/10.1017/9781108942607.004 Published online by Cambridge University Press

**Proposition 1.19** (properties of matrix adjoints)**.** *Let* $\mathsf{A} \in \mathbb{C}^{m\times p}$ *and* $\mathsf{B} \in \mathbb{C}^{p\times n}$. *Then* $(\mathsf{AB})^{\mathsf{H}} = \mathsf{B}^{\mathsf{H}}\mathsf{A}^{\mathsf{H}}$ *and* $(\mathsf{A}^{\mathsf{H}})^{\mathsf{H}} = \mathsf{A}$.

*Proof.* See Problem 1.8. □

**Remark 1.20** (notation)**.** Observe that, above, we have naturally extended the domain of definition of the operator $\cdot^{\intercal}$. Let $\boldsymbol{x} = [x_1, \ldots, x_n]^{\intercal} \in \mathbb{C}^n$. The *conjugate transpose* of $\boldsymbol{x}$ is defined as the row vector $\boldsymbol{x}^{\mathsf{H}} = [\bar{x}_1, \ldots, \bar{x}_n]$. This conforms to the definition above, provided that we view any column $n$-vector as a matrix with $n$ rows and one column. A direct computation shows that $(\boldsymbol{x}^{\mathsf{H}})^{\mathsf{H}} = \boldsymbol{x}$ for all $\boldsymbol{x} \in \mathbb{C}^n$. Moreover, upon identifying $\mathbb{C}^{1\times 1}$ with $\mathbb{C}$, if $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^m$,

$$(\boldsymbol{x}, \boldsymbol{y})_{\ell^2(\mathbb{C}^m)} = (\boldsymbol{x}, \boldsymbol{y})_2 = \boldsymbol{y}^{\mathsf{H}}\boldsymbol{x} = \overline{\boldsymbol{x}^{\mathsf{H}}\boldsymbol{y}} = \overline{(\boldsymbol{y}, \boldsymbol{x})_2} = \overline{(\boldsymbol{y}, \boldsymbol{x})_{\ell^2(\mathbb{C}^m)}}.$$

Furthermore, if $\mathsf{A} \in \mathbb{C}^{m\times n}$, $\boldsymbol{x} \in \mathbb{C}^n$ and $\boldsymbol{y} \in \mathbb{C}^m$, then it follows that

$$(\mathsf{A}\boldsymbol{x}, \boldsymbol{y})_{\ell^2(\mathbb{C}^m)} = \boldsymbol{y}^{\mathsf{H}}\mathsf{A}\boldsymbol{x} = \left(\mathsf{A}^{\mathsf{H}}\boldsymbol{y}\right)^{\mathsf{H}}\boldsymbol{x} = (\boldsymbol{x}, \mathsf{A}^{\mathsf{H}}\boldsymbol{y})_{\ell^2(\mathbb{C}^n)},$$

where $(\,\cdot\,, \,\cdot\,)_{\ell^2(\mathbb{C}^m)}$ is the Euclidean inner product on $\mathbb{C}^m$. For any $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{x}^{\mathsf{H}} = \boldsymbol{x}^{\intercal}$, and for $\mathsf{A} \in \mathbb{R}^{m\times n}$, the conjugate transpose coincides with the transpose, $\mathsf{A}^{\intercal}$.

**Theorem 1.21** (properties of the conjugate transpose)**.** *Let* $\mathsf{A} \in \mathbb{C}^{m\times n}$. *Then*

1. $\operatorname{rank}(\mathsf{A}) = \operatorname{rank}(\mathsf{A}^{\mathsf{H}}) = \operatorname{rank}(\mathsf{A}^{\intercal})$.
2. $\ker(\mathsf{A}) = \operatorname{im}(\mathsf{A}^{\mathsf{H}})^{\perp}$.
3. $\operatorname{im}(\mathsf{A})^{\perp} = \ker(\mathsf{A}^{\mathsf{H}})$.

*Proof.* We prove the second result and leave the first and last to exercises; see Problem 1.10.

($\subseteq$) Let $\boldsymbol{x} \in \ker(\mathsf{A})$. By definition, $\mathsf{A}\boldsymbol{x} = \boldsymbol{0} \in \mathbb{C}^m$. Let $\boldsymbol{z} \in \operatorname{im}(\mathsf{A}^{\mathsf{H}})$, i.e., $\exists \boldsymbol{y} \in \mathbb{C}^m$ for which $\boldsymbol{z} = \mathsf{A}^{\mathsf{H}}\boldsymbol{y}$. Now compute

$$(\boldsymbol{z}, \boldsymbol{x})_2 = (\mathsf{A}^{\mathsf{H}}\boldsymbol{y}, \boldsymbol{x})_2 = (\boldsymbol{y}, \mathsf{A}\boldsymbol{x})_2 = 0,$$

which shows that $\boldsymbol{x} \in \operatorname{im}(\mathsf{A}^{\mathsf{H}})^{\perp}$.

($\supseteq$) Conversely, if $\boldsymbol{x} \in \operatorname{im}(\mathsf{A}^{\mathsf{H}})^{\perp}$, then $0 = (\boldsymbol{x}, \mathsf{A}^{\mathsf{H}}\boldsymbol{y})_2 = (\mathsf{A}\boldsymbol{x}, \boldsymbol{y})_2$ for every $\boldsymbol{y} \in \mathbb{C}^m$. Thus, $\mathsf{A}\boldsymbol{x} = \boldsymbol{0}$. □

**Definition 1.22** (identity)**.** The matrix $\mathsf{I}_n \in \mathbb{C}^{n\times n}$, defined by

$$[\mathsf{I}_n]_{i,j} = \delta_{i,j},$$

is known as the **matrix identity** of order $n$.

**Definition 1.23** (inverse)**.** Let $\mathsf{A} \in \mathbb{C}^{n\times n}$. If there is $\mathsf{B} \in \mathbb{C}^{n\times n}$ such that $\mathsf{AB} = \mathsf{BA} = \mathsf{I}_n$, then we say that $\mathsf{A}$ is **invertible** and call the matrix $\mathsf{B}$ an inverse of $\mathsf{A}$.

In light of Problem 1.13, we denote the inverse of $\mathsf{A}$ by $\mathsf{A}^{-1}$.

