Vector coordinates, matrix elements, changes of basis, and matrix diagonalization

1. Coordinates of vectors and matrix elements of linear operators

Let V be an n-dimensional real (or complex) vector space. Vectors that live in V are usually represented by a single column of n real (or complex) numbers. A linear transformation (also called a linear operator) acting on V is a "machine" that acts on a vector and and produces another vector. Linear operators are represented by square $n \times n$ real (or complex) matrices.[1]

If it is not specified, the representations of vectors and matrices described above implicitly assume that the *standard basis* has been chosen. That is, all vectors in V can be expressed as linear combinations of basis vectors:[2]

$$\begin{aligned}\mathcal{B}_s &= \big\{\widehat{\boldsymbol{e}}_1\,,\,\widehat{\boldsymbol{e}}_2\,,\,\widehat{\boldsymbol{e}}_3\,,\,\ldots\,,\,\widehat{\boldsymbol{e}}_n\big\} \\ &= \big\{(1,0,0,\ldots,0)^{\mathsf{T}}\,,\,(0,1,0,\ldots,0)^{\mathsf{T}}\,,\,(0,0,1,\ldots,0)^{\mathsf{T}}\,,\,\ldots\,,\,(0,0,0,\ldots,1)^{\mathsf{T}}\big\}\,.\end{aligned}$$

The subscript s indicates that this is the standard basis. The superscript T (which stands for *transpose*) turns the row vectors into column vectors. Thus,

$$\vec{\boldsymbol{v}} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_n \end{pmatrix} = v_1 \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + v_2 \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ v_n \end{pmatrix} + v_3 \begin{pmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + v_n \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}\,. \tag{1}$$

The v_i are the components of the vector $\vec{\boldsymbol{v}}$. However, it is more precise to say that the v_i are the *coordinates* of the abstract vector $\vec{\boldsymbol{v}}$ with respect to the standard basis.

Consider a linear operator A. The corresponding matrix representation is given by $A = [a_{ij}]$. For example, if $\vec{\boldsymbol{w}} = A\vec{\boldsymbol{v}}$, then

$$w_i = \sum_{j=1}^{n} a_{ij} v_j\,, \tag{2}$$

[1] We can generalize this slightly by viewing a linear operator as a function whose input is taken from vectors that live in V and whose output is a vector that lives in another vector space W. If V is n-dimensional and W is m-dimensional, then a linear operator is represented by an $m \times n$ real (or complex) matrix. In these notes, we will simplify the discussion by always taking $W = V$.

[2] If $V = \mathbb{R}^3$ (*i.e.*, three-dimensional Euclidean space), then it is traditional to designate $\widehat{\boldsymbol{e}}_1 = \widehat{\boldsymbol{i}}$, $\widehat{\boldsymbol{e}}_2 = \widehat{\boldsymbol{j}}$ and $\widehat{\boldsymbol{e}}_3 = \widehat{\boldsymbol{k}}$.

where v_i and w_i are the coordinates of $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ with respect to the standard basis and a_{ij} are the matrix elements of A with respect to the standard basis. If we express $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ as linear combinations of basis vectors, then

$$\vec{\boldsymbol{v}} = \sum_{j=1}^{n} v_j \widehat{\boldsymbol{e}}_j \,, \qquad\qquad \vec{\boldsymbol{w}} = \sum_{i=1}^{n} w_i \widehat{\boldsymbol{e}}_i \,,$$

then $\vec{\boldsymbol{w}} = A\vec{\boldsymbol{v}}$ implies that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} v_j \widehat{\boldsymbol{e}}_i = A \sum_{j=1}^{n} v_j \widehat{\boldsymbol{e}}_j \,,$$

where we have used eq. (2) to substitute for w_i. It follows that:

$$\sum_{j=1}^{n} \left(A\widehat{\boldsymbol{e}}_j - \sum_{i=1}^{n} a_{ij} \widehat{\boldsymbol{e}}_i \right) v_j = 0 \,. \tag{3}$$

Eq. (3) must be true for *any* vector $\vec{\boldsymbol{v}} \in V$; that is, for any choice of coordinates v_j. Thus, the coefficient inside the parentheses in eq. (3) must vanish. We conclude that:

$$\boxed{A\widehat{\boldsymbol{e}}_j = \sum_{i=1}^{n} a_{ij} \widehat{\boldsymbol{e}}_i \,.} \tag{4}$$

Eq. (4) can be used as the definition of the matrix elements a_{ij} with respect to the standard basis of a linear operator A.

To appreciate the meaning of eq. (4), consider the following example. If we choose $j = 1$ then eq. (4) is equivalent to

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix}$$

$$= \sum_{i=1}^{n} \left\{ a_{11} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + a_{21} \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \ldots + a_{n1} \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \right\} . \tag{5}$$

One can write similar results for the other possible choices of $j = 2, 3, \ldots, n$ in eq. (4). Note that by writing $A = [a_{ij}]$, one has implicitly made a choice of the standard basis.

However, there is nothing sacrosanct about the choice of the standard basis. One can expand a vector as a linear combination of any set of n linearly independent vectors. Thus, we generalize the above discussion by introducing a basis

$$\mathcal{B} = \left\{ \vec{\boldsymbol{b}}_1 \,,\, \vec{\boldsymbol{b}}_2 \,,\, \vec{\boldsymbol{b}}_3 \,,\, \ldots \,,\, \vec{\boldsymbol{b}}_n \right\} .$$

Because the basis $\mathcal{B}$ consists of n linearly independent vectors, one can find a unique set of coefficients v'_i for any vector $\vec{\boldsymbol{v}} \in V$ such that

$$\vec{\boldsymbol{v}} = \sum_{j=1}^{n} v'_j \, \vec{\boldsymbol{b}}_j \, . \tag{6}$$

The v'_i are the *coordinates* of $\vec{\boldsymbol{v}}$ with respect to the basis $\mathcal{B}$.[3] Likewise, in analogy with eq. (4), given a linear operator A one can write

$$A\vec{\boldsymbol{b}}_j = \sum_{i=1}^{n} a'_{ij} \vec{\boldsymbol{b}}_i \tag{7}$$

where the a'_{ij} are the *matrix elements* of the linear operator A with respect to the basis $\mathcal{B}$.[4]

Clearly, these more general definitions reduce to the previous ones given in the case of the standard basis. Moreover, one can easily compute $A\vec{\boldsymbol{v}} \equiv \vec{\boldsymbol{w}}$ using the results of eqs. (6) and (7):

$$A\vec{\boldsymbol{v}} = \sum_{i=1}^{n} \sum_{j=1}^{n} a'_{ij} v'_j \vec{\boldsymbol{b}}_i = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} a'_{ij} v'_j \right) \vec{\boldsymbol{b}}_i = \sum_{i=1}^{n} w'_i \vec{\boldsymbol{b}}_i = \vec{\boldsymbol{w}} \, ,$$

which implies that the coordinates of the vector $\vec{\boldsymbol{w}} = A\vec{\boldsymbol{v}}$ with respect to the basis $\mathcal{B}$ are given by:

$$w'_i = \sum_{j=1}^{n} a'_{ij} v'_j \, .$$

Thus, the relation between the coordinates of $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ with respect to the basis $\mathcal{B}$ is the same as the relation obtained with respect to the standard basis [see eq. (2)]. One must simply be consistent and always employ the *same* basis for defining the vector coordinates and the matrix elements of a linear operator in a practical calculation.

