5 Matrices

So far, our theoretical investigation has been dealing mostly with abstract vectors and abstract operators. As we have seen in examples and problems, concrete representations of vectors and operators are necessary in most applications. Such representations are obtained by choosing a basis and expressing all operations in terms of components of vectors and matrix representations of operators.

5.1 Representing Vectors and Operators

Let us choose a basis $B_V = \{|a_i\rangle\}_{i=1}^N$ of a vector space $\mathcal{V}_N$, and express an arbitrary vector $|x\rangle$ in this basis: $|x\rangle = \sum_{i=1}^N \xi_i |a_i\rangle$. We write

$$\mathsf{x} = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_N \end{pmatrix} \tag{5.1}$$

and say that the column vector x **represents** $|x\rangle$ **in** B_V. We can also have a linear transformation $\mathbf{A} \in \mathcal{L}(\mathcal{V}_N, \mathcal{W}_M)$ act on the basis vectors in B_V to give vectors in the M-dimensional vector space $\mathcal{W}_M$: $|w_k\rangle = \mathbf{A}|a_k\rangle$. The latter can be written as a linear combination of basis vectors $B_W = \{|b_j\rangle\}_{j=1}^M$ in $\mathcal{W}_M$:

representation of vectors

$$|w_1\rangle = \sum_{j=1}^M \alpha_{j1}|b_j\rangle, \qquad |w_2\rangle = \sum_{j=1}^M \alpha_{j2}|b_j\rangle, \quad \ldots, \quad |w_N\rangle = \sum_{j=1}^M \alpha_{jN}|b_j\rangle.$$

Note that the components have an extra subscript to denote which of the N vectors $\{|w_i\rangle\}_{i=1}^N$ they are representing. The components can be arranged in a column as before to give a representation of the corresponding vectors:

$$\mathsf{w}_1 = \begin{pmatrix} \alpha_{11} \\ \alpha_{21} \\ \vdots \\ \alpha_{M1} \end{pmatrix}, \qquad \mathsf{w}_2 = \begin{pmatrix} \alpha_{12} \\ \alpha_{22} \\ \vdots \\ \alpha_{M2} \end{pmatrix}, \quad \ldots, \quad \mathsf{w}_N = \begin{pmatrix} \alpha_{1N} \\ \alpha_{2N} \\ \vdots \\ \alpha_{MN} \end{pmatrix}.$$

S. Hassani, *Mathematical Physics*, DOI 10.1007/978-3-319-01195-0_5,

The operator itself is determined by the collection of all these vectors, i.e., by a matrix. We write this as

$$\mathsf{A} = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1N} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2N} \\ \vdots & \vdots & & \vdots \\ \alpha_{M1} & \alpha_{M2} & \dots & \alpha_{MN} \end{pmatrix} \tag{5.2}$$

representation of operators

and call A the **matrix representing $\mathbf{A}$ in bases B_V and B_W**. This statement is also summarized symbolically as

$$\mathbf{A}|a_i\rangle = \sum_{j=1}^{M} \alpha_{ji}|b_j\rangle, \quad i = 1, 2, \dots, N. \tag{5.3}$$

We thus have the following rule:

Box 5.1.1 *To find the matrix A representing $\mathbf{A}$ in bases $B_V = \{|a_i\rangle\}_{i=1}^N$ and $B_W = \{|b_j\rangle\}_{j=1}^M$, express $\mathbf{A}|a_i\rangle$ as a linear combination of the vectors in B_W. The components form the ith column of A.*

Now consider the vector $|y\rangle = \mathbf{A}|x\rangle$ in $\mathcal{W}_M$. This vector can be written in two ways: On the one hand, $|y\rangle = \sum_{j=1}^M \eta_j|b_j\rangle$. On the other hand,

$$\begin{aligned} |y\rangle &= \mathbf{A}|x\rangle = \mathbf{A}\sum_{i=1}^{N} \xi_i|a_i\rangle = \sum_{i=1}^{N} \xi_i \mathbf{A}|a_i\rangle \\ &= \sum_{i=1}^{N} \xi_i \left(\sum_{j=1}^{M} \alpha_{ji}|b_j\rangle\right) = \sum_{j=1}^{M} \left(\sum_{i=1}^{N} \xi_i \alpha_{ji}\right)|b_j\rangle. \end{aligned}$$

Since $|y\rangle$ has a unique set of components in the basis B_W, we conclude that

$$\eta_j = \sum_{i=1}^{N} \alpha_{ji}\xi_i, \quad j = 1, 2, \dots, M. \tag{5.4}$$

This is written as

$$\begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_M \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1N} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2N} \\ \vdots & \vdots & & \vdots \\ \alpha_{M1} & \alpha_{M2} & \dots & \alpha_{MN} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_N \end{pmatrix} \quad \Rightarrow \quad \mathsf{y} = \mathsf{Ax}, \tag{5.5}$$

in which the usual matrix multiplication rule is understood. This matrix equation is the representation of the operator equation $|y\rangle = \mathbf{A}|x\rangle$ in the bases B_V and B_W.

The construction above indicates that—once the bases are fixed in the two vector spaces—to every operator there corresponds a *unique* matrix. This uniqueness is the result of the uniqueness of the components of vectors in a basis. On the other hand, given an $M \times N$ matrix A with elements α_{ij}, one can construct a unique linear operator $\mathbf{T}_A$ defined by its action on the basis vectors (see Box 2.3.6): $\mathbf{T}_A|a_i\rangle \equiv \sum_{j=1}^{M} \alpha_{ji}|b_j\rangle$. Thus, there is a one-to-one correspondence between operators and matrices. This correspondence is in fact a linear isomorphism:

The operator $\mathbf{T}_A$ associated with a matrix A

Proposition 5.1.2 *The two vector spaces $\mathcal{L}(\mathcal{V}_N, \mathcal{W}_M)$ and $\mathcal{M}^{M\times N}$ are isomorphic. An explicit isomorphism is established only when a basis is chosen for each vector space, in which case, an operator is identified with its matrix representation.*

Example 5.1.3 In this example, we construct a matrix representation of the complex structure $\mathbf{J}$ on a real vector space $\mathcal{V}$ introduced in Sect. 2.4. There are two common representations, each corresponding to a different ordering of the vectors in the basis $\{|e_i\rangle, \mathbf{J}|e_i\rangle\}_{i=1}^{m}$ of $\mathcal{V}$. One ordering is to let $\mathbf{J}|e_i\rangle$ come right after $|e_i\rangle$. The other is to collect all the $\mathbf{J}|e_i\rangle$ after the $|e_i\rangle$ in the same order. We consider the first ordering in this example, and leave the other for the reader to construct.

matrix representation of the complex structure $\mathbf{J}$

In the first ordering, for each $|e_i\rangle$, we let $|e_{i+1}\rangle = \mathbf{J}|e_i\rangle$. Starting with $|e_1\rangle$, we have

$$\mathbf{J}|e_1\rangle = |e_2\rangle = 0\cdot|e_1\rangle + 1\cdot|e_2\rangle + 0\cdot|e_3\rangle + \cdots + 0\cdot|e_{2m}\rangle,$$
$$\mathbf{J}|e_2\rangle = \mathbf{J}^2|e_1\rangle = -|e_1\rangle = -1\cdot|e_1\rangle + 0\cdot|e_2\rangle + 0\cdot|e_3\rangle + \cdots + 0\cdot|e_{2m}\rangle.$$

These two equations give the first two columns as

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}.$$

For the third and fourth basis vectors, we get

$$\mathbf{J}|e_3\rangle = |e_4\rangle = 0\cdot|e_1\rangle + 0\cdot|e_2\rangle + 0\cdot|e_3\rangle + 1\cdot|e_4\rangle + \cdots + 0\cdot|e_{2m}\rangle$$
$$\mathbf{J}|e_4\rangle = \mathbf{J}^2|e_3\rangle = -|e_3\rangle = 0\cdot|e_1\rangle + 0\cdot|e_2\rangle - 1\cdot|e_3\rangle + \cdots + 0\cdot|e_{2m}\rangle,$$

giving rise to the following third and fourth columns:

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & -1 \\ 1 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}.$$

It should now be clear that the matrix representation of $\mathbf{J}$ is of the form

$$\mathsf{J} = \begin{pmatrix} \mathsf{R}_1 & 0 & \dots & 0 \\ 0 & \mathsf{R}_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \mathsf{R}_m \end{pmatrix},$$

where the zeros are the 2×2 zero matrices and $\mathsf{R}_k = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ for all k.

Notation 5.1.4 *Let $\mathbf{A} \in \mathcal{L}(\mathcal{V}_N, \mathcal{W}_M)$. Choose a bases B_V for $\mathcal{V}$ and B_W for $\mathcal{W}$. We denote the matrix representing $\mathbf{A}$ in these bases by $\mathbf{M}_{B_V}^{B_W}(\mathbf{A})$, where*

$$\mathbf{M}_{B_V}^{B_W} : \mathcal{L}(\mathcal{V}_N, \mathcal{W}_M) \to \mathcal{M}^{M \times N}$$

is the basis-dependent linear isomorphism. When $\mathcal{V} = \mathcal{W}$, we leave out the subscripts and superscripts of $\mathbf{M}$, keeping in mind that all matrices are representations in a single basis.

Given the linear transformations $\mathbf{A} : \mathcal{V}_N \to \mathcal{W}_M$ and $\mathbf{B} : \mathcal{W}_M \to \mathcal{U}_K$, we can form the composite linear transformation $\mathbf{B} \circ \mathbf{A} : \mathcal{V}_N \to \mathcal{U}_K$. We can also choose bases $B_V = \{|a_i\rangle\}_{i=1}^N$, $B_W = \{|b_i\rangle\}_{i=1}^M$, $B_U = \{|c_i\rangle\}_{i=1}^K$ for $\mathcal{V}$, $\mathcal{W}$, and $\mathcal{U}$, respectively. Then $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{B} \circ \mathbf{A}$ will be represented by an $M \times N$, a $K \times M$, and a $K \times N$ matrix, respectively, and we have

$$\mathbf{M}_{B_V}^{B_U}(\mathbf{B} \circ \mathbf{A}) = \mathbf{M}_{B_W}^{B_U}(\mathbf{B})\mathbf{M}_{B_V}^{B_W}(\mathbf{A}), \tag{5.6}$$

where on the right-hand side the product is defined as the usual product of matrices. If $\mathcal{V} = \mathcal{W} = \mathcal{U}$, we write (5.6) as

$$\mathbf{M}(\mathbf{B} \circ \mathbf{A}) = \mathbf{M}(\mathbf{B})\mathbf{M}(\mathbf{A}) \tag{5.7}$$

Matrices are determined entirely by their elements. For this reason a matrix A whose elements are $\alpha_{11}, \alpha_{12}, \dots$ is sometimes denoted by (α_{ij}). Similarly, the elements of this matrix are denoted by $(\mathsf{A})_{ij}$. So, on the one hand, we have $(\alpha_{ij}) = \mathsf{A}$, and on the other hand $(\mathsf{A})_{ij} = \alpha_{ij}$. In the context of this notation, therefore, we can write

$$\begin{aligned} (\mathsf{A} + \mathsf{B})_{ij} = (\mathsf{A})_{ij} + (\mathsf{B})_{ij} &\Rightarrow (\alpha_{ij} + \beta_{ij}) = (\alpha_{ij}) + (\beta_{ij}), \\ (\gamma \mathsf{A})_{ij} = \gamma(\mathsf{A})_{ij} &\Rightarrow \gamma(\alpha_{ij}) = (\gamma\alpha_{ij}), \\ (\mathsf{0})_{ij} &= 0, \\ (\mathsf{1})_{ij} &= \delta_{ij}. \end{aligned}$$

A matrix, as a representation of a linear operator, is well-defined only in reference to a specific basis. A collection of rows and columns of numbers by themselves have no operational meaning. When we manipulate matrices and attach meaning to them, we make an unannounced assumption regarding the basis: We have the standard basis of $\mathbb{C}^n$ (or $\mathbb{R}^n$) in mind. The following example should clarify this subtlety.

Example 5.1.5 Let us find the matrix representation of the linear operator $\mathbf{A} \in \mathcal{L}(\mathbb{R}^3)$, given by

$$\mathbf{A}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x - y + 2z \\ 3x - z \\ 2y + z \end{pmatrix} \tag{5.8}$$

in the basis

$$B = \left\{ |a_1\rangle = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix},\ |a_2\rangle = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix},\ |a_3\rangle = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\}.$$

There is a tendency to associate the matrix

$$\begin{pmatrix} 1 & -1 & 2 \\ 3 & 0 & -1 \\ 0 & 2 & 1 \end{pmatrix}$$

with the operator $\mathbf{A}$. The following discussion will show that this is false.