**Theorem 1.24** (properties of the inverse)**.** *Let* $\mathsf{A} \in \mathbb{C}^{n\times n}$. *Then* $\mathsf{A}$ *is invertible if and only if* $\operatorname{rank}(\mathsf{A}) = n$. *Moreover, if* $\mathsf{A}$ *is invertible,*

1. $\mathsf{A}^{-1}$ *is invertible and* $(\mathsf{A}^{-1})^{-1} = \mathsf{A}$.

https://doi.org/10.1017/9781108942607.004 Published online by Cambridge University Press

2. $\mathsf{A}^{\mathsf{H}}$ *is invertible and* $(\mathsf{A}^{\mathsf{H}})^{-1} = (\mathsf{A}^{-1})^{\mathsf{H}}$. *In this case, we write*
$$\mathsf{A}^{-\mathsf{H}} = (\mathsf{A}^{\mathsf{H}})^{-1}.$$
3. $\mathsf{A}^{\intercal}$ *is invertible and* $(\mathsf{A}^{\intercal})^{-1} = (\mathsf{A}^{-1})^{\intercal}$. *In this case, we write*
$$\mathsf{A}^{-\intercal} = (\mathsf{A}^{\intercal})^{-1}.$$
4. *For all* $\alpha \in \mathbb{C}_\star = \mathbb{C}\backslash\{0\}$, $\alpha\mathsf{A}$ *is invertible and* $(\alpha\mathsf{A})^{-1} = \frac{1}{\alpha}\mathsf{A}^{-1}$.
5. *If* $\mathsf{B} \in \mathbb{C}^{n\times n}$ *is also invertible, then the product* $\mathsf{AB}$ *is invertible and* $(\mathsf{AB})^{-1} = \mathsf{B}^{-1}\mathsf{A}^{-1}$.

*Proof.* See Problem 1.14. □

**Definition 1.25** (unitary matrices)**.** Let $\mathsf{A} \in \mathbb{R}^{m\times m}$. We say that $\mathsf{A}$ is **orthogonal** if and only if $\mathsf{A}^{-1} = \mathsf{A}^{\intercal}$. Similarly, for $\mathsf{A} \in \mathbb{C}^{m\times m}$, we say that $\mathsf{A}$ is **unitary** if and only if $\mathsf{A}^{\mathsf{H}} = \mathsf{A}^{-1}$.

## 1.2 Matrix Norms

Since, for any two vector spaces $\mathbb{V}$ and $\mathbb{W}$, the set $\mathfrak{L}(\mathbb{V}, \mathbb{W})$ is a vector space itself, we can think of ways of norming it. An immediate way of doing so is by simply considering elements of $\mathbb{C}^{m\times n}$ as a collection of $mn$ numbers, i.e., by identifying $\mathbb{C}^{m\times n}$ with $\mathbb{C}^{m\cdot n}$.

**Definition 1.26** (Frobenius norm[2])**.** Let $\mathsf{A} = [a_{i,j}] \in \mathbb{C}^{m\times n}$. The **Frobenius norm** is defined via
$$\|\mathsf{A}\|_F^2 = \sum_{i=1}^{m}\sum_{j=1}^{n} |a_{i,j}|^2.$$

**Definition 1.27** (max norm)**.** The **matrix max norm** is defined via
$$\|\mathsf{A}\|_{\max} = \max_{\substack{1\le i\le m\\ 1\le j\le n}} |a_{i,j}|$$
for all $\mathsf{A} = [a_{i,j}] \in \mathbb{C}^{m\times n}$.

However, it turns out that it is often more useful when the norms on $\mathfrak{L}(\mathbb{V}, \mathbb{W})$ are, in a sense, compatible with those of $\mathbb{V}$ and $\mathbb{W}$.

**Definition 1.28** (induced norm)**.** Let $(\mathbb{V}, \|\cdot\|_{\mathbb{V}})$ and $(\mathbb{W}, \|\cdot\|_{\mathbb{W}})$ be complex, finite-dimensional normed vector spaces. The **induced norm** on $\mathfrak{L}(\mathbb{V}, \mathbb{W})$ is
$$\|A\|_{\mathfrak{L}(\mathbb{V},\mathbb{W})} = \sup_{x\in\mathbb{V}_\star} \frac{\|Ax\|_{\mathbb{W}}}{\|x\|_{\mathbb{V}}}, \quad \forall A \in \mathfrak{L}(\mathbb{V}, \mathbb{W}),$$
where $\mathbb{V}_\star = \mathbb{V}\backslash\{0\}$. When $\mathbb{V} = \mathbb{W}$ it is understood that $\|\cdot\|_{\mathbb{V}} = \|\cdot\|_{\mathbb{W}}$ as well.

[2] Named in honor of the German mathematician Ferdinand Georg Frobenius (1849–1917).

https://doi.org/10.1017/9781108942607.004 Published online by Cambridge University Press

**Remark 1.29** (convention)**.** Regarding the last point, in our presentation, the following object would *not* define an induced matrix norm:

$$\|\mathsf{A}\|_{\mathfrak{L}(\ell^p(\mathbb{C}^n),\ell^q(\mathbb{C}^n))} = \sup_{\boldsymbol{x}\in\mathbb{C}^n_\star} \frac{\|\mathsf{A}\boldsymbol{x}\|_{\ell^q(\mathbb{C}^n)}}{\|\boldsymbol{x}\|_{\ell^p(\mathbb{C}^n)}}, \quad \forall \mathsf{A} \in \mathbb{C}^{n\times n}$$

for $p \neq q$. While this definition is meaningful for every $p, q \in [1, \infty]$, and it indeed defines a norm, we will only consider it to be an induced norm for $p = q$.

**Definition 1.30** (matrix $p$-norm)**.** Let $\mathsf{A} \in \mathbb{C}^{m\times n}$ be given and $p \in [1, \infty]$. The induced $\mathfrak{L}(\ell^p(\mathbb{C}^n), \ell^p(\mathbb{C}^m))$ norm, called simply the **induced matrix** $\boldsymbol{p}$**-norm**, is denoted $\|\mathsf{A}\|_p$ and is defined as

$$\|\mathsf{A}\|_p = \sup_{\boldsymbol{x}\in\mathbb{C}^n_\star} \frac{\|\mathsf{A}\boldsymbol{x}\|_{\ell^p(\mathbb{C}^m)}}{\|\boldsymbol{x}\|_{\ell^p(\mathbb{C}^n)}}.$$