2. Change of basis and its effects on coordinates and matrix elements

The choice of basis is arbitrary. The existence of vectors and linear operators does not depend on the choice of basis. However, a choice of basis is very convenient since it permits explicit calculations involving vectors and matrices. Suppose we start with some basis choice $\mathcal{B}$ and then later we decide to employ a different basis choice $\mathcal{C}$,

$$\mathcal{C} = \left\{ \vec{\boldsymbol{c}}_1 \, , \, \vec{\boldsymbol{c}}_2 \, , \, \vec{\boldsymbol{c}}_3 \, , \, \ldots \, , \, \vec{\boldsymbol{c}}_n \right\} .$$

[3]We write v'_i to distinguish these from the coordinates with respect to the Standard basis, which were denoted by v_i in eq. (1)

[4]We write a'_{ij} to distinguish these from the matrix elements with respect to the Standard basis, which were denoted by a_{ij} in eq. (4)

In particular, suppose $\mathcal{B} = \mathcal{B}_s$ is the standard basis. Then to change from $\mathcal{B}_s$ to $\mathcal{C}$ is geometrically equivalent to starting with a definition of the x, y and z axis, and then defining a new set of axes. Note that we have not yet introduced the concept of an inner product or norm, so there is no concept of orthogonality or unit vectors. The new set of axes may be quite skewed (although such a concept also requires an inner product).

Thus, we pose the following question. If the coordinates of a vector $\vec{\boldsymbol{v}}$ and the matrix elements of a linear operator A are known with respect to a basis $\mathcal{B}$ (which need not be the standard basis), what are the coordinates of the vector $\vec{\boldsymbol{v}}$ and the matrix elements of a linear operator A with respect to a basis $\mathcal{C}$? To answer this question, we must describe the relation between $\mathcal{B}$ and $\mathcal{C}$. We do this as follows. The basis vectors of $\mathcal{C}$ can be expressed as linear combinations of the basis vectors $\vec{\boldsymbol{b}}_i$, since the latter span the vector space V. We shall denote these coefficients as follows:

$$\vec{\boldsymbol{c}}_j = \sum_{i=1}^{n} P_{ij}\vec{\boldsymbol{b}}_i\,, \qquad j = 1, 2, 3, \ldots, n\,. \tag{8}$$

Note that eq. (8) is a shorthand for n separate equations, and provides the coefficients P_{i1}, $P_{i2}, \ldots$, P_{in} needed to expand $\vec{\boldsymbol{c}}_1$, $\vec{\boldsymbol{c}}_2, \ldots$, $\vec{\boldsymbol{c}}_n$, respectively, as linear combinations of the $\vec{\boldsymbol{b}}_i$. We can assemble the P_{ij} into a matrix. A crucial observation is that this matrix P is invertible. This must be true, since one can reverse the process and express the basis vectors of $\mathcal{B}$ as linear combinations of the basis vectors $\vec{\boldsymbol{c}}_i$ (which again follows from the fact that the latter span the vector space V). Explicitly,

$$\vec{\boldsymbol{b}}_k = \sum_{j=1}^{n} (P^{-1})_{jk}\vec{\boldsymbol{c}}_j\,, \qquad k = 1, 2, 3, \ldots, n\,. \tag{9}$$

We are now in the position to determine the coordinates of a vector $\vec{\boldsymbol{v}}$ and the matrix elements of a linear operator A with respect to a basis $\mathcal{C}$. Assume that the coordinates of $\vec{\boldsymbol{v}}$ with respect to $\mathcal{B}$ are given by v_i and the matrix elements of A with respect to $\mathcal{B}$ are given by a_{ij}. With respect to $\mathcal{C}$, we shall denote the vector coordinates by v'_i and the matrix elements by a'_{ij}. Then, using the definition of vector coordinates [eq. (6)] and matrix elements [eq. (7)],

$$\vec{\boldsymbol{v}} = \sum_{j=1}^{n} v'_j \vec{\boldsymbol{c}}_j = \sum_{j=1}^{n} v'_j \sum_{i=1}^{n} P_{ij}\vec{\boldsymbol{b}}_i = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} P_{ij}v'_j\right)\vec{\boldsymbol{b}}_i = \sum_{i=1}^{n} v_i\vec{\boldsymbol{b}}_i\,, \tag{10}$$

where we have used eq. (8) to express the $\vec{\boldsymbol{c}}_j$ in terms of the $\vec{\boldsymbol{b}}_i$. Note that the last step in eq. (10) can be rewritten as:

$$\sum_{i=1}^{n}\left(v_i - \sum_{j=1}^{n} P_{ij}v'_j\right)\vec{\boldsymbol{b}}_i = 0\,. \tag{11}$$

Since the $\vec{\boldsymbol{b}}_i$ are linearly independent, the coefficient inside the parentheses in eq. (11) must vanish. Hence,

$$v_i = \sum_{j=1}^{n} P_{ij}v'_j\,, \qquad \text{or equivalently} \qquad \boxed{[\vec{\boldsymbol{v}}]_{\mathcal{B}} = P[\vec{\boldsymbol{v}}]_{\mathcal{C}}}\,. \tag{12}$$

Here we have introduced the notation $[\vec{\boldsymbol{v}}]_{\mathcal{B}}$ to indicate the vector $\vec{\boldsymbol{v}}$ represented in terms of its coordinates with respect to the basis $\mathcal{B}$. Inverting this result yields:

$$v'_j = \sum_{k=1}^{n} (P^{-1})_{jk} v_k \,, \qquad \text{or equivalently} \qquad \boxed{[\vec{\boldsymbol{v}}]_{\mathcal{C}} = P^{-1}[\vec{\boldsymbol{v}}]_{\mathcal{B}}} \,. \tag{13}$$

Thus, we have determined the relation between the coordinates of $\vec{\boldsymbol{v}}$ with respect to the bases $\mathcal{B}$ and $\mathcal{C}$.