To obtain the first column of the matrix representing $\mathbf{A}$, we note that

$$\begin{aligned} \mathbf{A}|a_1\rangle = \mathbf{A}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} &= \begin{pmatrix} 0 \\ 3 \\ 2 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} - \frac{1}{2}\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + \frac{5}{2}\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \\ &= \frac{1}{2}|a_1\rangle - \frac{1}{2}|a_2\rangle + \frac{5}{2}|a_3\rangle. \end{aligned}$$

So, by Box 5.1.1, the first column of the matrix is

$$\begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{5}{2} \end{pmatrix}.$$

The other two columns are obtained from

$$\mathbf{A}|a_2\rangle = \mathbf{A}\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} = 2\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + 0\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix},$$

$$\mathbf{A}|a_3\rangle = \mathbf{A}\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 3 \end{pmatrix} = -\frac{3}{2}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \frac{5}{2}\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix},$$

giving the second and the third columns, respectively. The whole matrix is then

$$\mathsf{A} = \begin{pmatrix} \frac{1}{2} & 2 & -\frac{3}{2} \\ -\frac{1}{2} & 1 & \frac{5}{2} \\ \frac{5}{2} & 0 & \frac{1}{2} \end{pmatrix}.$$

As long as all vectors are represented by columns whose entries are expansion coefficients of the vectors in B, $\mathbf{A}$ and A are indistinguishable. However, the action of A on the column vector $\left(\begin{smallmatrix} x \\ y \\ z \end{smallmatrix}\right)$ will not yield the RHS of Eq. (5.8)! Although this is not usually emphasized, the column vector on the LHS of Eq. (5.8) is really the vector

$$x\begin{pmatrix}1\\0\\0\end{pmatrix} + y\begin{pmatrix}0\\1\\0\end{pmatrix} + z\begin{pmatrix}0\\0\\1\end{pmatrix},$$

which is an expansion in terms of the *standard basis* of $\mathbb{R}^3$ rather than in terms of B.

We can expand $\mathbf{A}\left(\begin{smallmatrix} x \\ y \\ z \end{smallmatrix}\right)$ in terms of B, yielding

$$\begin{aligned}\mathbf{A}\begin{pmatrix}x\\y\\z\end{pmatrix} &= \begin{pmatrix}x-y+2z\\3x-z\\2y+z\end{pmatrix}\\ &= \left(2x-\frac{3}{2}y\right)\begin{pmatrix}1\\1\\0\end{pmatrix} + \left(-x+\frac{1}{2}y+2z\right)\begin{pmatrix}1\\0\\1\end{pmatrix}\\ &\quad + \left(x+\frac{3}{2}y-z\right)\begin{pmatrix}0\\1\\1\end{pmatrix}.\end{aligned}$$

This says that in the basis B this vector has the representation

$$\left(\mathbf{A}\begin{pmatrix}x\\y\\z\end{pmatrix}\right)_B = \begin{pmatrix}2x-\frac{3}{2}y\\-x+\frac{1}{2}y+2z\\x+\frac{3}{2}y-z\end{pmatrix}. \tag{5.9}$$

Similarly, $\left(\begin{smallmatrix} x \\ y \\ z \end{smallmatrix}\right)$ is represented by

$$\left(\begin{pmatrix}x\\y\\z\end{pmatrix}\right)_B = \begin{pmatrix}\frac{1}{2}x+\frac{1}{2}y-\frac{1}{2}z\\\frac{1}{2}x-\frac{1}{2}y+\frac{1}{2}z\\-\frac{1}{2}x+\frac{1}{2}y+\frac{1}{2}z\end{pmatrix}. \tag{5.10}$$

Applying A to the RHS of (5.10) yields the RHS of (5.9), as it should.

5.2 Operations on Matrices

There are two basic operations that one can perform on a matrix to obtain a new one; these are transposition and complex conjugation. The **transpose**

transpose of a matrix

of an $M \times N$ matrix A is an $N \times M$ matrix A^t obtained by interchanging the rows and columns of A:

$$\left(\mathsf{A}^t\right)_{ij} = (\mathsf{A})_{ji}, \quad \text{or} \quad (\alpha_{ij})^t = (\alpha_{ji}). \tag{5.11}$$

The following theorem, whose proof follows immediately from the definition of transpose, summarizes the important properties of the operation of transposition.

Theorem 5.2.1 *Let* A *and* B *be two matrices for which the operation of addition and/or multiplication are defined. Then*

(a) $(\mathsf{A} + \mathsf{B})^t = \mathsf{A}^t + \mathsf{B}^t$,
(b) $(\mathsf{AB})^t = \mathsf{B}^t \mathsf{A}^t$,
(c) $(\mathsf{A}^t)^t = \mathsf{A}$.

Let $\mathbf{T} \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ and $B_V = \{|a_i\rangle\}_{i=1}^N$ and $B_W = \{|b_i\rangle\}_{i=1}^M$ bases in $\mathcal{V}$ and $\mathcal{W}$. Then

$$\mathbf{T}|a_i\rangle = \sum_{j=1}^{M} (\mathsf{T})_{ji} |b_j\rangle,$$

where $\mathsf{T} = \mathbf{M}_{B_V}^{B_W}(\mathbf{T})$. Let $\mathbf{T}^* \in \mathcal{L}(\mathcal{W}^*, \mathcal{V}^*)$ be the pull-back of $\mathbf{T}$ and $B_V^* = \{\boldsymbol{\theta}_k\}_{k=1}^N$ and $B_W^* = \{\boldsymbol{\phi}_l\}_{l=1}^M$ bases dual to B_V and B_W. Then

$$\mathbf{T}^* \boldsymbol{\phi}_l = \sum_{k=1}^{N} \left(\mathsf{T}^*\right)_{kl} \boldsymbol{\theta}_k.$$

Apply both sides of this equation to $|a_i\rangle$ to get

$$\begin{aligned}
\text{LHS} &= \left(\mathbf{T}^* \boldsymbol{\phi}_l\right)|a_i\rangle \equiv \boldsymbol{\phi}_l\left(\mathbf{T}|a_i\rangle\right) \\
&= \boldsymbol{\phi}_l\left(\sum_{j=1}^{M} (\mathsf{T})_{ji} |b_j\rangle\right) = \sum_{j=1}^{M} (\mathsf{T})_{ji} \overbrace{\boldsymbol{\phi}_l\left(|b_j\rangle\right)}^{=\delta_{lj}} = (\mathsf{T})_{li}
\end{aligned}$$

and

$$\text{RHS} = \sum_{k=1}^{N} \left(\mathsf{T}^*\right)_{kl} \boldsymbol{\theta}_k |a_i\rangle = \sum_{k=1}^{N} \left(\mathsf{T}^*\right)_{kl} \delta_{ki} = \left(\mathsf{T}^*\right)_{il}.$$

Comparing the last two equations, we have

Matrix of pullback of **T** is transpose of matrix of **T**.

Proposition 5.2.2 *Let* $\mathbf{T} \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ *and* B_V *and* B_W *be bases in* $\mathcal{V}$ *and* $\mathcal{W}$. *Let* $\mathbf{T}^*$, B_V^*, *and* B_W^* *be duals to* $\mathbf{T}$, B_V, *and* B_W, *respectively. Let* $\mathsf{T} = \mathbf{M}_{B_V}^{B_W}(\mathbf{T})$ *and* $\mathsf{T}^* = \mathbf{M}_{B_W^*}^{B_V^*}(\mathbf{T}^*)$. *Then* $\mathsf{T}^* = \mathsf{T}^t$.

Of special interest is a matrix that is equal to either its transpose or the negative of its transpose. Such matrices occur frequently in physics.

symmetric and antisymmetric matrices

Definition 5.2.3 A matrix S is **symmetric** if $\mathsf{S}^t = \mathsf{S}$. Similarly, a matrix A is **antisymmetric** if $\mathsf{A}^t = -\mathsf{A}$.

Any matrix A can be written as $\mathsf{A} = \frac{1}{2}(\mathsf{A} + \mathsf{A}^t) + \frac{1}{2}(\mathsf{A} - \mathsf{A}^t)$, where the first term is symmetric and the second is antisymmetric.

The elements of a symmetric matrix A satisfy the relation $\alpha_{ji} = (\mathsf{A}^t)_{ij} = (\mathsf{A})_{ij} = \alpha_{ij}$; i.e., the matrix is symmetric under reflection through the main diagonal. On the other hand, for an antisymmetric matrix we have $\alpha_{ji} = -\alpha_{ij}$. In particular, the diagonal elements of an antisymmetric matrix are all zero.

orthogonal matrix

A (real) matrix satisfying $\mathsf{A}^t\mathsf{A} = \mathsf{A}\mathsf{A}^t = 1$ is called **orthogonal**.

complex conjugation

Complex conjugation is an operation under which all elements of a matrix are complex conjugated. Denoting the complex conjugate of A by A^*, we have $(\mathsf{A}^*)_{ij} = (\mathsf{A})^*_{ij}$, or $(\alpha_{ij})^* = (\alpha^*_{ij})$. A matrix is real if and only if $\mathsf{A}^* = \mathsf{A}$. Clearly, $(\mathsf{A}^*)^* = \mathsf{A}$.

hermitian conjugate

Under the combined operation of complex conjugation and transposition, the rows and columns of a matrix are interchanged and all of its elements are complex conjugated. This combined operation is called the **adjoint** operation, or **hermitian conjugation**, and is denoted by †, as with operators. Thus, we have

$$\begin{aligned} \mathsf{A}^\dagger &= \left(\mathsf{A}^t\right)^* = \left(\mathsf{A}^*\right)^t, \\ \left(\mathsf{A}^\dagger\right)_{ij} &= (\mathsf{A})^*_{ji} \quad \text{or} \quad (\alpha_{ij})^\dagger = \left(\alpha^*_{ji}\right). \end{aligned} \tag{5.12}$$

Two types of matrices are important enough to warrant a separate definition.

hermitian and unitary matrices

Definition 5.2.4 A **hermitian** matrix H satisfies $\mathsf{H}^\dagger = \mathsf{H}$, or, in terms of elements, $\eta^*_{ij} = \eta_{ji}$. A **unitary** matrix U satisfies $\mathsf{U}^\dagger\mathsf{U} = \mathsf{U}\mathsf{U}^\dagger = 1$, or, in terms of elements, $\sum_{k=1}^N \mu_{ik}\mu^*_{jk} = \sum_{k=1}^N \mu^*_{ki}\mu_{kj} = \delta_{ij}$.

Remarks It follows immediately from this definition that

1. The diagonal elements of a hermitian matrix are real.
2. The kth column of a hermitian matrix is the complex conjugate of its kth row, and vice versa.
3. A real hermitian matrix is symmetric.
4. The rows of an $N \times N$ unitary matrix, when considered as vectors in $\mathbb{C}^N$, form an orthonormal set, as do the columns.
5. A real unitary matrix is orthogonal.

It is sometimes possible (and desirable) to transform a matrix into a form in which all of its off-diagonal elements are zero. Such a matrix is called a **diagonal** matrix.

diagonal matrices

Box 5.2.5 *A diagonal matrix whose diagonal elements are* $\{\lambda_k\}_{k=1}^N$ *is denoted by* $\operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_N)$.

Example 5.2.6 In this example, we derive a useful identity for functions of a diagonal matrix. Let $\mathsf{D} = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ be a diagonal matrix, and $f(x)$ a function that has a Taylor series expansion $f(x) = \sum_{k=0}^{\infty} a_k x^k$. The same function of D can be written as

$$\begin{aligned} f(\mathsf{D}) &= \sum_{k=0}^{\infty} a_k \mathsf{D}^k = \sum_{k=0}^{\infty} a_k \bigl[\mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)\bigr]^k \\ &= \sum_{k=0}^{\infty} a_k\, \mathrm{diag}\bigl(\lambda_1^k, \lambda_2^k, \dots, \lambda_n^k\bigr) \\ &= \mathrm{diag}\left(\sum_{k=0}^{\infty} a_k \lambda_1^k, \sum_{k=0}^{\infty} a_k \lambda_2^k, \dots, \sum_{k=0}^{\infty} a_k \lambda_n^k\right) \\ &= \mathrm{diag}\bigl(f(\lambda_1), f(\lambda_2), \dots, f(\lambda_n)\bigr). \end{aligned}$$

In words, the function of a diagonal matrix is equal to a diagonal matrix whose entries are the same function of the corresponding entries of the original matrix. In the above derivation, we used the following obvious properties of diagonal matrices:

$$\begin{aligned} &a\, \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) = \mathrm{diag}(a\lambda_1, a\lambda_2, \dots, a\lambda_n), \\ &\mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) + \mathrm{diag}(\omega_1, \omega_2, \dots, \omega_n) \\ &\quad = \mathrm{diag}(\lambda_1 + \omega_1, \dots, \lambda_n + \omega_n), \\ &\mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) \cdot \mathrm{diag}(\omega_1, \omega_2, \dots, \omega_n) = \mathrm{diag}(\lambda_1\omega_1, \dots, \lambda_n\omega_n). \end{aligned}$$

Example 5.2.7 In this example, we list some familiar matrices in physics.