**Proposition 1.31** (matrix 1-norm)**.** *Let* $\mathsf{A} = [a_{i,j}] = [a_{i,j}] = [\boldsymbol{c}_1, \dots, \boldsymbol{c}_n] \in \mathbb{C}^{m\times n}$ *be arbitrary. The induced matrix 1-norm, which is, by definition,*

$$\|\mathsf{A}\|_1 = \sup_{\boldsymbol{x}\in\mathbb{C}^n_\star} \frac{\|\mathsf{A}\boldsymbol{x}\|_{\ell^1(\mathbb{C}^m)}}{\|\boldsymbol{x}\|_{\ell^1(\mathbb{C}^n)}},$$

*may be calculated via the following formula:*

$$\|\mathsf{A}\|_1 = \max_{j=1}^{n} \left( \sum_{i=1}^{m} |a_{i,j}| \right).$$

*Proof.* Given any $\boldsymbol{x} = [x_1, \dots, x_n]^\intercal \in \mathbb{C}^n$,

$$\begin{aligned}
\|\mathsf{A}\boldsymbol{x}\|_{\ell^1(\mathbb{C}^m)} &= \left\| \sum_{j=1}^{n} x_j \boldsymbol{c}_j \right\|_{\ell^1(\mathbb{C}^m)} \\
&\leq \sum_{j=1}^{n} |x_j| \|\boldsymbol{c}_j\|_{\ell^1(\mathbb{C}^m)} \\
&\leq \max_{j=1}^{n} \|\boldsymbol{c}_j\|_{\ell_1(\mathbb{C}^m)} \sum_{j=1}^{n} |x_j| \\
&= \max_{j=1}^{n} \|\boldsymbol{c}_j\|_{\ell_1(\mathbb{C}^m)} \|\boldsymbol{x}\|_{\ell^1(\mathbb{C}^n)}.
\end{aligned}$$

This shows that

$$\|\mathsf{A}\|_1 \leq \max_{j=1}^{n} \|\boldsymbol{c}_j\|_{\ell_1(\mathbb{C}^m)} = \max_{j=1}^{n} \left( \sum_{i=1}^{m} |a_{i,j}| \right).$$

On the other hand, there must be an index $j_0$ where the maximum in the previous inequality is attained. Choose $\boldsymbol{x} = \boldsymbol{e}_{j_0}$, the $j_0$th canonical basis vector, and notice then that

$$\|\mathsf{A}\boldsymbol{x}\|_{\ell^1(\mathbb{C}^m)} = \|\boldsymbol{c}_{j_0}\|_{\ell^1(\mathbb{C}^m)}.$$

https://doi.org/10.1017/9781108942607.004 Published online by Cambridge University Press

It is not difficult to see that the supremum in the definition of induced norm is attained at this vector. This implies that the norm is the maximum absolute column sum, i.e.,

$$\|A\|_1 = \max_{j=1}^{n} \left( \sum_{i=1}^{m} |a_{i,j}| \right). \qquad \square$$

**Definition 1.32** (sub-multiplicativity)**.** Suppose that $\|\cdot\| : \mathbb{C}^{n\times n} \to \mathbb{R}$ is a matrix norm, i.e., a norm on the vector space $\mathfrak{L}(\mathbb{C}^n)$. We say that the norm is **sub-multiplicative** if and only if

$$\|AB\| \le \|A\| \, \|B\|, \quad \forall A, B \in \mathbb{C}^{n\times n}.$$

**Definition 1.33** (consistency)**.** Suppose that $\|\cdot\|_{\mathbb{C}^n} : \mathbb{C}^n \to \mathbb{R}$ and $\|\cdot\|_{\mathbb{C}^m} : \mathbb{C}^m \to \mathbb{R}$ are norms, and $\|\cdot\| : \mathbb{C}^{m\times n} \to \mathbb{R}$ is a matrix norm. We say that $\|\cdot\|$ is **consistent** with respect to the norms $\|\cdot\|_{\mathbb{C}^n}$ and $\|\cdot\|_{\mathbb{C}^m}$ if and only if

$$\|A\boldsymbol{x}\|_{\mathbb{C}^m} \le \|A\| \, \|\boldsymbol{x}\|_{\mathbb{C}^n}$$

for all $A \in \mathbb{C}^{m\times n}$ and $\boldsymbol{x} \in \mathbb{C}^n$.

**Proposition 1.34** (property of induced norms)**.** *Suppose that* $\|\cdot\|_{\mathbb{C}^n} : \mathbb{C}^n \to \mathbb{R}$ *is a norm on* $\mathbb{C}^n$ *and* $\|\cdot\| : \mathbb{C}^{n\times n} \to \mathbb{R}$ *is the induced matrix norm*

$$\|A\| = \sup_{\boldsymbol{x}\in\mathbb{C}^n_\star} \frac{\|A\boldsymbol{x}\|_{\mathbb{C}^n}}{\|\boldsymbol{x}\|_{\mathbb{C}^n}}, \quad \forall A \in \mathbb{C}^{n\times n}.$$

*Then* $\|\cdot\|$ *is a sub-multiplicative norm, and it is consistent with respect to* $\|\cdot\|_{\mathbb{C}^n}$.

*Proof.* See Problem 1.27. □

---

**Example 1.1** Let $A \in \mathbb{C}^{1\times n}$, i.e., $A = \boldsymbol{a}^{\mathrm{H}}$ for some $\boldsymbol{a} \in \mathbb{C}^n$. Then $A\boldsymbol{x} = (\boldsymbol{x}, \boldsymbol{a})_2$, so that

$$|A\boldsymbol{x}| = |(\boldsymbol{x}, \boldsymbol{a})_2| \le \|\boldsymbol{x}\|_2 \|\boldsymbol{a}\|_2.$$

In addition,

$$|A\boldsymbol{a}| = |(\boldsymbol{a}, \boldsymbol{a})_2| = \|\boldsymbol{a}\|_2^2,$$

from which we may conclude that $\|A\|_2 = \|\boldsymbol{a}\|_2$. This matrix $A\colon \mathbb{C}^n \to \mathbb{C}$ is a prototype of an object called a *linear functional*.

---

**Proposition 1.35** (norm of a unitary matrix)**.** *Let* $A \in \mathbb{C}^{m\times n}$ *be arbitrary and* $Q \in \mathbb{C}^{m\times m}$ *be unitary. Then we have*

$$\|QA\|_2 = \|A\|_2.$$

*Proof.* Recall that, owing to Problem 1.16, for any unitary matrix we have $\|Q\boldsymbol{x}\|_2 = \|\boldsymbol{x}\|_2$. The result follows from this fact. □

https://doi.org/10.1017/9781108942607.004 Published online by Cambridge University Press

## 1.3 Eigenvalues and Spectral Decomposition

As a final topic in this chapter we discuss eigenvalues and spectral decomposition of square matrices. We begin with a definition.