A similar computation can determine the relation between the matrix elements of A with respect to the basis $\mathcal{B}$, which we denote by a_{ij} [see eq. (7)], and the matrix elements of A with respect to the basis $\mathcal{C}$, which we denote by a'_{ij}:

$$A\vec{\boldsymbol{c}}_j = \sum_{i=1}^{n} a'_{ij} \vec{\boldsymbol{c}}_i \,. \tag{14}$$

The desired relation can be obtained by evaluating $A\vec{\boldsymbol{b}}_\ell$:

$$\begin{aligned} A\vec{\boldsymbol{b}}_\ell &= A \sum_{j=1}^{n} (P^{-1})_{j\ell} \vec{\boldsymbol{c}}_j = \sum_{j=1}^{n} (P^{-1})_{j\ell}\, A\vec{\boldsymbol{c}}_j = \sum_{j=1}^{n} (P^{-1})_{j\ell} \sum_{i=1}^{n} a'_{ij} \vec{\boldsymbol{c}}_i \\ &= \sum_{j=1}^{n} (P^{-1})_{j\ell} \sum_{i=1}^{n} a'_{ij} \sum_{k=1}^{n} P_{ki} \vec{\boldsymbol{b}}_k = \sum_{k=1}^{n} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} P_{ki} a'_{ij} (P^{-1})_{j\ell} \right) \vec{\boldsymbol{b}}_k \,, \end{aligned}$$

where we have used eqs. (8) and (9) and the definition of the matrix elements of A with respect to the basis $\mathcal{C}$ [eq. (14)]. Comparing this result with eq. (7), it follows that

$$\sum_{k=1}^{n} \left(a_{k\ell} - \sum_{i=1}^{n} \sum_{j=1}^{n} P_{ki} a'_{ij} (P^{-1})_{j\ell} \right) \vec{\boldsymbol{b}}_k = 0 \,.$$

Since the $\vec{\boldsymbol{b}}_k$ are linearly independent, we conclude that

$$a_{k\ell} = \sum_{i=1}^{n} \sum_{j=1}^{n} P_{ki} a'_{ij} (P^{-1})_{j\ell} \,.$$

The double sum above corresponds to the matrix multiplication of three matrices, so it is convenient to write this result symbolically as:

$$\boxed{[A]_{\mathcal{B}} = P[A]_{\mathcal{C}} P^{-1} \,.} \tag{15}$$

In particular, the matrix formed by the matrix elements of A with respect to the basis $\mathcal{B}$ is related to the matrix formed by the matrix elements of A with respect to the basis $\mathcal{C}$ by the *similarity transformation* given by eq. (15). We can invert eq. (15) to obtain:

$$\boxed{[A]_{\mathcal{C}} = P^{-1}[A]_{\mathcal{B}} P \,.} \tag{16}$$

In fact, there is a much faster method to derive eqs. (15) and (16). Consider the equation $\vec{\boldsymbol{w}} = A\vec{\boldsymbol{v}}$ evaluated with respect to bases $\mathcal{B}$ and $\mathcal{C}$, respectively:

$$[\vec{\boldsymbol{w}}]_{\mathcal{B}} = [A]_{\mathcal{B}}[\vec{\boldsymbol{v}}]_{\mathcal{B}}\,, \qquad\qquad [\vec{\boldsymbol{w}}]_{\mathcal{C}} = [A]_{\mathcal{C}}[\vec{\boldsymbol{v}}]_{\mathcal{C}}\,.$$

Using eq. (12), $[\vec{\boldsymbol{w}}]_{\mathcal{B}} = [A]_{\mathcal{B}}[\vec{\boldsymbol{v}}]_{\mathcal{B}}$ can be rewritten as:

$$P[\vec{\boldsymbol{w}}]_{\mathcal{C}} = [A]_{\mathcal{B}}P[\vec{\boldsymbol{v}}]_{\mathcal{C}}\,.$$

Hence,

$$[\vec{\boldsymbol{w}}]_{\mathcal{C}} = [A]_{\mathcal{C}}[\vec{\boldsymbol{v}}]_{\mathcal{C}} = P^{-1}[A]_{\mathcal{B}}P[\vec{\boldsymbol{v}}]_{\mathcal{C}}\,.$$

It then follows that

$$\left\{[A]_{\mathcal{C}} - P^{-1}[A]_{\mathcal{B}}P\right\}[\vec{\boldsymbol{v}}]_{\mathcal{C}} = 0\,. \tag{17}$$

Since this equation must be true for all $\vec{\boldsymbol{v}} \in V$ (and thus for any choice of $[\vec{\boldsymbol{v}}]_{\mathcal{C}}$), it follows that the quantity inside the parentheses in eq. (17) must vanish. This yields eq. (16).

The significance of eq. (16) is as follows. If two matrices are related by a similarity transformation, then these matrices may represent the *same* linear operator with respect to two different choices of basis. These two choices are related by eq. (8).[5]

Example: Let $\mathcal{B} = \mathcal{B}_s$ be the standard basis and let $\mathcal{C} = \left\{(1,0,0)\,,\,(1,1,0)\,,\,(1,1,1)\right\}$. Suppose that we are given a linear operator A whose matrix elements with respect to the standard basis $\mathcal{B}_s$ are:

$$[A]_{\mathcal{B}} = \begin{pmatrix} 1 & 2 & -1 \\ 0 & -1 & 0 \\ 1 & 0 & 7 \end{pmatrix}\,.$$

In order to determine $[A]_{\mathcal{C}}$, one must first identify the matrix P. Noting that:

$$\vec{\boldsymbol{c}}_1 = \vec{\boldsymbol{b}}_1\,, \qquad \vec{\boldsymbol{c}}_2 = \vec{\boldsymbol{b}}_1 + \vec{\boldsymbol{b}}_2\,, \qquad \vec{\boldsymbol{c}}_3 = \vec{\boldsymbol{b}}_1 + \vec{\boldsymbol{b}}_2 + \vec{\boldsymbol{b}}_3\,, \tag{18}$$

it follows from eq. (8) that,

$$P = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}\,.$$

One quick way to compute P^{-1} is to invert the relations given in eq. (18) to obtain,

$$\vec{\boldsymbol{b}}_1 = \vec{\boldsymbol{c}}_1\,, \qquad \vec{\boldsymbol{b}}_2 = \vec{\boldsymbol{c}}_2 - \vec{\boldsymbol{c}}_1\,, \qquad \vec{\boldsymbol{b}}_3 = \vec{\boldsymbol{c}}_3 - \vec{\boldsymbol{c}}_2\,. \tag{19}$$

[5]However, it would *not* be correct to conclude that two matrices that are related by a similarity transformation cannot represent different linear operators. In fact, one could also interpret these two matrices as representing (with respect to the *same* basis) two different linear operators that are related by a similarity transformation. That is, given two linear operators A and B and an invertible linear operator P, it is clear that if $B = P^{-1}AP$ then the matrix elements of A and B with respect to a fixed basis are related by the same similarity transformation.