(a) A prototypical symmetric matrix is that of the moment of inertia encountered in mechanics. The ijth element of this matrix is defined as $I_{ij} \equiv \iiint \rho(x_1, x_2, x_3) x_i x_j \, dV$, where x_i is the ith Cartesian coordinate of a point in the distribution of mass described by the volume density $\rho(x_1, x_2, x_3)$. It is clear that $I_{ij} = I_{ji}$, or $\mathsf{I} = \mathsf{I}^t$. The moment of inertia matrix can be represented as

$$\mathsf{I} = \begin{pmatrix} I_{11} & I_{12} & I_{13} \\ I_{12} & I_{22} & I_{23} \\ I_{13} & I_{23} & I_{33} \end{pmatrix}.$$

It has six independent elements.

(b) An example of an antisymmetric matrix is the electromagnetic field tensor given by

$$\mathsf{F} = \begin{pmatrix} 0 & -B_3 & B_2 & E_1 \\ B_3 & 0 & -B_1 & E_2 \\ -B_2 & B_1 & 0 & E_3 \\ -E_1 & -E_2 & -E_3 & 0 \end{pmatrix}.$$

(c) Examples of hermitian matrices are the 2×2 **Pauli spin matrices**:

Pauli spin matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

(d) The most frequently encountered orthogonal matrices are rotations. One such matrix represents the rotation of a 3-dimensional rigid body in terms of **Euler angles** and is used in mechanics. Attaching a coordinate system to the body, a general rotation can be decomposed into a rotation of angle φ about the z-axis, followed by a rotation of angle θ about the *new* x-axis, followed by a rotation of angle ψ about the *new* z-axis. We simply exhibit this matrix in terms of these angles and leave it to the reader to show that it is indeed orthogonal.

Euler angles

$$\begin{pmatrix} \cos\psi\cos\varphi - \sin\psi\cos\theta\sin\varphi & -\cos\psi\sin\varphi - \sin\psi\cos\theta\cos\varphi & \sin\psi\sin\theta \\ \sin\psi\cos\varphi + \cos\psi\cos\theta\sin\varphi & -\sin\psi\sin\varphi + \cos\psi\cos\theta\cos\varphi & -\cos\psi\sin\theta \\ \sin\theta\sin\varphi & \sin\theta\cos\varphi & \cos\theta \end{pmatrix}.$$

5.3 Orthonormal Bases

The matrix representation of $\mathbf{A} \in \operatorname{End}(\mathcal{V})$ is facilitated by choosing an *orthonormal* basis $B = \{|e_i\rangle\}_{i=1}^N$. The matrix elements of $\mathbf{A}$ can be found in such a basis by "multiplying" both sides of $\mathbf{A}|e_i\rangle = \sum_{k=1}^N \alpha_{ki}|e_k\rangle$ on the left by $\langle e_j|$:

$$\langle e_j|\mathbf{A}|e_i\rangle = \langle e_j|\left(\sum_{k=1}^N \alpha_{ki}|e_k\rangle\right) = \sum_{k=1}^N \alpha_{ki} \underbrace{\langle e_j|e_k\rangle}_{=\delta_{jk}} = \alpha_{ji},$$

or

$$(\mathsf{A})_{ij} = \alpha_{ij} = \langle e_i|\mathbf{A}|e_j\rangle. \tag{5.13}$$

We can also show that in an orthonormal basis, the ith component ξ_i of a vector is found by multiplying the vector by $\langle e_i|$. This expression for ξ_i allows us to write the expansion of $|x\rangle$ as

$$|x\rangle = \sum_{j=1}^N \underbrace{\langle e_j|x\rangle}_{\xi_j}|e_j\rangle = \sum_{j=1}^N |e_j\rangle\langle e_j|x\rangle \quad \Rightarrow \quad \mathbf{1} = \sum_{j=1}^N |e_j\rangle\langle e_j|, \tag{5.14}$$

which is the same as in Proposition 4.4.6.

Let us now investigate the representation of the special operators discussed in Chap. 4 and find the connection between those operators and the matrices encountered in the last section. We begin by calculating the matrix representing the hermitian conjugate of an operator $\mathbf{T}$. In an orthonormal basis, the elements of this matrix are given by Eq. (5.13), $\tau_{ij} = \langle e_i|\mathbf{T}|e_j\rangle$. Taking the complex conjugate of this equation and using the definition of $\mathbf{T}^\dagger$ given in Eq. (4.11), we obtain

$$\tau_{ij}^* = \langle e_i|\mathbf{T}|e_j\rangle^* = \langle e_j|\mathbf{T}^\dagger|e_i\rangle, \quad \text{or} \quad \left(\mathsf{T}^\dagger\right)_{ij} = \tau_{ji}^*.$$

This is precisely how the adjoint of a matrix was defined. Note how crucially this conclusion depends on the orthonormality of the basis vectors. If the basis were not orthonormal, we could not use Eq. (5.13) on which the conclusion is based. Therefore,

> **Box 5.3.1** *Only in an orthonormal basis is the adjoint of an operator represented by the adjoint of the matrix representing that operator.*

In particular, a hermitian operator is represented by a hermitian matrix only if an orthonormal basis is used. The following example illustrates this point.

Example 5.3.2 Consider the matrix representation of the hermitian operator $\mathbf{H}$ in a general—not orthonormal—basis $B = \{|a_i\rangle\}_{i=1}^N$. The elements of the matrix corresponding to $\mathbf{H}$ are given by

$$\mathbf{H}|a_k\rangle = \sum_{j=1}^N \eta_{jk}|a_j\rangle, \quad \text{or} \quad \mathbf{H}|a_i\rangle = \sum_{j=1}^N \eta_{ji}|a_j\rangle. \tag{5.15}$$

Taking the product of the first equation with $\langle a_i|$ and complex-conjugating the result gives

$$\langle a_i|\mathbf{H}|a_k\rangle^* = \left(\sum_{j=1}^N \eta_{jk}\langle a_i|a_j\rangle\right)^* = \sum_{j=1}^N \eta_{jk}^*\langle a_j|a_i\rangle.$$

But by the definition of a hermitian operator,

$$\langle a_i|\mathbf{H}|a_k\rangle^* = \langle a_k|\mathbf{H}^\dagger|a_i\rangle = \langle a_k|\mathbf{H}|a_i\rangle.$$

So we have $\langle a_k|\mathbf{H}|a_i\rangle = \sum_{j=1}^N \eta_{jk}^*\langle a_j|a_i\rangle$.

On the other hand, multiplying the second equation in (5.15) by $\langle a_k|$ gives

$$\langle a_k|\mathbf{H}|a_i\rangle = \sum_{j=1}^N \eta_{ji}\langle a_k|a_j\rangle.$$

The only conclusion we can draw from this discussion is

$$\sum_{j=1}^N \eta_{jk}^*\langle a_j|a_i\rangle = \sum_{j=1}^N \eta_{ji}\langle a_k|a_j\rangle.$$

Because this equation does not say anything about each individual η_{ij}, we cannot conclude, in general, that $\eta_{ij}^* = \eta_{ji}$. However, if the $|a_i\rangle$'s are orthonormal, then $\langle a_j|a_i\rangle = \delta_{ji}$ and $\langle a_k|a_j\rangle = \delta_{kj}$, and we obtain $\sum_{j=1}^N \eta_{jk}^*\delta_{ji} = \sum_{j=1}^N \eta_{ji}\delta_{kj}$, or $\eta_{ik}^* = \eta_{ki}$, as expected of a hermitian matrix.

Similarly, we expect the matrices representing unitary operators to be unitary only if the basis is orthonormal. This is an immediate consequence of Eq. (5.12), but we shall prove it in order to provide yet another example of how the completeness relation, Eq. (5.14), is used. Since $\mathbf{U}\mathbf{U}^\dagger = \mathbf{1}$, we have

$$\langle e_i|\mathbf{U}\mathbf{U}^\dagger|e_j\rangle = \langle e_i|\mathbf{1}|e_j\rangle = \langle e_i|e_j\rangle = \delta_{ij}.$$

We insert the completeness relation $\mathbf{1} = \sum_{k=1}^N |e_k\rangle\langle e_k|$ between $\mathbf{U}$ and $\mathbf{U}^\dagger$ on the LHS:

$$\langle e_i|\mathbf{U}\left(\sum_{k=1}^N |e_k\rangle\langle e_k|\right)\mathbf{U}^\dagger|e_j\rangle = \sum_{k=1}^N \underbrace{\langle e_i|\mathbf{U}|e_k\rangle}_{\equiv\mu_{ik}} \underbrace{\langle e_k|\mathbf{U}^\dagger|e_j\rangle}_{\equiv\mu^*_{jk}} = \delta_{ij}.$$

This equation gives the first half of the requirement for a unitary matrix given in Definition 5.2.4. By redoing the calculation for $\mathbf{U}^\dagger\mathbf{U}$, we could obtain the second half of that requirement.

5.4 Change of Basis

It is often advantageous to describe a physical problem in a particular basis because it takes a simpler form there, but the general form of the result may still be of importance. In such cases the problem is solved in one basis, and the result is transformed to other bases. Let us investigate this point in some detail.

Given a basis $B = \{|a_i\rangle\}_{i=1}^N$, we can write an arbitrary vector $|a\rangle$ with components $\{\alpha_i\}_{i=1}^N$ in B as $|a\rangle = \sum_{i=1}^N \alpha_i|a_i\rangle$. Now suppose that we change the basis to $B' = \{|a'_j\rangle\}_{j=1}^N$. How are the components of $|a\rangle$ in B' related to those in B? To answer this question, we write $|a_i\rangle$ in terms of B' vectors,

$$|a_i\rangle = \sum_{j=1}^N \rho_{ji}|a'_j\rangle, \quad i = 1, 2, \ldots, N,$$

which can also be abbreviated as

$$\begin{pmatrix} |a_1\rangle \\ |a_2\rangle \\ \vdots \\ |a_N\rangle \end{pmatrix} = \begin{pmatrix} \rho_{11} & \rho_{21} & \cdots & \rho_{N1} \\ \rho_{12} & \rho_{22} & \cdots & \rho_{N2} \\ \vdots & \vdots & & \vdots \\ \rho_{1N} & \rho_{2N} & \cdots & \rho_{NN} \end{pmatrix} \begin{pmatrix} |a'_1\rangle \\ |a'_2\rangle \\ \vdots \\ |a'_N\rangle \end{pmatrix}. \tag{5.16}$$

In this notation, we also have

$$|a\rangle = \begin{pmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_N \end{pmatrix} \begin{pmatrix} |a_1\rangle \\ |a_2\rangle \\ \vdots \\ |a_N\rangle \end{pmatrix} \equiv \mathsf{a}^t \begin{pmatrix} |a_1\rangle \\ |a_2\rangle \\ \vdots \\ |a_N\rangle \end{pmatrix},$$

where a is the column representation of $|a\rangle$ in B. Now multiply both sides of (5.16) by a^t to get

$$|a\rangle = \mathsf{a}^t \begin{pmatrix} |a_1\rangle \\ |a_2\rangle \\ \vdots \\ |a_N\rangle \end{pmatrix} \equiv \mathsf{a}^t \mathsf{R}^t \begin{pmatrix} |a'_1\rangle \\ |a'_2\rangle \\ \vdots \\ |a'_N\rangle \end{pmatrix} \equiv \underbrace{\begin{pmatrix} \alpha'_1 & \alpha'_2 & \dots & \alpha'_N \end{pmatrix}}_{\equiv \mathsf{a}'^t} \begin{pmatrix} |a'_1\rangle \\ |a'_2\rangle \\ \vdots \\ |a'_N\rangle \end{pmatrix}$$

where R is the transpose of the $N \times N$ matrix of Eq. (5.16), and the last equality expresses $|a\rangle$ in B'. We therefore conclude that

$$\mathsf{a}'^t \equiv \mathsf{a}^t \mathsf{R}^t,$$

where a' designates a column vector with elements α'_j, the components of $|a\rangle$ in B'. Taking the transpose of the last equation yields

$$\mathsf{a}' = \mathsf{R}\mathsf{a} \quad \text{or} \quad \begin{pmatrix} \alpha'_1 \\ \alpha'_2 \\ \vdots \\ \alpha'_N \end{pmatrix} = \begin{pmatrix} \rho_{11} & \rho_{12} & \dots & \rho_{1N} \\ \rho_{21} & \rho_{22} & \dots & \rho_{2N} \\ \vdots & \vdots & & \vdots \\ \rho_{N1} & \rho_{N2} & \dots & \rho_{NN} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{pmatrix}, \tag{5.17}$$

which in component form can be written as

$$\alpha'_j = \sum_{i=1}^{N} \rho_{ji} \alpha_i \quad \text{for } j = 1, 2, \dots, N. \tag{5.18}$$

The matrix R is called the **basis transformation matrix**. It is invertible because it is a linear transformation that maps one basis onto another (see Proposition 4.1.3).