**Definition 1.36** (spectrum)**.** Let $\mathsf{A} \in \mathbb{C}^{n\times n}$. We say that $\lambda \in \mathbb{C}$ is an **eigenvalue** of $\mathsf{A}$ if and only if there exists a vector $\boldsymbol{x} \in \mathbb{C}_\star^n = \mathbb{C}^n \backslash \{\mathbf{0}\}$ such that

$$\mathsf{A}\boldsymbol{x} = \lambda \boldsymbol{x}.$$

This vector is called an **eigenvector** of $\mathsf{A}$ associated with $\lambda$. The **spectrum** of $\mathsf{A}$, denoted by $\sigma(\mathsf{A})$, is the collection of all eigenvalues of $\mathsf{A}$. The pair $(\lambda, \boldsymbol{x})$ is called an **eigenpair** of $\mathsf{A}$.

**Theorem 1.37** (properties of the spectrum)**.** *Let* $\mathsf{A} \in \mathbb{C}^{n\times n}$*. Then*

1. $\lambda \in \sigma(\mathsf{A})$ *if and only if* $\bar{\lambda} \in \sigma(\mathsf{A}^{\mathsf{H}})$*.*
2. $\mathsf{A}$ *is invertible if and only if* $0 \notin \sigma(\mathsf{A})$*.*
3. *The eigenvectors corresponding to distinct eigenvalues are linearly independent.*
4. $\lambda \in \sigma(\mathsf{A})$ *if and only if* $\chi_{\mathsf{A}}(\lambda) = 0$*, where* $\chi_{\mathsf{A}}$ *is a polynomial of degree* $n$*, defined via*

$$\chi_{\mathsf{A}}(\lambda) = \det(\lambda \mathsf{I}_n - \mathsf{A}).$$

   $\chi_{\mathsf{A}}$ *is called the* characteristic polynomial*.*
5. *There are at most* $n$ *distinct complex-valued eigenvalues of* $\mathsf{A}$*.*

*Proof.* See Problem 1.28. □

Since we are dealing with matrices with complex entries, the fundamental theorem of algebra (see [18, Section 2.8]) implies that the characteristic polynomial can be written as a product of factors, i.e.,

$$\chi_{\mathsf{A}}(\lambda) = \prod_{i=1}^{L} (\lambda - \lambda_i)^{m_i} \tag{1.1}$$

with $n = \sum_{i=1}^{L} m_i$.

**Definition 1.38** (algebraic multiplicity)**.** Let $\mathsf{A} \in \mathbb{C}^{n\times n}$ be given. The number $m_i$ in (1.1) is called the **algebraic multiplicity** of the eigenvalue $\lambda_i$.

**Definition 1.39** (geometric multiplicity)**.** Let $\mathsf{A} \in \mathbb{C}^{n\times n}$ and $\lambda \in \sigma(\mathsf{A})$. Define the eigenspace

$$E(\lambda, \mathsf{A}) = \{\boldsymbol{x} \in \mathbb{C}^n \mid \mathsf{A}\boldsymbol{x} = \lambda \boldsymbol{x}\}.$$

This is a vector subspace of $\mathbb{C}^n$; its dimension $\dim(E(\lambda, \mathsf{A}))$ is called the **geometric multiplicity** of $\lambda$.

The following result gives a relation between the algebraic and geometric multiplicities of an eigenvalue. For a proof of this result, see [44].

**Theorem 1.40** (relation between multiplicities)**.** *Let* $\mathsf{A} \in \mathbb{C}^{n\times n}$ *and* $\lambda \in \sigma(\mathsf{A})$*. The geometric multiplicity of* $\lambda$ *is not larger than the algebraic multiplicity of* $\lambda$*.*

https://doi.org/10.1017/9781108942607.004 Published online by Cambridge University Press

**Definition 1.41** (triangular matrices)**.** The square matrix $\mathsf{A} = [a_{i,j}] \in \mathbb{C}^{n\times n}$ is called **upper triangular** if and only if $a_{i,j} = 0$ for all $i > j$. A is called **lower triangular** if and only if $a_{i,j} = 0$ for all $i < j$. A matrix is called **triangular** if and only if it is either upper or lower triangular. A is called **diagonal** if and only if $a_{i,j} = 0$ for all $i \neq j$. A matrix $\mathsf{A} = [a_{i,j}] \in \mathbb{C}^{n\times n}$ is called **unit lower triangular** (**unit upper triangular**) if and only if it is lower (upper) triangular and $a_{i,i} = 1$, $i = 1, \dots, n$.

**Definition 1.42** (similarity)**.** Let $\mathsf{A}, \mathsf{B} \in \mathbb{C}^{n\times n}$. We say that A and B are **similar**, denoted by $\mathsf{A} \asymp \mathsf{B}$, if and only if there is an invertible matrix S such that

$$\mathsf{A} = \mathsf{S}^{-1}\mathsf{B}\mathsf{S}.$$

We say that matrix A is **diagonalizable** if it is similar to a diagonal matrix.

**Proposition 1.43** (spectrum of similar matrices)**.** *Let* $\mathsf{A}, \mathsf{B} \in \mathbb{C}^{n\times n}$ *be such that* $\mathsf{A} \asymp \mathsf{B}$*. Then* $\chi_\mathsf{A} = \chi_\mathsf{B}$ *and, consequently,* $\sigma(\mathsf{A}) = \sigma(\mathsf{B})$*. Furthermore,* $\det(\mathsf{A}) = \det(\mathsf{B})$ *and* $\operatorname{tr}(\mathsf{A}) = \operatorname{tr}(\mathsf{B})$*.*

*Proof.* See Problem 1.34. $\square$

**Definition 1.44** (defective matrix)**.** A matrix $\mathsf{A} \in \mathbb{C}^{n\times n}$ is called **defective** if and only if there is an eigenvalue $\lambda_k$ with geometric multiplicity strictly smaller than the algebraic multiplicity. Otherwise, the matrix is called **nondefective**.

One of the main results in the spectral theory of matrices is the following.