Hence, eq. (9) yields,

$$P^{-1} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} .$$

Finally, we make use of eq. (16) to obtain.

$$[A]_{\mathcal{C}} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & -1 \\ 0 & -1 & 0 \\ 1 & 0 & 7 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 4 & 3 \\ -1 & -2 & -9 \\ 1 & 1 & 8 \end{pmatrix} .$$

One last observation is noteworthy. Although the matrix elements of a matrix A depend on the choice of basis, the eigenvalues of A are independent of the basis choice. To prove this assertion, note that the eigenvalues of $[A]_{\mathcal{B}}$ are the roots of the characteristic equation, $\det\big\{[A]_{\mathcal{B}} - \lambda\mathbf{I}\big\} = 0$. However, in light of eq. (16), the characteristic equations of $[A]_{\mathcal{B}}$ and $[A]_{\mathcal{C}}$ are the same, since

$$\det\big\{[A]_{\mathcal{C}} - \lambda\mathbf{I}\big\} = \det\big\{P^{-1}[A]_{\mathcal{B}}P - \lambda\mathbf{I}\big\} = \det\big\{P^{-1}\big([A]_{\mathcal{B}} - \lambda\mathbf{I}\big)P\big\} = \det\big\{[A]_{\mathcal{B}} - \lambda\mathbf{I}\big\} ,$$

after using $PP^{-1} = \mathbf{I}$ and the result of problem 3.11–9 on p. 159 of Boas.

3. Applications to matrix diagonalization

Consider a matrix $A \equiv [A]_{\mathcal{B}_s}$, whose matrix elements are defined with respect to the standard basis, $\mathcal{B}_s = \big\{\widehat{\boldsymbol{e}}_1\,,\,\widehat{\boldsymbol{e}}_2\,,\,\widehat{\boldsymbol{e}}_3\,,\,\ldots\,,\,\widehat{\boldsymbol{e}}_n\big\}$. The eigenvalue problem for the matrix A consists of finding all complex λ_i such that

$$A\vec{\boldsymbol{v}}_i = \lambda_i\vec{\boldsymbol{v}}_i\,, \qquad \vec{\boldsymbol{v}}_i \neq 0 \quad \text{for} \quad i = 1, 2, \ldots, n\,. \tag{20}$$

The λ_i are the roots of the characteristic equation $\det(A - \lambda\mathbf{I}) = 0$. This is an nth order polynomial equation which has n (possibly complex) roots, although some of the roots could be degenerate. If the roots are nondegenerate, then A is called *simple*. In this case, the n eigenvectors are linearly independent and span the vector space V.[6] If some of the roots are degenerate, then the corresponding n eigenvectors may or may not be linearly independent. In general, if A possesses n linearly independent eigenvectors, then A is called *semisimple*.[7] If some of the eigenvalues of A are degenerate and its eigenvectors do *not* span the vector space V, then we say that A is *defective*. A is diagonalizable if and only if it is semisimple.

In cases where the eigenvectors of a matrix A span the vector space V (i.e., A is semisimple), we may define a new basis consisting of the eigenvectors of A, which we shall denote by $\mathcal{C} = \big\{\vec{\boldsymbol{v}}_1\,,\,\vec{\boldsymbol{v}}_2\,,\,\vec{\boldsymbol{v}}_3\,,\,\ldots\,,\,\vec{\boldsymbol{v}}_n\big\}$. The matrix elements of A with respect to the basis $\mathcal{C}$, denoted by $[A]_{\mathcal{C}}$, is obtained by employing eq. (14):

$$A\vec{\boldsymbol{v}}_j = \sum_{i=1}^{n} a'_{ij}\vec{\boldsymbol{v}}_i\,.$$

[6] This result is proved in Appendix A.

[7] Note that if A is semisimple, then A is also simple only if all the eigenvalues of A are nondegenerate.

But, eq. (20) implies that $a'_{ij} = \lambda_j \delta_{ij}$. That is,

$$[A]_{\mathcal{C}} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} .$$

The relation between A and $[A]_{\mathcal{C}}$ can be obtained from eq. (16). Thus, we must determine the matrix P that governs the relation between $\mathcal{B}_s$ and $\mathcal{C}$ [cf. eq. (8)]. Consider the coordinates of $\vec{\boldsymbol{v}}_j$ with respect to the standard basis $\mathcal{B}_s$:

$$\vec{\boldsymbol{v}}_j = \sum_{i=1}^{n} (\vec{\boldsymbol{v}}_j)_i \, \widehat{\boldsymbol{e}}_i = \sum_{i=1}^{n} P_{ij} \widehat{\boldsymbol{e}}_i \,, \tag{21}$$

where $(\vec{\boldsymbol{v}}_j)_i$ is the ith coordinate of the jth eigenvector. Using eq. (21), we identify $P_{ij} = (\vec{\boldsymbol{v}}_j)_i$. In matrix form,

$$P = \begin{pmatrix} (v_1)_1 & (v_2)_1 & \cdots & (v_n)_1 \\ (v_1)_2 & (v_2)_2 & \cdots & (v_n)_2 \\ \vdots & \vdots & \ddots & \vdots \\ (v_1)_n & (v_2)_n & \cdots & (v_n)_n \end{pmatrix} . \tag{22}$$

Finally, we use eq. (16) to conclude that $[A]_{\mathcal{C}} = P^{-1}[A]_{\mathcal{B}_s} P$. Denoting the diagonalized matrix by $D \equiv [A]_{\mathcal{C}}$ and the matrix A with respect to the standard basis by $A \equiv [A]_{\mathcal{B}_s}$, then it follows that,

$$P^{-1}AP = D \,, \tag{23}$$

where P is the matrix whose columns are the eigenvectors of A and D is the diagonal matrix whose diagonal elements are the eigenvalues of A. Thus, we have succeeded in diagonalizing an arbitrary semisimple matrix.