basis transformation matrix

What happens to a matrix representation of an operator when we transform the basis? Consider the equation $|b\rangle = \mathbf{A}|a\rangle$, where $|a\rangle$ and $|b\rangle$ have components $\{\alpha_i\}_{i=1}^N$ and $\{\beta_i\}_{i=1}^N$, respectively, in B. This equation has a corresponding matrix equation $\mathsf{b} = \mathsf{A}\mathsf{a}$. Now, if we change the basis, the columns of the components of $|a\rangle$ and $|b\rangle$ will change to those of a' and b', respectively. We seek a matrix A' such that $\mathsf{b}' = \mathsf{A}'\mathsf{a}'$. This matrix will be the transform of A. Using Eq. (5.17), we write $\mathsf{R}\mathsf{b} = \mathsf{A}'\mathsf{R}\mathsf{a}$, or $\mathsf{b} = \mathsf{R}^{-1}\mathsf{A}'\mathsf{R}\mathsf{a}$. Comparing this with $\mathsf{b} = \mathsf{A}\mathsf{a}$ and applying the fact that both equations hold for arbitrary a and b, we conclude that

$$\mathsf{R}^{-1}\mathsf{A}'\mathsf{R} = \mathsf{A}, \quad \text{or} \quad \mathsf{A}' = \mathsf{R}\mathsf{A}\mathsf{R}^{-1}. \tag{5.19}$$

This is called a **similarity transformation** on A, and A' is said to be **similar** to A.

similarity transformation

The transformation matrix R can easily be found for orthonormal bases $B = \{|e_i\rangle\}_{i=1}^N$ and $B' = \{|e'_i\rangle\}_{i=1}^N$. We have $|e_i\rangle = \sum_{k=1}^N \rho_{ki} |e'_k\rangle$. Multiplying this equation by $\langle e'_j|$, we obtain

$$\langle e'_j \mid e_i\rangle = \sum_{k=1}^{N} \rho_{ki} \langle e'_j \mid e'_k\rangle = \sum_{k=1}^{N} \rho_{ki} \delta_{jk} = \rho_{ji}. \tag{5.20}$$

That is,

> **Box 5.4.1** *To find the ijth element of the matrix that changes the components of a vector in the orthonormal basis B to those in the orthonormal basis B', take the jth ket in B and multiply it by the ith bra in B'.*

To find the ijth element of the matrix that changes B' into B, we take the jth ket in B' and multiply it by the ith bra in B: $\rho'_{ij} = \langle e_i | e'_j \rangle$. However, the matrix R' must be R^{-1}, as can be seen from Eq. (5.17). On the other hand, $(\rho'_{ij})^* = \langle e_i | e'_j \rangle^* = \langle e'_j | e_i \rangle = \rho_{ji}$, or

$$\left(\mathsf{R}^{-1}\right)^*_{ij} = \rho_{ji}, \quad \text{or} \quad \left(\mathsf{R}^{-1}\right)_{ij} = \rho^*_{ji} = \left(\mathsf{R}^\dagger\right)_{ij}. \tag{5.21}$$

This shows that R is a unitary matrix and yields an important result.

Theorem 5.4.2 *The matrix that transforms one orthonormal basis into another is necessarily unitary.*

From Eqs. (5.20) and (5.21) we have $(\mathsf{R}^\dagger)_{ij} = \langle e_i | e'_j \rangle$. Thus,

> **Box 5.4.3** *To obtain the jth column of $\mathsf{R}^\dagger$, we take the jth vector in the* new *basis and successively "multiply" it by $\langle e_i|$ for $i = 1, 2, \dots, N$.*

In particular, if the original basis is the standard basis of $\mathbb{C}^N$ and $|e'_j\rangle$ is represented by a column vector in that basis, then the jth column of $\mathsf{R}^\dagger$ is simply the vector $|e'_j\rangle$.

Example 5.4.4 In this example, we show that the similarity transform of a function of a matrix is the same function of the similarity transform of the matrix:

$$\mathsf{R} f(\mathsf{A})\mathsf{R}^{-1} = f\left(\mathsf{R}\mathsf{A}\mathsf{R}^{-1}\right).$$

The proof involves inserting $\mathsf{1} = \mathsf{R}^{-1}\mathsf{R}$ between factors of A in the Taylor series expansion of $f(\mathsf{A})$:

$$\begin{aligned}
\mathsf{R} f(\mathsf{A})\mathsf{R}^{-1} &= \mathsf{R}\left(\sum_{k=0}^\infty a_k \mathsf{A}^k\right)\mathsf{R}^{-1} = \sum_{k=0}^\infty a_k \mathsf{R}\mathsf{A}^k\mathsf{R}^{-1} = \sum_{k=0}^\infty a_k \mathsf{R}\overbrace{\mathsf{A}\mathsf{A}\cdots\mathsf{A}}^{k \text{ times}}\mathsf{R}^{-1} \\
&= \sum_{k=0}^\infty a_k \overbrace{\mathsf{R}\mathsf{A}\mathsf{R}^{-1}\mathsf{R}\mathsf{A}\mathsf{R}^{-1}\cdots\mathsf{R}\mathsf{A}\mathsf{R}^{-1}}^{k \text{ times}} = \sum_{k=0}^\infty a_k \left(\mathsf{R}\mathsf{A}\mathsf{R}^{-1}\right)^k \\
&= f\left(\mathsf{R}\mathsf{A}\mathsf{R}^{-1}\right).
\end{aligned}$$

This completes the proof.

5.5 Determinant of a Matrix

An important concept associated with linear operators is the determinant, which we have already discussed in Sect. 2.6.1. Determinants are also defined for matrices. If A is representing $\mathbf{A}$ in some basis, then we set $\det\mathsf{A} = \det\mathbf{A}$. That this relation is basis-independent is, of course, obvious from Definition 2.6.10 and the discussion preceding it. However, it can also be shown directly, as we shall do later in this chapter.

Let $\mathbf{A}$ be a linear operator on $\mathcal{V}$. Let $\{|e_k\rangle\}_{k=1}^N$ be a basis of $\mathcal{V}$ in which $\mathbf{A}$ is represented by A. Then the left-hand side of Eq. (2.32) becomes

$$\begin{aligned}\text{LHS} &= \boldsymbol{\Delta}\big(\mathbf{A}|e_1\rangle, \dots, \mathbf{A}|e_N\rangle\big) = \boldsymbol{\Delta}\left(\sum_{i_1=1}^N \alpha_{i_1 1}|e_{i_1}\rangle, \dots, \sum_{i_N=1}^N \alpha_{i_N N}|e_{i_N}\rangle\right) \\ &= \sum_{i_1 \dots i_N=1}^N \alpha_{i_1 1}\dots\alpha_{i_N N}\boldsymbol{\Delta}\big(|e_{i_1}\rangle, \dots, |e_{i_N}\rangle\big) \\ &= \sum_\pi \alpha_{\pi(1)1}\dots\alpha_{\pi(N)N}\boldsymbol{\Delta}\big(|e_{\pi(1)}\rangle, \dots, |e_{\pi(N)}\rangle\big) \\ &= \sum_\pi \alpha_{\pi(1)1}\dots\alpha_{\pi(N)N}\epsilon_\pi \cdot \boldsymbol{\Delta}\big(|e_1\rangle, \dots, |e_N\rangle\big),\end{aligned}$$

where π is the permutation taking k to i_k. The right-hand side of Eq. (2.32) is just the product of $\det\mathbf{A}$ and $\boldsymbol{\Delta}(|e\rangle_1, \dots, |e\rangle_N)$. Hence,

$$\det\mathbf{A} = \det\mathsf{A} = \sum_\pi \epsilon_\pi \alpha_{\pi(1)1}\dots\alpha_{\pi(N)N} \equiv \sum_\pi \epsilon_\pi \prod_{k=1}^N (\mathsf{A})_{\pi(k)k}. \tag{5.22}$$

Since $\pi(k) = i_k$, the product in the sum can be written as

$$\prod_{k=1}^N (\mathsf{A})_{i_k k} = \prod_{k=1}^N (\mathsf{A})_{i_k \pi^{-1}(i_k)} = \prod_{k=1}^N (\mathsf{A})_{k\pi^{-1}(k)} = \prod_{k=1}^N \big(\mathsf{A}^t\big)_{\pi^{-1}(k)k},$$

where the second equality follows because we can commute the numbers until $(\mathsf{A})_{1\pi^{-1}(1)}$ becomes the first term of the product, $(\mathsf{A})_{2\pi^{-1}(2)}$, the second term, and so on. Substituting this in (5.22) and noting that $\sum_\pi = \sum_{\pi^{-1}}$ and $\epsilon_{\pi^{-1}} = \epsilon_\pi$, we have

Theorem 5.5.1 *Let $\mathbf{A} \in \mathcal{L}(\mathcal{V})$ and A its representation in any basis of $\mathcal{V}$. Then*

$$\det\mathbf{A} = \det\mathsf{A} = \sum_\pi \epsilon_\pi \prod_{j=1}^N (\mathsf{A})_{\pi(j)j} = \sum_{i_1\dots i_N} \varepsilon_{i_1 i_2 \dots i_N}(\mathsf{A})_{i_1 1}\dots(\mathsf{A})_{i_N N}$$

$$\det\mathbf{A} = \det\mathsf{A}^t = \sum_\pi \epsilon_\pi \prod_{j=1}^N (\mathsf{A})_{j\pi(j)} = \sum_{i_1\dots i_N} \varepsilon_{i_1 i_2 \dots i_N}(\mathsf{A})_{1 i_1}\dots(\mathsf{A})_{N i_N}$$

where $\varepsilon_{i_1 i_2 \dots i_N}$ is the symbol introduced in (2.29). *In particular,* $\det\mathsf{A}^t = \det\mathsf{A}$.

Let A be any $N \times N$ matrix. Let $|v_j\rangle \in \mathbb{R}^N$ be the jth column of A. Define the linear operator $\mathbf{A} \in \operatorname{End}(\mathbb{R}^N)$ by

$$\mathbf{A}|e_j\rangle = |v_j\rangle, \quad j = 1, \ldots, N, \tag{5.23}$$

where $\{|e_j\rangle\}_{j=1}^N$ is the standard basis of $\mathbb{R}^N$. Then A is the matrix representing $\mathbf{A}$ in the standard basis. Now let $\boldsymbol{\Delta}$ be a determinant function in $\mathbb{R}^N$ whose value is one at the standard basis. Then

$$\begin{aligned}\boldsymbol{\Delta}\big(|v_1\rangle, \ldots, |v_N\rangle\big) &= \boldsymbol{\Delta}\big(\mathbf{A}|e_1\rangle, \ldots, \mathbf{A}|e_N\rangle\big)\\ &= \det \mathbf{A} \cdot \boldsymbol{\Delta}\big(|e_1\rangle, \ldots, |e_N\rangle\big) = \det \mathbf{A}\end{aligned}$$

and, therefore,

$$\det \mathsf{A} = \boldsymbol{\Delta}\big(|v_1\rangle, \ldots, |v_N\rangle\big). \tag{5.24}$$

If instead of columns, we use rows $|u_j\rangle$, we obtain $\det \mathsf{A}^t = \boldsymbol{\Delta}(|u_1\rangle, \ldots, |u_N\rangle)$. Since $\boldsymbol{\Delta}$ is a multilinear skew-symmetric function, and $\det \mathsf{A}^t = \det \mathsf{A}$, we have the following familiar theorem.

Theorem 5.5.2 *Let* A *be a square matrix. Then*

1. $\det \mathsf{A}$ *is linear with respect to any row or column vector of* A.
2. *If any two rows or two columns of* A *are interchanged*, $\det \mathsf{A}$ *changes sign*.
3. *Adding a multiple of one row* (*column*) *of* A *to another row* (*column*) *of* A *does not change* $\det \mathsf{A}$.
4. $\det \mathsf{A} = 0$ *iff the rows* (*columns*) *are linearly dependent*.