**Theorem 1.45** (diagonalizability criterion)**.** *Let* $\mathsf{A} \in \mathbb{C}^{n\times n}$ *be nondefective. Then it is diagonalizable.*

*Proof.* Let $\sigma(\mathsf{A}) = \{\lambda_k\}_{k=1}^{L}$, where $\lambda_k \neq \lambda_j$, $k \neq j$. For each $k$,

$$E(\lambda_k, \mathsf{A}) = \operatorname{span}(\{\boldsymbol{x}_1^{(k)}, \dots, \boldsymbol{x}_{m_k}^{(k)}\}) = \operatorname{span}(S_k),$$

where the set $S_k = \{\boldsymbol{x}_1^{(k)}, \dots, \boldsymbol{x}_{m_k}^{(k)}\}$ is linearly independent. Then $S = \cup_{k=1}^{L} S_k$ is a basis of $\mathbb{C}^n$. Indeed, item 3 of Theorem 1.37 shows that the set $S$ is linearly independent. Moreover, $\#(S) = \sum_{k=1}^{L} m_k = n$, since the matrix A is nondefective.

Now set $\mathsf{D} = \operatorname{diag}(\lambda_1, \dots, \lambda_1, \dots, \lambda_L, \dots, \lambda_L)$ and

$$\mathsf{X} = \begin{bmatrix} | & & | & & | & & | \\ \boldsymbol{x}_1^{(1)} & \cdots & \boldsymbol{x}_{m_1}^{(1)} & \cdots & \boldsymbol{x}_1^{(L)} & \cdots & \boldsymbol{x}_{m_L}^{(L)} \\ | & & | & & | & & | \end{bmatrix},$$

where in D each eigenvalue $\lambda_k$ appears exactly $m_k$ times. Notice now that, since all the columns of X are linearly independent, we have $\operatorname{rank}(\mathsf{X}) = n$ and this implies that X is invertible.

Since, for all $j = 1, \dots, m_k$, we have $\mathsf{A}\boldsymbol{x}_j^{(k)} = \lambda_k \boldsymbol{x}_j^{(k)}$, we see that

$$\mathsf{A}\mathsf{X} = \mathsf{A}[\boldsymbol{x}_1, \dots, \boldsymbol{x}_n] = [\mathsf{A}\boldsymbol{x}_1, \dots, \mathsf{A}\boldsymbol{x}_n] \quad \text{and} \quad \mathsf{X}\mathsf{D} = [\lambda_1\boldsymbol{x}_1, \dots, \lambda_n\boldsymbol{x}_n].$$

This implies that $\mathsf{A}\mathsf{X} = \mathsf{X}\mathsf{D}$, or, since $X$ is invertible, $\mathsf{A} = \mathsf{X}\mathsf{D}\mathsf{X}^{-1}$. In conclusion, A is diagonalizable. $\square$

https://doi.org/10.1017/9781108942607.004 Published online by Cambridge University Press

An important class of nondefective matrices are those that are self-adjoint, or Hermitian. To investigate these, we use the Schur factorization. For a proof, again, we refer to [44].

**Lemma 1.46** (Schur normal form[3]). *Let* $\mathsf{A} \in \mathbb{C}^{n\times n}$*. There are, not necessarily unique, matrices* $\mathsf{U}, \mathsf{R} \in \mathbb{C}^{n\times n}$*, with* $\mathsf{U}$ *unitary and* $\mathsf{R}$ *upper triangular, such that*

$$\mathsf{A} = \mathsf{URU}^{\mathsf{H}}.$$

Notice that, in the setting of Lemma 1.46, we have that $\mathsf{A} \asymp \mathsf{R}$ and that, since $\mathsf{R}$ is upper triangular, its diagonal entries coincide with its spectrum.

**Proposition 1.47** (Spectral Decomposition Theorem). *Let* $\mathsf{A} \in \mathbb{C}^{n\times n}$ *be self-adjoint (Hermitian), i.e.,* $\mathsf{A}^{\mathsf{H}} = \mathsf{A}$*. Then* $\sigma(\mathsf{A}) \subseteq \mathbb{R}$ *and there is a unitary* $\mathsf{U} \in \mathbb{C}^{n\times n}$ *such that*

$$\mathsf{A} = \mathsf{UDU}^{\mathsf{H}},$$

*where the matrix* $\mathsf{D} = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$*. Furthermore, there exists an orthonormal basis* $B = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\}$ *of eigenvectors of* $\mathsf{A}$ *for the space* $\mathbb{C}^n$ *and* $\mathsf{A}\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i$*,* $i = 1, \dots, n$*.*

*Proof.* From Lemma 1.46 we are guaranteed that there is a unitary matrix $\mathsf{U} \in \mathbb{C}^{n\times n}$ and an upper triangular matrix $\mathsf{D} \in \mathbb{C}^{n\times n}$ such that

$$\mathsf{A} = \mathsf{UDU}^{\mathsf{H}}.$$

But, since $\mathsf{A}$ is self-adjoint,

$$\mathsf{A}^{\mathsf{H}} = \mathsf{UD}^{\mathsf{H}}\mathsf{U}^{\mathsf{H}} = \mathsf{UDU}^{\mathsf{H}} = \mathsf{A}.$$

This implies that $\mathsf{D}^{\mathsf{H}} = \mathsf{D}$, i.e., $\mathsf{D}$ is self-adjoint. Since $\mathsf{D}$ is triangular, it must be diagonal. Furthermore, the diagonal elements of $\mathsf{D}$ must be real. Otherwise, $\mathsf{D}$ could not be self-adjoint. Therefore, we have the desired factorization.

Now the eigenvalues of a diagonal matrix are precisely its diagonal entries. Since $\mathsf{A}$ is similar to the diagonal matrix $\mathsf{D}$, the eigenvalues of $\mathsf{A}$ are precisely $\lambda_i = d_{i,i} \in \mathbb{R}$, $i = 1, \dots, n$.

Finally, observe that the columns of $\mathsf{U}$ form an orthonormal basis for $\mathbb{C}^n$. Indeed, suppose that the $k$th column of $\mathsf{U}$ is denoted $\boldsymbol{u}_k$. Then $\mathsf{AU} = \mathsf{UD}$ if and only if

$$\mathsf{A}\boldsymbol{u}_k = d_{k,k}\boldsymbol{u}_k = \lambda_k \boldsymbol{u}_k.$$

Thus, the eigenvectors of $\mathsf{A}$, namely $\boldsymbol{u}_k$, $k = 1, \dots, n$, form an orthonormal basis for $\mathbb{C}^n$: $(\boldsymbol{u}_k, \boldsymbol{u}_j)_2 = \boldsymbol{u}_j^{\mathsf{H}}\boldsymbol{u}_k = \delta_{k,j}$, $k, j = 1, \dots, n$. $\square$

Notice that the previous result shows that, for $\mathsf{A}$ self-adjoint, there exists an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors of $\mathsf{A}$. This is a result that is used countless times in the text.