Note that the matrix P is not unique. In eq. (22), the matrix P was specified by identifying the columns of P with the eigenvectors of A. First, one can permute the columns of A. This will have the effect of permuting the diagonal elements of D. Second, each of the eigenvectors is not unique, since if $\vec{\boldsymbol{v}}$ is an eigenvector of A corresponding to eigenvalue λ, then $c\,\vec{\boldsymbol{v}}$ is also an eigenvector for any value of $c \neq 0$. Thus, if we multiply any column of P by a nonzero constant, then both P and P^{-1} will change, but the diagonal matrix D does not change (since the eigenvalues of A are not affected by this operation). In the case of a degenerate eigenvalue, one can go further. Consider an m-fold degenerate eigenvalue. Corresponding to this eigenvalue, there exist m linearly independent eigenvectors.[8] One choice of the m linearly independent eigenvectors has been made in assembling the corresponding m column of P. However any linear combination of these eigenvectors is also an eigenvector, so there are an infinite number of possible choices for the m linearly independent columns of P. Once again, constructing the matrix P with a different choice for these m columns does not affect the diagonal matrix D in eq. (23).

[8]If fewer eigenvectors existed, then the total number of linearly independent eigenvectors of A would be less than n, in which case A would be a defective matrix and hence nondiagonalizable.

The discussion of matrix diagonalization can be extended as follows. Suppose that A and B are both semisimple matrices and hence diagonalizable. Under what circumstances can A and B be *simultaneously diagonalized*? That is, when does an invertible matrix P exist such that both $P^{-1}AP = D_1$ and $P^{-1}BP = D_2$ are diagonal? Note that if this scenario is realized, the two diagonal matrices, D_1 and D_2, can be different. Simultaneous diagonalization occurs if both $[A]_{\mathcal{C}}$ and $[B]_{\mathcal{C}}$ are diagonal with respect to the *same* basis $\mathcal{C} = \{\vec{\boldsymbol{v}}_1\,, \vec{\boldsymbol{v}}_2\,, \vec{\boldsymbol{v}}_3\,, \ldots\,, \vec{\boldsymbol{v}}_n\}$. That is, the matrices A and B must share the same eigenvectors that comprise the basis $\mathcal{C}$. The relevant theorem states that[9]

> Two semisimple matrices A and B are simultaneously diagonalizable if and only if A and B commute (i.e., $AB = BA$).

Boas addresses the essential points on p. 158 at the end of Section 11 of Chapter 3. Indeed, if all the eigenvalues of A and B are nondegenerate, then the proof that A and B must share the same eigenvectors is almost immediate. Simply observe that if $A\vec{\boldsymbol{v}} = \lambda\vec{\boldsymbol{v}}$, then $BA\vec{\boldsymbol{v}} = \lambda B\vec{\boldsymbol{v}}$. Since A and B commute, we also have $BA\vec{\boldsymbol{v}} = AB\vec{\boldsymbol{v}}$. Hence,

$$A(B\vec{\boldsymbol{v}}) = \lambda(B\vec{\boldsymbol{v}})\,.$$

Our judicious placement of the parentheses is meant to convince you that either $B\vec{\boldsymbol{v}}$ is an eigenvector of A or $B\vec{\boldsymbol{v}} = \vec{\boldsymbol{0}}$. Since λ is a nondegenerate eigenvalue of A by assumption, one can conclude that $B\vec{\boldsymbol{v}}$ must be proportional to $\vec{\boldsymbol{v}}$. That is, $B\vec{\boldsymbol{v}} = c\,\vec{\boldsymbol{v}}$ for some constant c, which implies that $\vec{\boldsymbol{v}}$ is also an eigenvector of B. The case of degenerate eigenvalues is more delicate and requires a more careful analysis.

Finally, we consider the scenario where the matrix A is defective. In this case, the eigenvectors of A do *not* span the vector space V. Hence, the eigenvectors do not provide a basis $\mathcal{C}$ for evaluating the matrix $[A]_{\mathcal{C}}$, and one must conclude that the matrix A is nondiagonalizable, i.e., no invertible matrix P exists such that $P^{-1}AP$ is diagonal. The simplest example of a defective matrix is

$$B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\,. \tag{24}$$

The characteristic equation is given by

$$\det\,(B - \lambda\mathbf{I}) = 0 \qquad \Longrightarrow \qquad \lambda^2 = 0\,.$$

Hence, the eigenvalues of B are given by the degenerate root $\lambda = 0$ of the characteristic equation. However, solving the eigenvalue equation, $B\vec{\boldsymbol{v}} = 0$, yields

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 0\,,$$

whose solution is $v_2 = 0$. Hence, there is only one linearly independent eigenvector which is proportional to $\left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$. Thus, we can conclude that B is not diagonalizable.

A test to see whether or not a general matrix A is diagonalizable is presented in Appendix B.

[9]For a rigorous proof of this theorem, see e.g., pp. 209–210 of Stephan Ramon Garcia and Roger A. Horn, *A Second Course in Linear Algebra* (Cambridge University Press, Cambridge, UK, 2017).

4. Implications of the inner product

Nothing in sections 1–3 requires the existence of an inner product. However, if an inner product is defined, then the vector space V is promoted to an inner product space. Denoting the scalar product of two vectors $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ by $\langle\vec{\boldsymbol{v}},\vec{\boldsymbol{w}}\rangle$, the scalar product must obey the following three properties:

1. $\langle\vec{\boldsymbol{u}}, a\vec{\boldsymbol{v}}+b\vec{\boldsymbol{w}}\rangle = a\langle\vec{\boldsymbol{u}},\vec{\boldsymbol{v}}\rangle + b\langle\vec{\boldsymbol{u}},\vec{\boldsymbol{w}}\rangle\,, \quad \text{for } a,b\in\mathbb{C}\,,$
2. $\langle\vec{\boldsymbol{v}},\vec{\boldsymbol{w}}\rangle = \langle\vec{\boldsymbol{w}},\vec{\boldsymbol{v}}\rangle^*\,,$
3. $\langle\vec{\boldsymbol{v}},\vec{\boldsymbol{v}}\rangle \geq 0$ (and $\langle\vec{\boldsymbol{v}},\vec{\boldsymbol{v}}\rangle = 0 \iff \vec{\boldsymbol{v}}=\vec{\boldsymbol{0}}$).

Note that properties 1 and 2 above imply that,

$$\langle a\vec{\boldsymbol{v}}+b\vec{\boldsymbol{w}},\vec{\boldsymbol{u}}\rangle = a^*\langle\vec{\boldsymbol{v}},\vec{\boldsymbol{u}}\rangle + b^*\langle\vec{\boldsymbol{w}},\vec{\boldsymbol{u}}\rangle\,, \quad \text{for } a,b\in\mathbb{C}, \tag{25}$$

where the asterisk denotes complex conjugation. Finally, the length (or magnitude) of a vector is defined by $\|v\| \equiv \langle\vec{\boldsymbol{v}},\vec{\boldsymbol{v}}\rangle^{1/2}$. In light of property 3 above, the length of any nonzero vector is real and positive, whereas the length of the zero vector is equal to zero.