5.5.1 Matrix of the Classical Adjoint

Since by Corollary 2.6.13, the classical adjoint of $\mathbf{A}$ is essentially the inverse of $\mathbf{A}$, we expect its matrix representation to be essentially the inverse of the matrix of $\mathbf{A}$. To find this matrix, choose a basis $\{|e_j\rangle\}_{j=1}^N$ which evaluates the determinant function of Eq. (2.33) to 1. Then $\operatorname{ad}(\mathbf{A})|e_i\rangle = c_{ji}|e_j\rangle$, with c_{ji} forming the representation matrix of $\operatorname{ad}(\mathbf{A})$. Thus, substituting $|e_i\rangle$ for $|v\rangle$ on both sides of (2.33) and using the fact that $\{|e_j\rangle\}_{j=1}^N$ are linearly independent, we get

$$(-1)^{j-1}\boldsymbol{\Delta}\big(|e_i\rangle, \mathbf{A}|e_1\rangle, \ldots, \widehat{\mathbf{A}|e_j\rangle}, \ldots, \mathbf{A}|e_N\rangle\big) = c_{ji}$$

or

$$\begin{aligned}c_{ji} &= (-1)^{j-1}\boldsymbol{\Delta}\left(|e_i\rangle, \sum_{k_1=1}^N (\mathsf{A})_{k_1 1}|e_{k_1}\rangle, \ldots, \widehat{\mathbf{A}|e_j\rangle}, \ldots, \sum_{k_N=1}^N (\mathsf{A})_{k_N N}|e_{k_N}\rangle\right)\\ &= (-1)^{j-1}\sum_{k_1 \ldots k_N} (\mathsf{A})_{k_1 1} \ldots (\mathsf{A})_{k_N N}\boldsymbol{\Delta}\big(|e_i\rangle, |e_{k_1}\rangle, \ldots, |e_{k_N}\rangle\big)\\ &= (-1)^{j-1}\sum_{k_1 \ldots k_N} (\mathsf{A})_{k_1 1} \ldots (\mathsf{A})_{k_N N}\epsilon_{i k_1 \ldots k_N}.\end{aligned}$$

The product in the sum does not include $(\mathsf{A})_{k_j j}$. This means that the entire jth column is missing in the product. Furthermore, because of the skew-symmetry of $\epsilon_{ik_1 \dots k_N}$, none of the k_m's can be i, and since k_m's label the rows, the ith row is also absent in the sum. Now move i from the first location to the ith location. This will introduce a factor of $(-1)^{i-1}$ due to the $i-1$ exchanges of indices. Inserting all this information in the previous equation, we obtain

$$c_{ji} = (-1)^{i+j} \sum_{k_1 \dots k_N} (\mathsf{A})_{k_1 1} \dots (\mathsf{A})_{k_N N} \epsilon_{k_1 \dots i \dots k_N}. \tag{5.25}$$

Now note that the sum is a determinant of an $(N-1) \times (N-1)$ matrix obtained from A by eliminating its ith row and jth column. This determinant is called a **minor of order** $N-1$ and denoted by M_{ij}. The product $(-1)^{i+j} M_{ij}$ is called the **cofactor of** $(\mathsf{A})_{ij}$, and denoted by $(\operatorname{cof} \mathsf{A})_{ij}$.

minor of order $N-1$
cofactor of an element of a matrix

With this and another obvious notation, (5.25) becomes

$$(\operatorname{ad} \mathsf{A})_{ji} \equiv c_{ji} = (-1)^{i+j} M_{ij} = (\operatorname{cof} \mathsf{A})_{ij}. \tag{5.26}$$

With the matrix of the adjoint at our disposal, we can write Eq. (2.34) in the matrix form. Doing so, and taking the ikth element of all sides, we get

$$\sum_{j=1}^{N} \operatorname{ad}(\mathsf{A})_{ij} (\mathsf{A})_{jk} = \det \mathbf{A} \cdot \delta_{ik} = \sum_{j=1}^{N} (\mathsf{A})_{ij} \operatorname{ad}(\mathsf{A})_{jk}.$$

Setting $k = i$ yields

$$\det \mathbf{A} = \sum_{j=1}^{N} \operatorname{ad}(\mathsf{A})_{ij} (\mathsf{A})_{ji} = \sum_{j=1}^{N} (\mathsf{A})_{ij} \operatorname{ad}(\mathsf{A})_{ji}$$

or, using (5.26),

$$\det \mathbf{A} = \sum_{j=1}^{N} (\mathsf{A})_{ji} (\operatorname{cof} \mathsf{A})_{ji} = \sum_{j=1}^{N} (\mathsf{A})_{ij} (\operatorname{cof} \mathsf{A})_{ij}. \tag{5.27}$$

This is the familiar expansion of a determinant by its ith column or ith row.

Historical Notes

Vandermonde, Alexandre-Thiéophile, also known as Alexis, Abnit, and Charles-Auguste Vandermonde (1735–1796) had a father, a physician who directed his sickly son toward a musical career. An acquaintanceship with Fontaine, however, so stimulated Vandermonde that in 1771 he was elected to the Académie des Sciences, to which he presented four mathematical papers (his total mathematical production) in 1771–1772. Later, Vandermonde wrote several papers on harmony, and it was said at that time that musicians considered Vandermonde to be a mathematician and that mathematicians viewed him as a musician.

Vandermonde's membership in the Academy led to a paper on experiments with cold, made with Bezout and Lavoisier in 1776, and a paper on the manufacture of steel with Berthollet and Monge in 1786. Vandermonde became an ardent and active revolutionary, being such a close friend of Monge that he was termed "femme de Monge". He was a member of the Commune of Paris and the club of the Jacobins. In 1782 he was director of the Conservatoire des Arts et Métiers and in 1792, chief of the Bureau de l'Habillement des Armies. He joined in the design of a course in political economy for the École Normale and in 1795 was named a member of the Institut National.

Vandermonde is best known for the theory of determinants. Lebesgue believed that the attribution of determinant to Vandermonde was due to a misreading of his notation. Nevertheless, Vandermonde's fourth paper was the first to give a connected exposition of determinants, because he (1) defined a contemporary symbolism that was more complete, simple, and appropriate than that of Leibniz; (2) defined determinants as functions apart from the solution of linear equations presented by Cramer but also treated by Vandermonde; and (3) gave a number of properties of these functions, such as the number and signs of the terms and the effect of interchanging two consecutive indices (rows or columns), which he used to show that a determinant is zero if two rows or columns are identical.

Vandermonde's real and unrecognized claim to fame was lodged in his first paper, in which he approached the general problem of the solvability of algebraic equations through a study of functions invariant under permutations of the roots of the equations. Cauchy assigned priority in this to Lagrange and Vandermonde. Vandermonde read his paper in November 1770, but he did not become a member of the Academy until 1771, and the paper was not published until 1774. Although Vandermonde's methods were close to those later developed by Abel and Galois for testing the solvability of equations, and although his treatment of the binomial equation $x^n - 1 = 0$ could easily have led to the anticipation of Gauss's results on constructible polygons, Vandermonde himself did not rigorously or completely establish his results, nor did he see the implications for geometry. Nevertheless, Kronecker dates the modern movement in algebra to Vandermonde's 1770 paper.

Unfortunately, Vandermonde's spurt of enthusiasm and creativity, which in two years produced four insightful mathematical papers at least two of which were of substantial importance, was quickly diverted by the exciting politics of the time and perhaps by poor health.

Example 5.5.3 Let O and U denote, respectively, an orthogonal and a unitary $n \times n$ matrix; that is, $\mathsf{O}\mathsf{O}^t = \mathsf{O}^t\mathsf{O} = \mathbf{1}$, and $\mathsf{U}\mathsf{U}^\dagger = \mathsf{U}^\dagger\mathsf{U} = \mathbf{1}$. Taking the determinant of the first equation and using Theorems 2.6.11 (with $\lambda = 1$) and 5.5.1, we obtain

$$(\det \mathsf{O})\big(\det \mathsf{O}^t\big) = (\det \mathsf{O})^2 = \det \mathbf{1} = 1.$$

Therefore, for an orthogonal matrix, we get $\det \mathsf{O} = \pm 1$.

Orthogonal transformations preserve a real inner product. Among such transformations are the so-called inversions, which, in their simplest form, multiply a vector by -1. In three dimensions this corresponds to a reflection through the origin. The matrix associated with this operation is $-\mathbf{1}$:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \to \begin{pmatrix} -x \\ -y \\ -z \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

which has a determinant of -1. This is a prototype of other, more complicated, orthogonal transformations whose determinants are -1. The set of orthogonal matrices in n dimensions is denoted by $O(n)$.

$O(n)$ and $SO(n)$

The other orthogonal transformations, whose determinants are $+1$, are of special interest because they correspond to rotations in three dimensions. The set of orthogonal matrices in n dimensions having determinant $+1$ is denoted by $SO(n)$. These matrices are special because they have the mathematical structure of a (continuous) group, which finds application in many areas of advanced physics. We shall come back to the topic of group theory later in the book.

We can obtain a similar result for unitary transformations. We take the determinant of both sides of $\mathsf{U}^\dagger\mathsf{U} = \mathbf{1}$:

$$\det\bigl(\mathsf{U}^*\bigr)^t \det\mathsf{U} = \det\mathsf{U}^* \det\mathsf{U} = (\det\mathsf{U})^*(\det\mathsf{U}) = |\det\mathsf{U}|^2 = 1.$$

Thus, we can generally write $\det\mathsf{U} = e^{i\alpha}$, with $\alpha \in \mathbb{R}$. The set of unitary matrices in n dimensions is denoted by $U(n)$. The set of those matrices with $\alpha = 0$ forms a group to which $\mathbf{1}$ belongs and that is denoted by $SU(n)$. This group has found applications in the description of fundamental forces and the dynamics of fundamental particles.

$U(n)$ and $SU(n)$

5.5.2 Inverse of a Matrix

Equation (5.26) shows that the matrix of the classical adjoint is the transpose of the cofactor matrix. Using this, and writing (2.34) in matrix form yields

$$(\operatorname{cof}\mathsf{A})^t\mathsf{A} = \det\mathbf{A}\cdot\mathbf{1} = \mathsf{A}(\operatorname{cof}\mathsf{A})^t.$$

Therefore, we have

Theorem 5.5.4 *The matrix* A *has an inverse if and only if* $\det\mathsf{A} \neq 0$. *Furthermore*,

inverse of a matrix

$$\mathsf{A}^{-1} = \frac{(\operatorname{cof}\mathsf{A})^t}{\det\mathsf{A}}. \tag{5.28}$$

This is the matrix form of the operator equation in Corollary 2.6.13.

Example 5.5.5 The inverse of a 2×2 matrix is easily found:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \tag{5.29}$$

if $ad - bc \neq 0$.

We defined the determinant of an operator intrinsically, i.e., independent of a basis. We have also connected this intrinsic property to the determinant of the matrix representing that operator in some basis. We can now show directly that the matrices representing an operator in two arbitrary bases have the same determinant. We leave this as exercise for the reader in Problem 5.23.

Algorithm for Calculating the Inverse of a Matrix

There is a more practical way of calculating the inverse of matrices. In the following discussion of this method, we shall confine ourselves simply to stating a couple of definitions and the main theorem, with no attempt at providing any proofs. The practical utility of the method will be illustrated by a detailed analysis of examples.

elementary row operation

Definition 5.5.6 An **elementary row operation** on a matrix is one of the following:

(a) interchange of two rows of the matrix,
(b) multiplication of a row by a nonzero number, and
(c) addition of a multiple of one row to another.

Elementary column operations are defined analogously.

triangular, or row-echelon form of a matrix

Definition 5.5.7 A matrix is in **triangular**, or row-echelon, form if it satisfies the following three conditions:

1. Any row consisting of only zeros is below any row that contains at least one nonzero element.
2. Going from left to right, the first nonzero entry of any row is to the left of the first nonzero entry of any lower row.
3. The first nonzero entry of each row is 1.

Theorem 5.5.8 *For any invertible $n \times n$ matrix* A, *the $n \times 2n$ matrix* $(\mathsf{A}|\mathsf{1})$ *can be transformed into the $n \times 2n$ matrix* $(\mathsf{1}|\mathsf{A}^{-1})$ *by means of a finite number of elementary row operations.*[1]

A systematic way of transforming $(\mathsf{A}|\mathsf{1})$ into $(\mathsf{1}|\mathsf{A}^{-1})$ is first to bring A into triangular form and then eliminate all nonzero elements of each column by elementary row operations.

Example 5.5.9
Let us evaluate the inverse of

$$\mathsf{A} = \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & -2 \\ 2 & 1 & -1 \end{pmatrix}.$$

We start with

$$\left(\begin{array}{ccc|ccc} 1 & 2 & -1 & 1 & 0 & 0 \\ 0 & 1 & -2 & 0 & 1 & 0 \\ 2 & 1 & -1 & 0 & 0 & 1 \end{array}\right) \equiv \mathsf{M}$$

and apply elementary row operations to M to bring the left half of it into triangular form. If we denote the kth row by (k) and the three operations of Definition 5.5.6, respectively, by $(k) \leftrightarrow (j)$, $\alpha(k)$, and $\alpha(k) + (j)$, we get

$$\mathsf{M} \xrightarrow[-2(1)+(3)]{} \left(\begin{array}{ccc|ccc} 1 & 2 & -1 & 1 & 0 & 0 \\ 0 & 1 & -2 & 0 & 1 & 0 \\ 0 & -3 & 1 & -2 & 0 & 1 \end{array}\right)$$

[1]The matrix $(\mathsf{A}|\mathsf{1})$ denotes the $n \times 2n$ matrix obtained by juxtaposing the $n \times n$ unit matrix to the right of A. It can easily be shown that if A, B, and C are $n \times n$ matrices, then $\mathsf{A}(\mathsf{B}|\mathsf{C}) = (\mathsf{AB}|\mathsf{AC})$.

$$\xrightarrow[3(2)+(3)]{} \left(\begin{array}{ccc|ccc} 1 & 2 & -1 & 1 & 0 & 0 \\ 0 & 1 & -2 & 0 & 1 & 0 \\ 0 & 0 & -5 & -2 & 3 & 1 \end{array}\right)$$

$$\xrightarrow[-\frac{1}{5}(3)]{} \left(\begin{array}{ccc|ccc} 1 & 2 & -1 & 1 & 0 & 0 \\ 0 & 1 & -2 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2/5 & -3/5 & -1/5 \end{array}\right) \equiv \mathsf{M}'.$$

The left half of M' is in triangular form. However, we want all entries above any 1 in a column to be zero as well, i.e., we want the left-hand matrix to be $\mathbf{1}$. We can do this by appropriate use of type 3 elementary row operations:

$$\mathsf{M}' \xrightarrow[-2(2)+(1)]{} \left(\begin{array}{ccc|ccc} 1 & 0 & 3 & 1 & -2 & 0 \\ 0 & 1 & -2 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2/5 & -3/5 & -1/5 \end{array}\right)$$

$$\xrightarrow[-3(3)+(1)]{} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -1/5 & -1/5 & 3/5 \\ 0 & 1 & -2 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2/5 & -3/5 & -1/5 \end{array}\right)$$

$$\xrightarrow[2(3)+(2)]{} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -1/5 & -1/5 & 3/5 \\ 0 & 1 & 0 & 4/5 & -1/5 & -2/5 \\ 0 & 0 & 1 & 2/5 & -3/5 & -1/5 \end{array}\right).$$

The right half of the resulting matrix is A^{-1}.