There are numerous generalizations of the last theorem. We will be interested in one that is rather straightforward to establish. First, we need what is perhaps an obvious definition.

[3] Named in honor of the Russian-born German–Israeli mathematician Issai Schur (1875–1941).

**Definition 1.48** (eigenvalue)**.** Suppose that $\mathbb{V}$ is a complex vector space and $A \in \mathfrak{L}(\mathbb{V})$. The scalar $\lambda \in \mathbb{C}$ for which there is $w \in \mathbb{V} \setminus \{0\}$ such that

$$Aw = \lambda w,$$

is called an **eigenvalue** of $A$ and $w$ is a corresponding **eigenvector**. The spectrum of $A$, $\sigma(A)$, is the set of all eigenvalues of $A$. The pair $(\lambda, w)$ is called an **eigenpair** of $A$.

For self-adjoint operators we have the following general result.

**Theorem 1.49** (Spectral Decomposition Theorem)**.** *Suppose that $(\mathbb{V}, (\cdot, \cdot))$ is an $n$-dimensional complex inner product space and $A \in \mathfrak{L}(\mathbb{V})$ is self-adjoint. Then there are precisely $n$ eigenvalues, counting multiplicities, and $\sigma(A) \subseteq \mathbb{R}$. Moreover, there is an orthonormal basis $B = \{w_1, \ldots, w_n\}$ of eigenvectors of $A$ for the space $\mathbb{V}$: $(w_i, w_j) = \delta_{i,j}$, $i, j = 1, \ldots, n$.*

*Proof.* A proof for this is, for instance, furnished by the theory developed in Chapter 7. □

Finally, the class of normal matrices, which contains as a proper subset the class of Hermitian matrices, is sometimes important.

**Definition 1.50** (normal matrix)**.** The square matrix $\mathsf{A} \in \mathbb{C}^{n\times n}$ is called **normal** if and only if $\mathsf{A}^{\mathsf{H}}\mathsf{A} = \mathsf{A}\mathsf{A}^{\mathsf{H}}$.

We will need the following technical lemma.

**Lemma 1.51** (normal and triangular)**.** *Suppose that $\mathsf{A} \in \mathbb{C}^{n\times n}$ is normal and upper triangular. Then it must be diagonal.*

*Proof.* See Problem 1.45. □

**Theorem 1.52** (diagonalization of normal matrices)**.** *Suppose that $\mathsf{A} \in \mathbb{C}^{n\times n}$ is normal. Then $\mathsf{A}$ is unitarily diagonalizable, i.e., there is a unitary matrix $\mathsf{U} \in \mathbb{C}^{n\times n}$ and a diagonal matrix $\mathsf{D} \in \mathbb{C}^{n\times n}$ such that*

$$\mathsf{A} = \mathsf{U}\mathsf{D}\mathsf{U}^{\mathsf{H}}.$$

*Proof.* Use the Schur factorization and Lemma 1.51. See Problem 1.46. □

**Corollary 1.53** (orthonormal basis)**.** *Suppose that $\mathsf{A} \in \mathbb{C}^{n\times n}$ is normal. There is an orthonormal basis of eigenvectors of $\mathsf{A}$ for $\mathbb{C}^n$.*

*Proof.* Repeat the construction of Proposition 1.47. □

## Problems

**1.1** Let $(\mathbb{V}, \|\cdot\|)$ be a finite-dimensional normed space and $A \in \mathfrak{L}(\mathbb{V})$. Does

$$\|\cdot\|_A = \|A\cdot\| : \mathbb{V} \to \mathbb{R}$$

define a norm? Why or why not?

**1.2** Prove Proposition 1.2.

https://doi.org/10.1017/9781108942607.004 Published online by Cambridge University Press

**1.3** Prove Proposition 1.4.

**1.4** For $\mathsf{A} \in \mathbb{C}^{m\times n}$, prove that $\mathrm{im}(\mathsf{A}) \leq \mathbb{C}^m$ (i.e., $\mathrm{im}(\mathsf{A})$ is a vector subspace of $\mathbb{C}^m$) and $\ker(\mathsf{A}) \leq \mathbb{C}^n$ (i.e., $\ker(\mathsf{A})$ is a vector subspace of $\mathbb{C}^n$).

**1.5** Suppose that $\mathsf{A} \in \mathbb{C}^{m\times n}$. Prove that $\mathrm{im}(\mathsf{A}) = C_{\mathsf{A}}$, where $\mathrm{im}(\mathsf{A})$ is the range of A and $C_{\mathsf{A}}$ is its column space.

**1.6** Suppose that $\mathsf{A} \in \mathbb{C}^{m\times n}$ with $m \geq n$. Prove that the following are equivalent:

a) $\mathrm{rank}(\mathsf{A}) = n$.

b) A maps no two distinct vectors in $\mathbb{C}^n$ to the same vector in $\mathbb{C}^m$.

c) $\ker(\mathsf{A}) = \{\mathbf{0}\}$.

**1.7** Let $\mathsf{A} \in \mathbb{C}^{m\times n}$. Prove that $\mathrm{im}(\mathsf{A})^\perp = \ker(\mathsf{A}^\mathsf{H})$.

**1.8** Prove Proposition 1.19.

**1.9** Show that the definitions "adjoint" and the "conjugate transpose" coincide for matrices when we use the canonical inner product

$$(\boldsymbol{x}, \boldsymbol{y})_{\ell^2(\mathbb{C}^m)} = (\boldsymbol{x}, \boldsymbol{y})_2 = \boldsymbol{y}^\mathsf{H}\boldsymbol{x}$$

for $\mathbb{C}^m$.

**1.10** Complete the proof of Theorem 1.21.

**1.11** Show that $\mathsf{I}_n \in \mathbb{C}^{n\times n}$ acts as multiplicative identity with respect to matrix multiplication. In other words, for every $\mathsf{A} \in \mathbb{C}^{n\times n}$, we have

$$\mathsf{A}\mathsf{I}_n = \mathsf{I}_n\mathsf{A} = \mathsf{A}.$$

**1.12** Suppose that $\mathsf{C} \in \mathbb{C}^{n\times n}$ is invertible and the set $S = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k\} \subseteq \mathbb{C}^n$ is linearly independent. Prove that $\mathsf{C}S = \{\mathsf{C}\boldsymbol{w}_1, \ldots, \mathsf{C}\boldsymbol{w}_k\} \subseteq \mathbb{C}^n$ is linearly independent.