One can now define the concepts of orthogonality and orthonormality. The vectors $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ are orthogonal if $\langle\vec{\boldsymbol{v}},\vec{\boldsymbol{w}}\rangle = 0$. If in addition $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ are unit vectors then $\|\vec{\boldsymbol{v}}\| = \|\vec{\boldsymbol{w}}\| = 1$. Two unit vectors that are orthogonal are called orthonormal.

Staring from a basis $\mathcal{B}$, one can employ the Gram-Schmidt process to construct an orthonormal basis. Thus, when working with inner product spaces, it is often convenient to choose an orthonormal basis. For example, the standard inner product in a complex n-dimensional vector space is $\langle\vec{\boldsymbol{v}},\vec{\boldsymbol{w}}\rangle = \sum_{i=1}^n v_i^* w_i$, where v_i and w_i are the components of the vectors $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$, respectively, with respect to any orthonormal basis.[10]

Even with the restriction of an orthonormal basis, one can examine the effect of changing basis from one orthonormal basis to another. All the considerations of Section 2 apply, with the constraint that the matrix P is now a unitary matrix. To prove this last assertion, start with eq. (8) that relates the two basis choices $\mathcal{B}$ and $\mathcal{C}$,

$$\vec{\boldsymbol{c}}_j = \sum_{i=1}^n P_{ij}\vec{\boldsymbol{b}}_i\,, \qquad j=1,2,3,\ldots,n\,. \tag{26}$$

Under the assumption that both $\mathcal{B}$ and $\mathcal{C}$ are orthonormal bases, these bases consists of mutually orthogonal unit vectors, $\langle\vec{\boldsymbol{b}}_i,\vec{\boldsymbol{b}}_\ell\rangle = \delta_{i\ell}$ and $\langle\vec{\boldsymbol{c}}_j,\vec{\boldsymbol{c}}_k\rangle = \delta_{jk}$. Using eq. (26) and the properties of the inner product listed above, it follows that

$$\begin{aligned}\delta_{jk} = \langle\vec{\boldsymbol{c}}_j,\vec{\boldsymbol{c}}_k\rangle &= \left\langle \sum_{i=1}^n P_{ij}\vec{\boldsymbol{b}}_i\,,\, \sum_{\ell=1}^n P_{\ell k}\vec{\boldsymbol{b}}_\ell\right\rangle = \sum_{i=1}^n\sum_{\ell=1}^n \langle P_{ij}\vec{\boldsymbol{b}}_i\,,\, P_{\ell k}\vec{\boldsymbol{b}}_\ell\rangle \\ &= \sum_{i=1}^n\sum_{\ell=1}^n P_{ij}^*P_{\ell k}\langle\vec{\boldsymbol{b}}_i,\vec{\boldsymbol{b}}_\ell\rangle = \sum_{i=1}^n\sum_{\ell=1}^n P_{ij}^*P_{\ell k}\delta_{i\ell} = \sum_{i=1}^n P_{ij}^*P_{ik}\,.\end{aligned} \tag{27}$$

[10] A useful exercise for the reader is to investigate how the standard inner product is modified if one employs a basis that is not orthonormal.

Since $P_{ij}^* = P_{ji}^{*\mathsf{T}} = P_{ji}^\dagger$ by the definition of the transpose and the hermitian conjugate, eq. (27) yields

$$\sum_{i=1}^n P_{ji}^\dagger P_{ik} = \sum_{i=1}^n (P^\dagger P)_{jk} = \delta_{jk}\,, \quad \Longrightarrow \quad P^\dagger P = \mathbf{I}\,, \tag{28}$$

after using the definition of matrix multiplication. Hence, P is a unitary matrix as claimed above.[11] Consequently, we can conclude that the transformation between any two orthonormal bases is unitary.

In Section 3, we showed that any $n\times n$ matrix A that possesses n linearly independent eigenvectors is diagonalizable. That is, $P^{-1}AP$ is diagonal, where P is the matrix that transforms the standard basis $\mathcal{B}_s$ to a basis consisting of the n eigenvectors of A. Note that the standard basis $\mathcal{B}_s$ is an orthonormal basis. The following question then naturally arises—which complex matrices have the property that their eigenvectors comprise an orthonormal basis that spans the inner product space V? This question is answered by eq. (11.28) on p. 154 of Boas, which states that:

> A matrix can be diagonalized by a unitary similarity transformation if and only if it is normal, i.e., the matrix commutes with its hermitian conjugate.[12]

In light of the discussion of Section 3, it then follows that for any normal matrix A (which satisfies $AA^\dagger = A^\dagger A$), there exists a diagonalizing matrix U such that

$$U^\dagger A U = D\,,$$

where U is a unitary matrix whose columns are the orthonormal eigenvectors of A and D is a diagonal matrix whose diagonal elements are the eigenvalues of A. The matrix U is not unique (details are left to the reader). Note that if A is hermitian ($A = A^\dagger$), then its eigenvalues, and hence the diagonal matrix D, are real, as shown in Appendix C.

In the case of a real vector space, a related question arises—which real matrices A have the property that their eigenvectors are real vectors that comprise an orthonormal basis that spans the real inner product space V? This question is answered by eq. (11.27) on p. 154 of Boas, which states that:

> A real matrix can be diagonalized by a real orthogonal similarity transformation if and only if it is a real symmetric matrix.[13]

That is, for any real symmetric matrix A (which satisfies $A = A^\mathsf{T}$) there exists a real diagonalizing matrix Q such that

$$Q^\mathsf{T} A Q = D\,,$$

where Q is a real orthogonal matrix whose columns are the orthonormal eigenvectors of A and D is a real diagonal matrix whose diagonal elements are the eigenvalues of A.

[11]Note that our analysis above applies both to real and complex vector spaces. If V is a real vector space then P is a real matrix, in which case eq. (28) implies that P is a real orthogonal matrix that satisfies $P^\mathsf{T} P = \mathbf{I}$. Hence, the transformation between two real orthonormal bases is orthogonal.

[12]A proof of this statement can be found in Philip A. Macklin, "Normal Matrices for Physicists," *American Journal of Physics* **52**, 513–515 (1984). A link to this article can be found in Section IX of the course website.

[13]Note that a real symmetric matrix is also hermitian since $A^\dagger = A^\mathsf{T}$ if the matrix A is real.