Example 5.5.10 It is instructive to start with a matrix that is not invertible and show that it is impossible to turn it into $\mathbf{1}$ by elementary row operations. Consider the matrix

$$\mathsf{B} = \begin{pmatrix} 2 & -1 & 3 \\ 1 & -2 & 1 \\ -1 & 5 & 0 \end{pmatrix}.$$

Let us systematically bring it into triangular form:

$$\mathsf{M} = \left(\begin{array}{ccc|ccc} 2 & -1 & 3 & 1 & 0 & 0 \\ 1 & -2 & 1 & 0 & 1 & 0 \\ -1 & 5 & 0 & 0 & 0 & 1 \end{array}\right) \xrightarrow[(1)\leftrightarrow(2)]{} \left(\begin{array}{ccc|ccc} 1 & -2 & 1 & 0 & 1 & 0 \\ 2 & -1 & 3 & 1 & 0 & 0 \\ -1 & 5 & 0 & 0 & 0 & 1 \end{array}\right)$$

$$\xrightarrow[-2(1)+(2)]{} \left(\begin{array}{ccc|ccc} 1 & -2 & 1 & 0 & 1 & 0 \\ 0 & 3 & 1 & 1 & -2 & 0 \\ -1 & 5 & 0 & 0 & 0 & 1 \end{array}\right)$$

$$\xrightarrow[(1)+(3)]{} \left(\begin{array}{ccc|ccc} 1 & -2 & 1 & 0 & 1 & 0 \\ 0 & 3 & 1 & 1 & -2 & 0 \\ 0 & 3 & 1 & 0 & 1 & 1 \end{array}\right)$$

$$\xrightarrow[-(2)+(3)]{} \left(\begin{array}{ccc|ccc} 1 & -2 & 1 & 0 & 1 & 0 \\ 0 & 3 & 1 & 1 & -2 & 0 \\ 0 & 0 & 0 & -1 & 3 & 1 \end{array}\right)$$

$$\xrightarrow[\frac{1}{3}(2)]{} \left(\begin{array}{ccc|ccc} 1 & -2 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1/3 & 1/3 & -2/3 & 0 \\ 0 & 0 & 0 & -1 & 3 & 1 \end{array}\right).$$

The matrix B is now in triangular form, but its third row contains all zeros. There is no way we can bring this into the form of a unit matrix. We therefore conclude that B is not invertible. This is, of course, obvious, since it can easily be verified that B has a vanishing determinant.

Rank of a Matrix

Given any $M \times N$ matrix A, an operator $\mathbf{T}_A \in \mathcal{L}(\mathcal{V}_N, \mathcal{W}_M)$ can be associated with A, and one can construct the kernel and the range of $\mathbf{T}_A$. The rank of $\mathbf{T}_A$ is called the **rank of** A. Since the rank of an operator is basis independent, this definition makes sense.

rank of a matrix

Now suppose that we choose a basis for the kernel of $\mathbf{T}_A$ and extend it to a basis of $\mathcal{V}$. Let $\mathcal{V}_1$ denote the span of the remaining basis vectors. Similarly, we choose a basis for $\mathbf{T}_A(\mathcal{V})$ and extend it to a basis for $\mathcal{W}$. In these two bases, the $M \times N$ matrix representing $\mathbf{T}_A$ will have all zeros except for an $r \times r$ submatrix, where r is the rank of $\mathbf{T}_A$. The reader may verify that this submatrix has a nonzero determinant. In fact, the submatrix represents the isomorphism between $\mathcal{V}_1$ and $\mathbf{T}_A(\mathcal{V})$, and, by its very construction, is the largest such matrix. Since the determinant of an operator is basis-independent, we have the following proposition.

Proposition 5.5.11 *The rank of a matrix is the dimension of the largest (square) submatrix whose determinant is not zero.*

5.5.3 Dual Determinant Function

Let $\mathcal{V}$ and $\mathcal{V}^*$ be N-dimensional dual vector spaces, and let $\boldsymbol{\Theta} : \mathcal{V}^N \times \mathcal{V}^{*N} \to \mathbb{C}$ be a function defined by

$$\boldsymbol{\Theta}\big(|v_1\rangle, \dots, |v_N\rangle, \boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_N\big) = \det\big(\boldsymbol{\phi}_i\big(|v_j\rangle\big)\big), \quad \boldsymbol{\phi}_i \in \mathcal{V}^*, \ |v_j\rangle \in \mathcal{V}. \tag{5.30}$$

By Theorem 5.5.2, $\boldsymbol{\Theta}$ is a skew-symmetric linear function in $|v_1\rangle, \dots, |v_N\rangle$ as well as in $\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_N$. Considering the first set of arguments and taking a nonzero determinant function $\boldsymbol{\Delta}$ in $\mathcal{V}$, we can write

$$\boldsymbol{\Theta}\big(|v_1\rangle, \dots, |v_N\rangle, \boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_N\big) = \boldsymbol{\Delta} \cdot \underbrace{\boldsymbol{\Omega}(\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_N)}_{\in \mathbb{C}}$$

by Corollary 2.6.8. We note that $\boldsymbol{\Omega}$ is a determinant function in $\mathcal{V}^*$. Thus, again by Corollary 2.6.8,

$$\boldsymbol{\Omega}(\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_N) = \beta \cdot \boldsymbol{\Delta}^*(\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_N),$$

for some nonzero determinant function $\boldsymbol{\Delta}^*$ in $\mathcal{V}^*$ and some $\beta \in \mathbb{C}$. Combining the last two equations, we obtain

$$\boldsymbol{\Theta}\big(|v_1\rangle, \ldots, |v_N\rangle, \boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_N\big) = \beta \boldsymbol{\Delta}\big(|v_1\rangle, \ldots, |v_N\rangle\big)\boldsymbol{\Delta}^*(\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_N). \tag{5.31}$$

Now let $\{\boldsymbol{\epsilon}\}_{i=1}^N$ and $\{|e_j\rangle\}_{j=1}^N$ be dual bases. Then Eq. (5.30) gives

$$\boldsymbol{\Theta}\big(|e_1\rangle, \ldots, |e_N\rangle, \boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_N\big) = \det(\delta_{ij}) = 1,$$

and Eq. (5.31) yields

$$1 = \beta \boldsymbol{\Delta}\big(|e_1\rangle, \ldots, |e_N\rangle\big)\boldsymbol{\Delta}^*(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_N).$$

This implies that $\beta \neq 0$. Multiplying both sides of (5.30) by $\alpha \equiv \beta^{-1}$ and using (5.31), we obtain

Proposition 5.5.12 *For any pair of nonzero determinant functions $\boldsymbol{\Delta}$ and $\boldsymbol{\Delta}^*$ in $\mathcal{V}$ and $\mathcal{V}^*$, respectively, there is a nonzero constant $\alpha \in \mathbb{C}$ such that*

$$\boldsymbol{\Delta}\big(|v_1\rangle, \ldots, |v_N\rangle\big)\boldsymbol{\Delta}^*(\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_N) = \alpha \det\big(\boldsymbol{\phi}_i\big(|v_j\rangle\big)\big)$$

for $|v_j\rangle \in \mathcal{V}$ and $\boldsymbol{\phi}_i \in \mathcal{V}^*$.

Definition 5.5.13 Two nonzero determinant function $\boldsymbol{\Delta}$ and $\boldsymbol{\Delta}^*$ in $\mathcal{V}$ and $\mathcal{V}^*$, respectively, are called **dual** if

dual determinant functions

$$\boldsymbol{\Delta}\big(|v_1\rangle, \ldots, |v_N\rangle\big)\boldsymbol{\Delta}^*(\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_N) = \det\big(\boldsymbol{\phi}_i\big(|v_j\rangle\big)\big).$$

It is clear that if $\boldsymbol{\Delta}$ and $\boldsymbol{\Delta}^*$ are any two determinant functions, then $\boldsymbol{\Delta}$ and $\alpha^{-1}\boldsymbol{\Delta}^*$ are dual. Furthermore, if $\boldsymbol{\Delta}_1^*$ and $\boldsymbol{\Delta}_2^*$ are dual to $\boldsymbol{\Delta}$, then $\boldsymbol{\Delta}_1^* = \boldsymbol{\Delta}_2^*$, because they both satisfy the equation of Definition 5.5.13 and $\boldsymbol{\Delta}$ is nonzero. We thus have

Proposition 5.5.14 *Every nonzero determinant function in $\mathcal{V}$ has a unique dual determinant function.*

Here is another way of proving the equality of the determinants of a matrix and its transpose:

Proposition 5.5.15 *Let $\mathbf{T}^* \in \operatorname{End}(\mathcal{V}^*)$ be the dual of $\mathbf{T} \in \operatorname{End}(\mathcal{V})$. Then $\det \mathbf{T}^* = \det \mathbf{T}$. In particular, $\det \mathsf{T}^t = \det \mathsf{T}$.*

Proof Use Definition 5.5.13 to get

$$\boldsymbol{\Delta}\big(|v_1\rangle, \ldots, |v_N\rangle\big)\boldsymbol{\Delta}^*\big(\mathbf{T}^*\boldsymbol{\phi}_1, \ldots, \mathbf{T}^*\boldsymbol{\phi}_N\big) = \det\big(\mathbf{T}^*\boldsymbol{\phi}_i\big(|v_j\rangle\big)\big)$$

or

$$\det \mathbf{T}^* \cdot \boldsymbol{\Delta}\big(|v_1\rangle, \ldots, |v_N\rangle\big)\boldsymbol{\Delta}^*(\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_N) = \det\big(\mathbf{T}^*\boldsymbol{\phi}_i\big(|v_j\rangle\big)\big).$$

Furthermore,

$$\boldsymbol{\Delta}\big(\mathbf{T}|v_1\rangle,\ldots,\mathbf{T}|v_N\rangle\big)\boldsymbol{\Delta}^*(\boldsymbol{\phi}_1,\ldots,\boldsymbol{\phi}_N)=\det\big(\boldsymbol{\phi}_i\big(\mathbf{T}|v_j\rangle\big)\big)$$

or

$$\det\mathbf{T}\cdot\boldsymbol{\Delta}\big(|v_1\rangle,\ldots,|v_N\rangle\big)\boldsymbol{\Delta}^*(\boldsymbol{\phi}_1,\ldots,\boldsymbol{\phi}_N)=\det\big(\boldsymbol{\phi}_i\big(\mathbf{T}|v_j\rangle\big)\big).$$

Now noting that $\mathbf{T}^*\boldsymbol{\phi}_i(|v_j\rangle)\equiv\boldsymbol{\phi}_i(\mathbf{T}|v_j\rangle)$, we obtain the equality of the determinant of $\mathbf{T}$ and $\mathbf{T}^*$, and by Proposition 5.2.2, the equality of the determinant of T and T^t. □

5.6 The Trace

Another intrinsic quantity associated with an operator that is usually defined in terms of matrices is given in the following definition.

trace of a square matrix

Definition 5.6.1 Let A be an $N\times N$ matrix. The mapping $\mathrm{tr}:\mathcal{M}^{N\times N}\to\mathbb{C}$ (or $\mathbb{R}$) given by $\mathrm{tr}\,\mathsf{A}=\sum_{i=1}^N\alpha_{ii}$ is called the **trace** of A.