**1.13** Suppose that $\mathsf{A} \in \mathbb{C}^{n\times n}$ is invertible. Prove that its inverse must be unique.

**1.14** Prove Theorem 1.24.

**1.15** Let $\mathsf{A} \in \mathbb{C}^{m\times n}$. Prove that $\mathrm{rank}(\mathsf{A}) = \mathrm{rank}(\mathsf{AB})$ for any $\mathsf{B} \in \mathbb{C}^{n\times n}$ that is invertible.

**1.16** Let $\mathsf{U} \in \mathbb{C}^{n\times n}$ be unitary. Show that, for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$, we have $(\mathsf{U}\boldsymbol{x}, \mathsf{U}\boldsymbol{y})_2 = (\boldsymbol{x}, \boldsymbol{y})_2$, so that $\|\mathsf{U}\boldsymbol{x}\|_2 = \|\boldsymbol{x}\|_2$.

**1.17** Show that the Frobenius and matrix max norms are indeed norms on the vector space $\mathfrak{L}(\mathbb{C}^n, \mathbb{C}^m)$.

**1.18** Show that

$$\|\mathsf{A}\|_F^2 = \mathrm{tr}(\mathsf{A}^\mathsf{H}\mathsf{A}) = \mathrm{tr}(\mathsf{A}\mathsf{A}^\mathsf{H}),$$

where, for any square matrix, $\mathsf{M} = [m_{i,j}] \in \mathbb{C}^{n\times n}$, $\mathrm{tr}(\mathsf{M}) = \sum_i^n m_{i,i}$ denotes its *trace*.

**1.19** Let $\mathbb{V}$ and $\mathbb{W}$ be finite-dimensional complex-normed vector spaces. Show that the induced norm is indeed a norm on the vector space $\mathfrak{L}(\mathbb{V}, \mathbb{W})$. Prove that

$$\|A\|_{\mathfrak{L}(\mathbb{V},\mathbb{W})} = \sup\left\{\|Ax\|_{\mathbb{W}} \mid x \in \mathbb{V},\ \|x\|_{\mathbb{V}} = 1\right\}.$$

**1.20** Let, for $a, b \in \mathbb{R}$,

$$\mathsf{A} = \begin{bmatrix} a & b \\ b & a \end{bmatrix}.$$

Show that $\|\mathsf{A}\|_1 = \|\mathsf{A}\|_2 = \|\mathsf{A}\|_\infty$.

https://doi.org/10.1017/9781108942607.004 Published online by Cambridge University Press

**1.21** Let, for $a, b \in \mathbb{R}$,

$$\mathsf{A} = \begin{bmatrix} a & b \\ b & -a \end{bmatrix}.$$

Show that $\|\mathsf{A}\|_2 = (a^2 + b^2)^{1/2}$.

**1.22** Show that

$$\|\mathsf{A}\|_\infty = \max_{1 \le i \le n} \sum_{j=1}^{n} |a_{ij}|, \quad \forall \mathsf{A} \in \mathbb{C}^{n \times n},$$

and also that $\|\mathsf{A}\|_1 = \left\|\mathsf{A}^{\mathsf{H}}\right\|_\infty$.

**1.23** Show that, for every $\mathsf{A} \in \mathbb{C}^{n \times n}$,

$$\frac{1}{\sqrt{n}}\|\mathsf{A}\|_2 \le \|\mathsf{A}\|_\infty \le \sqrt{n}\|\mathsf{A}\|_2.$$

**1.24** Show that

$$\|\mathsf{A}\|_2^2 \le \|\mathsf{A}\|_1 \|\mathsf{A}\|_\infty, \quad \forall \mathsf{A} \in \mathbb{C}^{n \times n}.$$

**1.25** Show that, for every $\mathsf{A} \in \mathbb{C}^{n \times n}$,

$$\|\mathsf{A}\|_{\max} \le \|\mathsf{A}\|_\infty \le n \|\mathsf{A}\|_{\max}, \tag{1.2}$$

where $\|\cdot\|_\infty$ is the induced matrix $\infty$-norm, and recall that

$$\|\mathsf{A}\|_{\max} = \max_{1 \le i,j \le n} |a_{i,j}|$$

is the matrix max norm.

**1.26** Let $\mathsf{A} \in \mathbb{R}^{n \times n}$ be such that $\mathsf{A}^{\intercal} = \mathsf{A}$ and $\operatorname{tr} \mathsf{A} = 0$. Show that

$$\|\mathsf{A}\|_2^2 \le \frac{n-1}{n} \|\mathsf{A}\|_F^2.$$

Is the assumption that $\operatorname{tr} \mathsf{A} = 0$ essential? You may justify your answer with an example or counterexample.

**1.27** Prove Proposition 1.34.

**1.28** Prove Theorem 1.37.

**1.29** Show that, for every $\mathsf{A} \in \mathbb{C}^{n \times n}$,

$$\|\mathsf{A}\|_2 = \max_{\lambda \in \sigma(\mathsf{A}^{\mathsf{H}}\mathsf{A})} \sqrt{\lambda}.$$

*Hint:* You need some facts about the eigenvalues and eigenvectors of Hermitian matrices.

**1.30** Suppose that $\|\cdot\| : \mathbb{C}^{m \times n} \to \mathbb{R}$ is the induced norm with respect to the vector norms $\|\cdot\|_{\mathbb{C}^m}$ and $\|\cdot\|_{\mathbb{C}^n}$ and that $\mathsf{A} \in \mathbb{C}^{m \times n}$. Prove that the function $\|\mathsf{A}(\cdot)\|_{\mathbb{C}^m} : \mathbb{C}^n \to \mathbb{R}$ is uniformly continuous. Use this fact to prove that there is a vector $\boldsymbol{x} \in \mathbb{S}_{\mathbb{C}^n}^{n-1}$ such that

$$\|\mathsf{A}\| = \|\mathsf{A}\boldsymbol{x}\|_{\mathbb{C}^m}.$$

https://doi.org/10.1017/9781108942607.004 Published online by Cambridge University Press