APPENDIX A: Proof that if all the eigenvalues of the matrix A are nondegenerate then the corresponding eigenvectors are linearly independent

If $\vec{\boldsymbol{v}}_i$ is an eigenvector of the $n \times n$ matrix A with eigenvalue λ_i, then

$$A\vec{\boldsymbol{v}}_i = \lambda_i \vec{\boldsymbol{v}}_i\,, \qquad \vec{\boldsymbol{v}}_i \neq 0 \quad \text{for} \quad i = 1, 2, \ldots, n\,. \tag{29}$$

We shall examine the case in which all the eigenvalues of the matrix A are nondegenerate (i.e., $\lambda_i \neq \lambda_j$ for $i \neq j$), in which case we say that A is simple. Our goal is to prove that the set consisting of the n eigenvectors of A must be a linearly independent set. The strategy of the following proof is to assume the opposite of our hypothesis and show that a contradiction ensues.

Suppose that the set consisting of the n eigenvectors of A is a linearly dependent set. Choose a subset of these eigenvectors that contains the maximal number of linearly independent vectors. This set will contain r vectors where $1 \leq r < n$. For convenience, we shall order the list of eigenvectors of A such that the set $\{\vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2 \ldots, \vec{\boldsymbol{v}}_r\}$ is a linearly independent set. This means that it must be possible to express $\vec{\boldsymbol{v}}_n$ as a linear combination of $\{\vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2 \ldots, \vec{\boldsymbol{v}}_r\}$. That is,

$$\vec{\boldsymbol{v}}_n = \sum_{i=1}^{r} a_i \vec{\boldsymbol{v}}_i\,, \tag{30}$$

where at least one of the a_i in eq. (30) must be nonzero. Multiplying eq. (30) on the left by $A - \lambda_n \mathbf{I}$, and using eq. (29) yields,

$$\vec{\mathbf{0}} = \sum_{i=1}^{r} a_i (A\vec{\boldsymbol{v}}_i - \lambda_n \vec{\boldsymbol{v}}_i) = \sum_{i=1}^{r} a_i (\lambda_i - \lambda_n) \vec{\boldsymbol{v}}_i\,. \tag{31}$$

Recall the definition of linear independence which states that the set $\{\vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2 \ldots, \vec{\boldsymbol{v}}_r\}$ is a linearly independent set if

$$\sum_{i=1}^{r} c_i \vec{\boldsymbol{v}}_i = \vec{\mathbf{0}} \quad \Longrightarrow \quad c_i = 0 \text{ for all } i = 1, 2, \ldots, r. \tag{32}$$

Since $\lambda_i \neq \lambda_n$ for $i = 1, 2, \ldots r$ (with $r < n$) by the condition specified below eq. (29), we can conclude from eqs. (31) and (32) that $a_i = 0$ for $i = 1, 2, \ldots, r$. Plugging this result back into eq. (30) then yields $\vec{\boldsymbol{v}}_n = \vec{\mathbf{0}}$, which is impossible since the zero vector is never an eigenvector.

We have arrived at a contradiction. Hence, the initial assumption that the set consisting of the n eigenvectors of A is a linearly dependent set must be false. Consequently, one can conclude that the set consisting of the n eigenvectors of A must be a linearly independent set.

REMARK: The converse of the result obtained above is not true in general. That is, if A possesses degenerate eigenvalues, then it is still possible that there exist n linearly independent eigenvectors. If such a linearly independent set exists, we say that A is semisimple. Otherwise, we say that A is defective. An example of the latter was provided by eq. (24).

APPENDIX B: A test for diagonalizability of a matrix

In Section 3, we stated that an $n \times n$ matrix A is diagonalizable if and only if it possesses n linearly independent eigenvectors. In this case, we say that A is semisimple. Note that a subset of semisimple matrices consists of those matrices that possess n nondegenerate eigenvalues, in which case we say that the matrices are simple and the theorem of Appendix A applies. Finally, matrices that possess m linearly independent eigenvectors, where $m < n$, are not diagonalizable and are called defective. Eq. (24) provides the simplest example of a defective matrix.

The question now arises: is there a simple test for the diagonalizability of a matrix. In general, the answer is negative. Based on the information of the previous paragraph above, in order to check whether or not an $n \times n$ matrix A is diagonalizable, one must first compute the eigenvalues of A. If the n eigenvalues are nondegenerate, then A is diagonalizable. If one or more of the eigenvalues of A are degenerate, then the theorem of Appendix A does not apply. Thus, in the case where degenerate eigenvalues are present, one must determine the corresponding eigenvectors and show that A possesses n linearly independent eigenvectors.[14] Returning to the example of eq. (24), we see that there is a doubly degenerate eigenvalue, $\lambda = 0$, but only one corresponding eigenvector. Hence, in this example the matrix is not diagonalizable.

However, one can devise a slightly more efficient algorithm for determining whether an $n \times n$ matrix A with degenerate eigenvalues is diagonalizable. This relies on the following theorem:

<u>Theorem</u>: Consider an $n \times n$ matrix A, with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. If some of the eigenvalues are degenerate, which means that there are some duplicate elements in the set $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$, then create a new list consisting of the distinct eigenvalues of A,

$$\{\lambda_1, \lambda_2, \ldots, \lambda_p\}, \quad \text{where } p \text{ is the number of distinct eigenvalues of } A. \tag{33}$$

The $n \times n$ matrix A is semisimple (and hence diagonalizable), i.e., it possesses n linearly independent eigenvectors, if and only if the following product of matrices is equal to the zero matrix (denoted by $\mathbf{0}$ below),

$$(A - \lambda_1 \mathbf{I})(A - \lambda_2 \mathbf{I}) \cdots (A - \lambda_p \mathbf{I}) = \mathbf{0}\,, \tag{34}$$

where only the p distinct eigenvalues of A appear on the left hand of eq. (34). Consequently, if the product of the matrices on the left hand side of eq. (34) is a nonzero matrix, then A is defective and thus is not diagonalizable.[15]

That is, eq. (34) provides a test for the diagonalizability of the $n \times n$ matrix A. For example, consider a 2×2 matrix with degenerate eigenvalues, $\lambda \equiv \lambda_1 = \lambda_2$. Then, eq. (34) implies that A is diagonalizable if and only if $A = \lambda \mathbf{I}$. In particular, the 2×2 matrix defined in eq. (24) is not of this form and hence is nondiagonalizable.

[14] Equivalently, one must show that there exist m linearly independent eigenvectors corresponding to each m-fold degenerate eigenvalue, respectively.

[15] A proof of this theorem can be found in Section 8.3.2 of the book by James B. Carrell, *Groups, Matrices, and Vector Spaces* (Springer Science, New York, NY, 2017).