Theorem 5.6.2 *The trace is a linear mapping. Furthermore,*

$$\mathrm{tr}(\mathsf{AB})=\mathrm{tr}(\mathsf{BA})\quad\textit{and}\quad\mathrm{tr}\,\mathsf{A}^t=\mathrm{tr}\,\mathsf{A}.$$

Proof To prove the first identity, we use the definitions of the trace and the matrix product:

$$\begin{aligned}\mathrm{tr}(\mathsf{AB})&=\sum_{i=1}^N(\mathsf{AB})_{ii}=\sum_{i=1}^N\sum_{j=1}^N(\mathsf{A})_{ij}(\mathsf{B})_{ji}=\sum_{i=1}^N\sum_{j=1}^N(\mathsf{B})_{ji}(\mathsf{A})_{ij}\\&=\sum_{j=1}^N\left(\sum_{i=1}^N(\mathsf{B})_{ji}(\mathsf{A})_{ij}\right)=\sum_{j=1}^N(\mathsf{BA})_{jj}=\mathrm{tr}(\mathsf{BA}).\end{aligned}$$

The linearity of the trace and the second identity follow directly from the definition. □

connection between trace and determinant

Example 5.6.3 In this example, we show a very useful connection between the trace and the determinant that holds when a matrix is only infinitesimally different from the unit matrix. Let us calculate the determinant of $1+\epsilon\mathsf{A}$ to first order in ϵ. Using the definition of determinant, we write

$$\begin{aligned}\det(1+\epsilon\mathsf{A})&=\sum_{i_1,\ldots,i_n=1}^n\epsilon_{i_1\ldots i_n}(\delta_{1i_1}+\epsilon\alpha_{1i_1})\ldots(\delta_{ni_n}+\epsilon\alpha_{ni_n})\\&=\sum_{i_1,\ldots,i_n=1}^n\epsilon_{i_1\ldots i_n}\delta_{1i_1}\ldots\delta_{ni_n}\\&\quad+\epsilon\sum_{k=1}^n\sum_{i_1,\ldots,i_n=1}^n\epsilon_{i_1\ldots i_n}\delta_{1i_1}\ldots\hat{\delta}_{ki_k}\ldots\delta_{ni_n}\alpha_{ki_k}.\end{aligned}$$

The first sum is just the product of all the Kronecker deltas. In the second sum, $\hat{\delta}_{ki_k}$ means that in the product of the deltas, δ_{ki_k} is absent. This term is obtained by multiplying the second term of the kth parentheses by the first term of all the rest. Since we are interested only in the first power of ϵ, we stop at this term. Now, the first sum is reduced to $\epsilon_{12\ldots n} = 1$ after all the Kronecker deltas are summed over. For the second sum, we get

$$\begin{aligned}
\epsilon \sum_{k=1}^{n} \sum_{i_1,\ldots,i_n=1}^{n} \epsilon_{i_1\ldots i_n} \delta_{1i_1} \ldots \hat{\delta}_{ki_k} \ldots \delta_{ni_n} \alpha_{ki_k} \\
&= \epsilon \sum_{k=1}^{n} \sum_{i_k=1}^{n} \epsilon_{12\ldots i_k \ldots n} \alpha_{ki_k} \\
&= \epsilon \sum_{k=1}^{n} \epsilon_{12\ldots k\ldots n} \alpha_{kk} = \epsilon \sum_{k=1}^{n} \alpha_{kk} = \epsilon \operatorname{tr} \mathsf{A},
\end{aligned} \tag{5.32}$$

where the last line follows from the fact that the only nonzero value for $\epsilon_{12\ldots i_k \ldots n}$ is obtained when i_k is equal to the missing index, i.e., k, in which case it will be 1. Thus $\det(\mathsf{1} + \epsilon \mathsf{A}) = 1 + \epsilon \operatorname{tr} \mathsf{A}$.

Similar matrices have the same trace: If $\mathsf{A}' = \mathsf{RAR}^{-1}$, then

$$\begin{aligned}
\operatorname{tr} \mathsf{A}' &= \operatorname{tr}\bigl(\mathsf{RAR}^{-1}\bigr) = \operatorname{tr}\bigl[\mathsf{R}\bigl(\mathsf{AR}^{-1}\bigr)\bigr] = \operatorname{tr}\bigl[\bigl(\mathsf{AR}^{-1}\bigr)\mathsf{R}\bigr] \\
&= \operatorname{tr}\bigl[\mathsf{A}\bigl(\mathsf{R}^{-1}\mathsf{R}\bigr)\bigr] = \operatorname{tr}(\mathsf{A1}) = \operatorname{tr} \mathsf{A}.
\end{aligned}$$

The preceding discussion is summarized in the following proposition.

Proposition 5.6.4 *To every operator* $\mathbf{A} \in \mathcal{L}(\mathcal{V})$ *are associated two intrinsic numbers*, $\det \mathbf{A}$ *and* $\operatorname{tr} \mathbf{A}$, *which are the determinant and trace of the matrix representation of the operator in any basis of* $\mathcal{V}$.

It follows from this proposition that the result of Example 5.6.3 can be written in terms of operators:

$$\det(\mathbf{1} + \epsilon \mathbf{A}) = 1 + \epsilon \operatorname{tr} \mathbf{A}. \tag{5.33}$$

A particularly useful formula that can be derived from this equation is the derivative at $t = 0$ of an operator $\mathbf{A}(t)$ depending on a single variable with the property that $\mathbf{A}(0) = \mathbf{1}$. To first order in t, we can write $\mathbf{A}(t) = \mathbf{1} + t\dot{\mathbf{A}}(0)$ where a dot represents differentiating with respect to t. Substituting this in Eq. (5.33) and differentiating with respect to t, we obtain the important result

$$\left.\frac{d}{dt} \det\bigl(\mathbf{A}(t)\bigr)\right|_{t=0} = \operatorname{tr} \dot{\mathbf{A}}(0). \tag{5.34}$$

Example 5.6.5 We have seen that the determinant of a *product* of matrices is the product of the determinants. On the other hand, the trace of a *sum* of matrices is the sum of traces. When dealing with numbers, products and

relation between determinant and trace

sums are related via the logarithm and exponential: $\alpha\beta = \exp\{\ln\alpha + \ln\beta\}$. A generalization of this relation exists for diagonalizable matrices, i.e., matrices which can be transformed into diagonal form by a suitable similarity transformation. Let A be such a matrix, i.e., let $\mathsf{D} = \mathsf{RAR}^{-1}$ for some similarity transformation R and some diagonal matrix $\mathsf{D} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. The determinant of a diagonal matrix is simply the product of its elements:

$$\det \mathsf{D} = \lambda_1\lambda_2 \ldots \lambda_n.$$

Taking the natural log of both sides and using the result of Example 5.2.6, we have

$$\ln(\det \mathsf{D}) = \ln\lambda_1 + \ln\lambda_2 + \cdots + \ln\lambda_n = \operatorname{tr}(\ln \mathsf{D}),$$

which can also be written as $\det \mathsf{D} = \exp\{\operatorname{tr}(\ln \mathsf{D})\}$.

In terms of A, this reads $\det(\mathsf{RAR}^{-1}) = \exp\{\operatorname{tr}(\ln(\mathsf{RAR}^{-1}))\}$. Now invoke the invariance of determinant and trace under similarity transformation and the result of Example 5.4.4 to obtain

$$\det \mathsf{A} = \exp\bigl\{\operatorname{tr}\bigl(\mathsf{R}(\ln \mathsf{A})\mathsf{R}^{-1}\bigr)\bigr\} = \exp\bigl\{\operatorname{tr}(\ln \mathsf{A})\bigr\}. \tag{5.35}$$

This is an important equation, which is sometimes used to define the determinant of operators in infinite-dimensional vector spaces.

Both the determinant and the trace are mappings from $\mathcal{M}^{N\times N}$ to $\mathbb{C}$. The determinant is not a linear mapping, but the trace is; and this opens up the possibility of defining an inner product in the vector space of $N \times N$ matrices in terms of the trace:

Proposition 5.6.6 *For any two matrices* $\mathsf{A}, \mathsf{B} \in \mathcal{M}^{N\times N}$, *the mapping*

$$g : \mathcal{M}^{N\times N} \times \mathcal{M}^{N\times N} \to \mathbb{C}$$

defined by $g(\mathsf{A}, \mathsf{B}) = \operatorname{tr}(\mathsf{A}^\dagger \mathsf{B})$ *is a sesquilinear inner product.*

Proof The proof follows directly from the linearity of trace and the definition of hermitian conjugate. □

Just as determinant of an operator was defined in terms of the operator itself (see Definition 2.6.10), the trace of an operator can be defined similarly as follows. Let $\mathbf{\Delta}$ be a nonzero determinant function in $\mathcal{V}$, and $\mathbf{T} \in \mathcal{L}(\mathcal{V})$. Define $\operatorname{tr}\mathbf{T}$ by

$$\sum_{i=1}^{N} \mathbf{\Delta}\bigl(|a_1\rangle, \ldots, \mathbf{T}|a_i\rangle, \ldots, |a_N\rangle\bigr) = (\operatorname{tr}\mathbf{T}) \cdot \mathbf{\Delta}\bigl(|a_1\rangle, \ldots, |a_N\rangle\bigr). \tag{5.36}$$

Then one can show that $\operatorname{tr}\mathbf{T} = \operatorname{tr}\mathsf{T}$, for any matrix T representing $\mathbf{T}$ in some basis of $\mathcal{V}$. The details are left as an exercise for the reader.

5.7 Problems

5.1 Show that if $|c\rangle = |a\rangle + |b\rangle$, then in any basis the components of $|c\rangle$ are equal to the sums of the corresponding components of $|a\rangle$ and $|b\rangle$. Also show that the elements of the matrix representing the sum of two operators are the sums of the elements of the matrices representing those two operators.

5.2 Show that the unit operator $\mathbf{1}$ is represented by the unit matrix in any basis.

5.3 The linear operator $\mathbf{A} : \mathbb{R}^3 \to \mathbb{R}^2$ is given by

$$\mathbf{A}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2x + y - 3z \\ x + y - z \end{pmatrix}.$$

Construct the matrix representing $\mathbf{A}$ in the standard bases of $\mathbb{R}^3$ and $\mathbb{R}^2$.

5.4 Find the matrix representation of the complex structure $\mathbf{J}$ on a real vector space $\mathcal{V}$ introduced in Sect. 2.4 in the basis

$$\big\{|e_1\rangle, |e_2\rangle, \ldots, |e_m\rangle, \mathbf{J}|e_1\rangle, \mathbf{J}|e_2\rangle, \ldots, \mathbf{J}|e_m\rangle\big\}.$$

5.5 The linear transformation $\mathbf{T} : \mathbb{R}^3 \to \mathbb{R}^3$ is defined as

$$\mathbf{T}(x_1, x_2, x_3) = (x_1 + x_2 - x_3, 2x_1 - x_3, x_1 + 2x_2).$$

Find the matrix representation of $\mathbf{T}$ in

(a) the standard basis of $\mathbb{R}^3$,
(b) the basis consisting of $|a_1\rangle = (1, 1, 0)$, $|a_2\rangle = (1, 0, -1)$, and $|a_3\rangle = (0, 2, 3)$.

5.6 Prove that for Eq. (5.6) to hold, we must have

$$\big(\mathbf{M}_{B_V}^{B_U}(\mathbf{B} \circ \mathbf{A})\big)_{kj} = \sum_{i=1}^{M} \big(\mathbf{M}_{B_W}^{B_U}(\mathbf{B})\big)_{ki} \big(\mathbf{M}_{B_V}^{B_W}(\mathbf{A})\big)_{ij}$$

5.7 Show that the diagonal elements of an antisymmetric matrix are all zero.

5.8 Show that the number of independent *real* parameters for an $N \times N$

(a) (real) symmetric matrix is $N(N + 1)/2$,
(b) (real) antisymmetric matrix is $N(N - 1)/2$,
(c) (real) orthogonal matrix is $N(N - 1)/2$,
(d) (complex) unitary matrix is N^2,
(e) (complex) hermitian matrix is N^2.

5.9 Show that an arbitrary orthogonal 2×2 matrix can be written in one of the following two forms:

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}.$$

The first is a pure rotation (its determinant is $+1$), and the second has determinant -1. The form of the choices is dictated by the assumption that the first entry of the matrix reduces to 1 when $\theta = 0$.

5.10 Derive the formulas

$$\cos(\theta_1 + \theta_2) = \cos\theta_1 \cos\theta_2 - \sin\theta_1 \sin\theta_2,$$
$$\sin(\theta_1 + \theta_2) = \sin\theta_1 \cos\theta_2 + \cos\theta_1 \sin\theta_2$$

by noting that the rotation of the angle $\theta_1 + \theta_2$ in the xy-plane is the product of two rotations. (See Problem 5.9.)

5.11 Prove that if a matrix M satisfies $\mathsf{M}\mathsf{M}^\dagger = 0$, then $\mathsf{M} = \mathsf{0}$. Note that in general, $\mathsf{M}^2 = \mathsf{0}$ does not imply that M is zero. Find a nonzero 2×2 matrix whose square is zero.