**1.31** Suppose that $\|\cdot\| : \mathbb{C}^{n\times n} \to \mathbb{R}$ is the induced norm with respect to the vector norm $\|\cdot\| : \mathbb{C}^n \to \mathbb{R}$. Let $\mathsf{A} \in \mathbb{C}^{n\times n}$ be invertible. Prove that

$$\frac{1}{\|\mathsf{A}^{-1}\|} = \min_{\boldsymbol{y}\in\mathbb{C}^n_\star} \frac{\|\mathsf{A}\boldsymbol{y}\|}{\|\boldsymbol{y}\|} .$$

**1.32** Let $\mathsf{T}_k, \mathsf{T} \in \mathbb{C}^{n\times n}$, $k = 1, 2$, be lower triangular matrices.

a) Show that $\mathsf{T}_1\mathsf{T}_2$ is lower triangular.
b) If $\mathsf{T}_1$ and $\mathsf{T}_2$ are unit lower triangular, show that $\mathsf{T}_1\mathsf{T}_2$ is unit lower triangular.
c) If $[\mathsf{T}]_{i,i} \neq 0$, show that $\mathsf{T}$ is invertible and $\mathsf{T}^{-1}$ is lower triangular.
d) If $\mathsf{T}$ is unit lower triangular, prove that it is invertible and $\mathsf{T}^{-1}$ is unit lower triangular.
e) If $[\mathsf{T}]_{i,i} > 0$, show that $\left[\mathsf{T}^{-1}\right]_{i,i} = \frac{1}{[\mathsf{T}]_{i,i}} > 0$.

**1.33** Show that if $\mathsf{A} \in \mathbb{C}^{n\times n}$ is both unitary (i.e., $\mathsf{A}\mathsf{A}^{\mathsf{H}} = \mathsf{A}^{\mathsf{H}}\mathsf{A} = \mathsf{I}_n$) and triangular, then it is diagonal.

*Hint:* You need a fact about the inverse of a triangular matrix.

**1.34** Prove Proposition 1.43.

**1.35** Let $\mathsf{A} \in \mathbb{C}^{n\times n}$. Suppose that $\lambda_1, \ldots, \lambda_k$ are distinct eigenvalues of $\mathsf{A}$ and suppose that $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ are eigenvectors associated with the respective eigenvalues. Prove that $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\}$ is linearly independent.

**1.36** Let $\mathsf{A} \in \mathbb{C}^{n\times n}$. Prove that if $\mathsf{A}$ has $n$ distinct eigenvalues, then $\mathsf{A}$ is diagonalizable.

**1.37** Let $\mathsf{A} \in \mathbb{C}^{n\times n}$ be Hermitian, i.e., $\mathsf{A}^{\mathsf{H}} = \mathsf{A}$.

a) Prove directly that all eigenvalues of $\mathsf{A}$ are real.
b) Prove that if $\boldsymbol{x}$ and $\boldsymbol{y}$ are eigenvectors associated with distinct eigenvalues, then they are orthogonal, i.e., $\boldsymbol{x}^{\mathsf{H}}\boldsymbol{y} = 0$.

**1.38** Let $\mathsf{A} \in \mathbb{C}^{m\times n}$ and $\mathsf{B} \in \mathbb{C}^{n\times m}$. Show that $\sigma(\mathsf{AB})\backslash\{0\} = \sigma(\mathsf{BA})\backslash\{0\}$, i.e., the nonzero eigenvalues of $\mathsf{AB}$ and $\mathsf{BA}$ coincide.

**1.39** Let $\mathsf{A} \in \mathbb{C}^{m\times n}$. Show that $\sigma(\mathsf{A}^{\mathsf{H}}\mathsf{A}) \cup \sigma(\mathsf{A}\mathsf{A}^{\mathsf{H}}) \subseteq [0, \infty)$.

**1.40** Let $\mathsf{A} \in \mathbb{C}^{n\times n}$ and $\lambda \in \sigma(\mathsf{A})$. The vector $\boldsymbol{y} \in \mathbb{C}^n_\star$ is called a left eigenvector associated with $\lambda$ if and only if $\boldsymbol{y}^{\mathsf{H}}\mathsf{A} = \lambda\boldsymbol{y}^{\mathsf{H}}$. Now suppose that $\lambda$, $\mu \in \sigma(\mathsf{A})$ are distinct. Let $\boldsymbol{y}$ be a left eigenvector associated with $\lambda$ and $\boldsymbol{x}$ be a right (usual) eigenvector associated with $\mu$. Prove that $\boldsymbol{y}^{\mathsf{H}}\boldsymbol{x} = 0$.

**1.41** Let $\mathsf{A} \in \mathbb{C}^{n\times n}$ be skew-Hermitian.

a) Prove directly that the eigenvalues of $\mathsf{A}$ are purely imaginary.
b) Prove that if $\boldsymbol{x}$ and $\boldsymbol{y}$ are eigenvectors associated with distinct eigenvalues, then they are orthogonal, i.e., $\boldsymbol{x}^{\mathsf{H}}\boldsymbol{y} = 0$.
c) Show that $\mathsf{I} - \mathsf{A}$ is nonsingular.
d) Prove that $\mathsf{Q} = (\mathsf{I} - \mathsf{A})^{-1}(\mathsf{I} + \mathsf{A})$ is unitary.

**1.42** Let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{C}^n$. Set $\mathsf{A} = \mathsf{I}_n + \boldsymbol{u}\boldsymbol{v}^{\mathsf{H}} \in \mathbb{C}^{n\times n}$.

a) Suppose that $\mathsf{A}$ is invertible. Prove that $\mathsf{A}^{-1} = \mathsf{I}_n + \alpha\boldsymbol{u}\boldsymbol{v}^{\mathsf{H}}$ for some $\alpha \in \mathbb{C}$. Give an expression for $\alpha$.
b) For what $\boldsymbol{u}$ and $\boldsymbol{v}$ is $\mathsf{A}$ singular, i.e., not invertible?
c) Suppose that $\mathsf{A}$ is singular. What is the kernel space of $\mathsf{A}$, $\ker(\mathsf{A})$, in this case?

**1.43** Suppose that $\boldsymbol{q} \in \mathbb{C}^n$, $\|\boldsymbol{q}\|_2 = 1$. Set $\mathsf{P} = \mathsf{I} - \boldsymbol{q}\boldsymbol{q}^{\mathsf{H}}$.

a) Find $\mathrm{im}(\mathsf{P})$.

b) Find ker(P).

c) Find the eigenvalues of P.

**1.44** Characterize the eigenvalues of a unitary matrix.

**1.45** Prove Lemma 1.51.

**1.46** Prove Theorem 1.52.

https://doi.org/10.1017/9781108942607.004 Published online by Cambridge University Press