Note that if A is a simple matrix, i.e., none of the eigenvalues of A are degenerate, then $p = n$ and eq. (34) is equivalent to the Cayley-Hamilton theorem, which states that any matrix A satisfies $p(A) = \mathbf{0}$, where $p(\lambda)$ is the characteristic equation of A.[16]

The theorem above provides a test for the diagonalizability of a general matrix A by a similarity transformation. That is, if eq. (34) is satisfied, then an invertible matrix P exists such that $P^{-1}AP$ is a diagonal matrix [cf. eq. (23)]. Note that since the columns of P consist of the n linearly-independent eigenvectors of A [cf. eq. (22)], one must determine explicitly the eigenvectors of A to complete the diagonalization procedure.

The following example illustrates the theorem quoted above. Consider the matrix,

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} . \tag{35}$$

The characteristic equation is

$$\det(A - \lambda\mathbf{I}) = \begin{pmatrix} -\lambda & 1 & 1 \\ 1 & -\lambda & 1 \\ 1 & 1 & -\lambda \end{pmatrix} = -\lambda^3 + 3\lambda + 2 = -(\lambda - 2)(\lambda + 1)^2 . \tag{36}$$

Thus, the eigenvalues are $\lambda_1 = 2$ and $\lambda_2 = \lambda_3 = -1$. That is, the eigenvalue $\lambda = -1$ is doubly degenerate. To check whether A is diagonalizable, we employ eq. (34),

$$(A - 2\mathbf{I})(A + \mathbf{I}) = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} . \tag{37}$$

Indeed, eq. (34) is satisfied, and hence we can conclude that the matrix A is diagonalizable. This means that A must possess three linearly independent eigenvectors. One can easily check that $(1, 1, 1)^{\mathsf{T}}$ is an eigenvector corresponding to the nondegenerate eigenvalue $\lambda = 2$ and any linear combination of the two vectors, $(1, 0, -1)^{\mathsf{T}}$ and $(0, 1, -1)^{\mathsf{T}}$, is an eigenvector corresponding to the degenerate eigenvalue $\lambda = -1$. It follows that one possible choice for the diagonalizing matrix P is given by

$$P = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \end{pmatrix} . \tag{38}$$

Finally, to provide the explicit diagonalization of A, one must compute P^{-1} and verify that,

$$P^{-1}AP = \frac{1}{3} \begin{pmatrix} 1 & 1 & 1 \\ 2 & -1 & -1 \\ -1 & 2 & -1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} . \tag{39}$$

As expected, the elements of the resulting diagonal matrix are the eigenvalues of A.

[16] See the class handout entitled *The characteristic polynomial*.

In contrast, the test for the unitary diagonalization of a $n \times n$ matrix A (or equivalently, the existence of n linearly independent eigenvectors that constitute an orthonormal set of vectors) is much simpler than the test given by eq. (34). We refer the reader back to Section 4, where the following result was given. The $n \times n$ matrix A is diagonalizable by a unitary similarity transformation (i.e., a unitary matrix U exists such that $U^\dagger AU$ is a diagonal matrix) if and only if A is normal (i.e., $AA^\dagger = A^\dagger A$). A special case of this result that is relevant for real matrices states that a real $n \times n$ matrix A is diagonalizable by a real orthogonal similarity transformation (i.e., a real orthogonal matrix Q exists such that $Q^{\mathsf{T}}AQ$ is a real diagonal matrix) if and only if A is a real symmetric matrix (i.e., $A^{\mathsf{T}} = A$). In both these cases, one must only verify that the matrix A satisfies a simple property (either A is a normal matrix or A is a real symmetric matrix) in order to conclude whether a unitary diagonalization of A is possible.

Since a unitary matrix satisfies $U^\dagger = U^{-1}$ and a real orthogonal matrix satisfies $Q^{\mathsf{T}} = Q^{-1}$, it follows that unitary diagonalization is a special case of the diagonalization of a matrix by a similarity transformation. Consequently, in the case of the unitary diagonalization of A, eq. (34) must still hold true. Nevertheless, it is much easier to determine whether it is possible to diagonalize a matrix A by a unitary or real orthogonal similarity transformation (by checking that $A^\dagger A = AA^\dagger$ or $A^{\mathsf{T}} = A$, respectively) as compared to the diagonalization of a general matrix by a similarity transformation (which is possible only if eq. (34) holds true). Of course, in the case of the unitary diagonalization of a normal matrix A, one must still determine the orthonormal eigenvectors of A in order to construct the columns of the unitary diagonalizing matrix U (or the real orthogonal diagonalizing matrix Q if the matrix A is real and symmetric).

Note that the matrix A given by eq. (35) is not a normal matrix, since $AA^\dagger \neq A^\dagger A$. Hence in this example, the unitary diagonalization of A is not possible.

APPENDIX C: The eigenvalues of an hermitian matrix are real

An hermitian matrix A satisfies the condition $A = A^\dagger$. Thus, A is a normal matrix and thus can be diagonalized by a unitary similarity transformation. But, hermitian matrices possess one additional feature not shared by nonhermitian normal matrices. Namely, the eigenvalues of an hermitian matrix must be real.

The proof of this statement is straightforward. Recall that if A is hermitian and $A = [a_{ij}]$, then $a_{ij} = a^*_{ji}$ with respect to any orthonormal basis. Thus, employing the standard inner product of Section 4,

$$\langle \vec{\boldsymbol{v}}, A\vec{\boldsymbol{v}} \rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} v_i^*(a_{ij}v_j) = \sum_{i=1}^{n}\sum_{j=1}^{n}(a^*_{ji}v_i^*)a_j = \langle A\vec{\boldsymbol{v}}, \vec{\boldsymbol{v}} \rangle \,. \tag{40}$$

Hence, if $\vec{\boldsymbol{v}}$ is a normalized eigenvector of A (i.e., $\langle \vec{\boldsymbol{v}}, \vec{\boldsymbol{v}} \rangle = 1$) with eigenvalue λ, then

$$\lambda = \langle \vec{\boldsymbol{v}}, \lambda\vec{\boldsymbol{v}} \rangle = \langle \vec{\boldsymbol{v}}, A\vec{\boldsymbol{v}} \rangle = \langle A\vec{\boldsymbol{v}}, \vec{\boldsymbol{v}} \rangle = \langle \lambda\vec{\boldsymbol{v}}, \vec{\boldsymbol{v}} \rangle = \lambda^* \,, \tag{41}$$

in light of eqs. (20) and (25). That is, $\lambda = \lambda^*$, which implies that λ is real. Likewise, the eigenvalues of a real symmetric matrix ($A = A^{\mathsf{T}}$) are real [by taking A real in eq. (41)].