5.12 Construct the matrix representations of

$$\mathbf{D} : \mathcal{P}_4^c[t] \to \mathcal{P}_4^c[t] \quad \text{and} \quad \mathbf{T} : \mathcal{P}_3^c[t] \to \mathcal{P}_4^c[t],$$

the derivative and multiplication-by-t operators. Choose $\{1, t, t^2, t^3\}$ as your basis of $\mathcal{P}_3^c[t]$ and $\{1, t, t^2, t^3, t^4\}$ as your basis of $\mathcal{P}_4^c[t]$. Use the matrix of $\mathbf{D}$ so obtained to find the first, second, third, fourth, and fifth derivatives of a general polynomial of degree 4.

5.13 Find the transformation matrix R that relates the (orthonormal) standard basis of $\mathbb{C}^3$ to the orthonormal basis obtained from the following vectors via the Gram-Schmidt process:

$$|a_1\rangle = \begin{pmatrix} 1 \\ i \\ 0 \end{pmatrix}, \qquad |a_2\rangle = \begin{pmatrix} 0 \\ 1 \\ -i \end{pmatrix}, \qquad |a_3\rangle = \begin{pmatrix} i \\ 0 \\ -1 \end{pmatrix}.$$

Verify that R is unitary, as expected from Theorem 5.4.2.

5.14 If the matrix representation of an endomorphism $\mathbf{T}$ of $\mathbb{C}^2$ with respect to the standard basis is $\left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)$, what is its matrix representation with respect to the basis $\left\{\left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right)\right\}$?

5.15 If the matrix representation of an endomorphism $\mathbf{T}$ of $\mathbb{C}^3$ with respect to the standard basis is

$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & -1 \\ -1 & -1 & 0 \end{pmatrix}$$

what is the representation of $\mathbf{T}$ with respect to the basis

$$\left\{\begin{pmatrix}0\\1\\-1\end{pmatrix}, \begin{pmatrix}1\\-1\\1\end{pmatrix}, \begin{pmatrix}-1\\1\\0\end{pmatrix}\right\}?$$

5.16 Using Theorem 5.5.1, calculate the determinant of a general 3×3 matrix and obtain the familiar expansion of such a determinant in terms of the first row of the matrix.

5.17 Using Theorem 5.5.1, show that if two rows (two columns) of a matrix are equal, then its determinant is zero.

5.18 Show that $\det(\alpha \mathsf{A}) = \alpha^N \det \mathsf{A}$ for an $N \times N$ matrix A and a complex number α.

5.19 Show that $\det \mathsf{1} = 1$ for any unit matrix.

5.20 Find a specific pair of matrices A and B such that $\det(\mathsf{A}+\mathsf{B}) \neq \det \mathsf{A} + \det \mathsf{B}$. Therefore, the determinant is *not* a linear mapping. Hint: *Any* pair of matrices will most likely work. In fact, the challenge is to find a pair such that $\det(\mathsf{A}+\mathsf{B}) = \det \mathsf{A} + \det \mathsf{B}$.

5.21 Let A be any $N \times N$ matrix. Replace its ith row (column) with any one of its other rows (columns), leaving the latter unchanged. Now expand the determinant of the new matrix by its ith row (column) to show that

$$\sum_{j=1}^{N} (\mathsf{A})_{ji} (\operatorname{cof} \mathsf{A})_{jk} = \sum_{j=1}^{N} (\mathsf{A})_{ij} (\operatorname{cof} \mathsf{A})_{kj} = 0, \quad k \neq i.$$

5.22 Demonstrate the result of Problem 5.21 using an arbitrary 4×4 matrix and evaluating the sum explicitly.

5.23 Suppose that $\mathbf{A}$ is represented by A in one basis and by A' in another, related to the first by a similarity transformation R. Show directly that $\det \mathsf{A}' = \det \mathsf{A}$.

5.24 Show explicitly that $\det(\mathsf{AB}) = \det \mathsf{A} \det \mathsf{B}$ for 2×2 matrices.

5.25 Given three $N \times N$ matrices A, B, and C such that $\mathsf{AB} = \mathsf{C}$ with C invertible, show that both A and B must be invertible. Thus, any two *operators* $\mathbf{A}$ and $\mathbf{B}$ on a *finite*-dimensional vector space satisfying $\mathbf{AB} = \mathbf{1}$ are invertible and each is the inverse of the other. Note: This is not true for infinite-dimensional vector spaces.

5.26 Show directly that the similarity transformation induced by R does not change the determinant or the trace of A where

$$\mathsf{R} = \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & -2 \\ 2 & 1 & -1 \end{pmatrix} \quad \text{and} \quad \mathsf{A} = \begin{pmatrix} 3 & -1 & 2 \\ 0 & 1 & -2 \\ 1 & -3 & -1 \end{pmatrix}.$$

5.27 Find the matrix that transforms the standard basis of $\mathbb{C}^3$ to the vectors

$$|a_1\rangle = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} \\ \frac{1+i}{\sqrt{6}} \end{pmatrix}, \qquad |a_2\rangle = \begin{pmatrix} \frac{-i}{\sqrt{2}} \\ \frac{i}{\sqrt{6}} \\ \frac{-1+i}{\sqrt{6}} \end{pmatrix}, \qquad |a_3\rangle = \begin{pmatrix} 0 \\ \frac{-2}{\sqrt{6}} \\ \frac{1+i}{\sqrt{6}} \end{pmatrix}.$$

Show that this matrix is unitary.

5.28 Consider the three operators $\mathbf{L}_1, \mathbf{L}_2$, and $\mathbf{L}_3$ satisfying

$$[\mathbf{L}_1, \mathbf{L}_2] = i\mathbf{L}_3, \qquad [\mathbf{L}_3, \mathbf{L}_1] = i\mathbf{L}_2, \qquad [\mathbf{L}_2, \mathbf{L}_3] = i\mathbf{L}_1.$$

Show that the trace of each of these operators is necessarily zero.

5.29 Show that in the expansion of the determinant given in Theorem 5.5.1, no two elements of the same row or the same column can appear in each term of the sum.

5.30 Find the inverse of the following matrices if they exist:

$$\mathsf{A} = \begin{pmatrix} 3 & -1 & 2 \\ 1 & 0 & -3 \\ -2 & 1 & -1 \end{pmatrix}, \qquad \mathsf{B} = \begin{pmatrix} 0 & 1 & -1 \\ 1 & 2 & 0 \\ -1 & -2 & 1 \end{pmatrix},$$

$$\mathsf{C} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & -1 \end{pmatrix}.$$

5.31 Find inverses for the following matrices using both methods discussed in this chapter.

$$\mathsf{A} = \begin{pmatrix} 2 & 1 & -1 \\ 2 & 1 & 2 \\ -1 & -2 & -2 \end{pmatrix}, \qquad \mathsf{B} = \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & -2 \\ 2 & 1 & -1 \end{pmatrix},$$

$$\mathsf{C} = \begin{pmatrix} 1 & -1 & 1 \\ -1 & 1 & 1 \\ 1 & -1 & -2 \end{pmatrix},$$

$$\mathsf{D} = \begin{pmatrix} 1/\sqrt{2} & 0 & (1-i)/(2\sqrt{2}) & (1+i)/(2\sqrt{2}) \\ 0 & 1/\sqrt{2} & (1-i)/(2\sqrt{2}) & -(1+i)/(2\sqrt{2}) \\ 1/\sqrt{2} & 0 & -(1-i)/(2\sqrt{2}) & -(1+i)/(2\sqrt{2}) \\ 0 & 1/\sqrt{2} & -(1-i)/(2\sqrt{2}) & (1+i)/(2\sqrt{2}) \end{pmatrix}.$$

5.32 Let $\mathbf{A}$ be an operator on $\mathcal{V}$. Show that if $\det \mathbf{A} = 0$, then there exists a nonzero vector $|x\rangle \in \mathcal{V}$ such that $\mathbf{A}|x\rangle = 0$.

5.33 For which values of α are the following matrices invertible? Find the inverses whenever they exist.

$$\mathsf{A} = \begin{pmatrix} 1 & \alpha & 0 \\ \alpha & 1 & \alpha \\ 0 & \alpha & 1 \end{pmatrix}, \qquad \mathsf{B} = \begin{pmatrix} \alpha & 1 & 0 \\ 1 & \alpha & 1 \\ 0 & 1 & \alpha \end{pmatrix},$$

$$\mathsf{C} = \begin{pmatrix} 0 & 1 & \alpha \\ 1 & \alpha & 0 \\ \alpha & 0 & 1 \end{pmatrix}, \qquad \mathsf{D} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & \alpha \\ 1 & \alpha & 1 \end{pmatrix}.$$

5.34 Let $\{\mathbf{a}_i\}_{i=1}^N$, be the set consisting of the N rows of an $N \times N$ matrix A and assume that the $\mathbf{a}_i$ are orthogonal to each other. Show that

$$|\det \mathsf{A}| = \|\mathbf{a}_1\| \, \|\mathbf{a}_2\| \cdots \|\mathbf{a}_N\|.$$

Hint: Consider $\mathsf{A}\mathsf{A}^\dagger$. What would the result be if A were a unitary matrix?

5.35 Prove that a set of n homogeneous linear equations in n unknowns has a nontrivial solution if and only if the determinant of the matrix of coefficients is zero.

5.36 Use determinants to show that an antisymmetric matrix whose dimension is odd cannot have an inverse.

5.37 Let $\mathcal{V}$ be a real inner product space. Let $\mathbf{\Theta} : \mathcal{V}^N \times \mathcal{V}^N \to \mathbb{R}$ be a function defined by

$$\mathbf{\Theta}\big(|v_1\rangle, \dots, |v_N\rangle, |u_1\rangle, \dots, |u_N\rangle\big) = \det\big(\langle u_i|v_j\rangle\big).$$

Follow the same procedure as in Sect. 5.5.3 to show that for any determinant function $\mathbf{\Delta}$ in $\mathcal{V}$ there is a nonzero constant $\alpha \in \mathbb{R}$ such that

$$\mathbf{\Delta}\big(|v_1\rangle, \dots, |v_N\rangle\big)\mathbf{\Delta}\big(|u_1\rangle, \dots, |u_N\rangle\big) = \alpha \det\big(\langle u_i|v_j\rangle\big)$$

for $|u_i\rangle, |v_j\rangle \in \mathcal{V}$.

5.38 Show that $\operatorname{tr}(|a\rangle\langle b|) = \langle b|a\rangle$. Hint: Evaluate the trace in an orthonormal basis.

5.39 Show that if two invertible $N \times N$ matrices A and B anticommute (that is, $\mathsf{AB} + \mathsf{BA} = 0$), then (a) N must be even, and (b) $\operatorname{tr}\mathsf{A} = \operatorname{tr}\mathsf{B} = 0$.

5.40 Show that for a spatial rotation $R_{\hat{n}}(\theta)$ of an angle θ about an *arbitrary* axis $\hat{n}$, $\operatorname{tr} R_{\hat{n}}(\theta) = 1 + 2\cos\theta$.

5.41 Express the sum of the squares of elements of a matrix as a trace. Show that this sum is invariant under an orthogonal transformation of the matrix.

5.42 Let S and A be a symmetric and an antisymmetric matrix, respectively, and let M be a general matrix. Show that

(a) $\operatorname{tr}\mathsf{M} = \operatorname{tr}\mathsf{M}^t$,
(b) $\operatorname{tr}(\mathsf{SA}) = 0$; in particular, $\operatorname{tr}\mathsf{A} = 0$,
(c) SA is antisymmetric if and only if $[\mathsf{S}, \mathsf{A}] = 0$,
(d) MSM^t is symmetric and MAM^t is antisymmetric,
(e) $\mathsf{MHM}^\dagger$ is hermitian if H is.

5.43 Find the trace of each of the following linear operators:

(a) $\mathbf{T} : \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$\mathbf{T}(x, y, z) = (x + y - z, 2x + 3y - 2z, x - y).$$

(b) $\mathbf{T} : \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$\mathbf{T}(x, y, z) = (y - z, x + 2y + z, z - y).$$

(c) $\mathbf{T} : \mathbb{C}^4 \to \mathbb{C}^4$ given by

$$\mathbf{T}(x, y, z, w) = (x + iy - z + iw, 2ix + 3y - 2iz - w, x - iy, z + iw).$$

5.44 Use Eq. (5.35) to derive Eq. (5.33).

5.45 Suppose that there are two operators $\mathbf{A}$ and $\mathbf{B}$ such that $[\mathbf{A}, \mathbf{B}] = c\mathbf{1}$, where c is a constant. Show that the vector space in which such operators are defined cannot be finite-dimensional. Conclude that the position and momentum operators of quantum mechanics can be defined only in infinite dimensions.

5.46 Use Eq. (5.36) to show that $\operatorname{tr}\mathbf{T} = \operatorname{tr}\mathsf{T}$, for any matrix T representing $\mathbf{T}$ in some basis of $\mathcal{V}$.