Coordinates, matrix elements and changes of basis

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.*

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:†

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

The subscript s indicates that this is the standard basis. The superscript T turns the row vectors into column vectors. Thus,

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

The v_i are called the *coordinates* of $\vec{\boldsymbol{v}}$ 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

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

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

*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 $W = V$.

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

express $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ as linear combinations of basis vectors, then

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

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

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

where we have used eq. (1) to substitute for w_i. It follows that:

$$\sum_{j=1}^{n} \left(A\widehat{\boldsymbol{x}}_j - \sum_{i=1}^{n} a_{ij} \widehat{\boldsymbol{x}}_i \right) v_j = 0 \,. \tag{2}$$

Eq. (2) must be true for *any* vector $\vec{\boldsymbol{v}} \in V$; that is, for any choice of components v_j. Thus, the coefficient inside the parentheses in eq. (2) must vanish. We conclude that:

$$A\widehat{\boldsymbol{x}}_j = \sum_{i=1}^{n} a_{ij} \widehat{\boldsymbol{x}}_i \,. \tag{3}$$

Eq. (3) can be used as the definition of the matrix elements a_{ij} with respect to the standard basis of a linear operator A.

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

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

For any vector $\vec{\boldsymbol{v}} \in V$, we can find a unique set of coefficients v_i such that

$$\vec{\boldsymbol{v}} = \sum_{j=1}^{n} v_j \, \vec{\boldsymbol{b}}_j \,. \tag{4}$$

The v_i are the *coordinates* of $\vec{\boldsymbol{v}}$ with respect to the basis $\mathcal{B}$. Likewise, for any linear operator A,

$$A\vec{\boldsymbol{b}}_j = \sum_{i=1}^{n} a_{ij} \vec{\boldsymbol{b}}_i \tag{5}$$

defines the *matrix elements* of the linear operator A with respect to the basis $\mathcal{B}$. Clearly, these more general definitions reduce to the previous ones given in the case of the standard basis. Moreover, we can easily compute $A\vec{\boldsymbol{v}} \equiv \vec{\boldsymbol{w}}$ using the results of eqs. (4) and (5):

$$A\vec{\boldsymbol{v}} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}v_j\vec{\boldsymbol{b}}_i = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} a_{ij}v_j\right)\vec{\boldsymbol{b}}_i = \sum_{i=1}^{n} w_i\vec{\boldsymbol{b}}_i = \vec{\boldsymbol{w}}\,,$$

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. (1)]. One must simply be consistent and always employ the *same* basis for defining the vector components and the matrix elements of a linear operator.

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 ones 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 decide to employ a different basis choice $\mathcal{C}$:

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

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 components 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 components 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:

$$\vec{\boldsymbol{c}}_j = \sum_{i=1}^{n} P_{ij}\vec{\boldsymbol{b}}_i\,,\qquad j = 1, 2, 3, \ldots, n\,. \tag{6}$$

Note that eq. (6) 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,

$$\vec{\boldsymbol{b}}_k = \sum_{j=1}^{n} (P^{-1})_{jk} \vec{\boldsymbol{c}}_j \,, \qquad k = 1, 2, 3, \ldots, n \,. \tag{7}$$

We are now in the position to determine the components 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 components 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 components by v'_i and the matrix elements by a'_{ij}. Then, using the definition of vector components [eq. (4)] and matrix elements [eq. (5)],

$$\vec{\boldsymbol{v}} = \sum_{j=1}^{n} v'_j \vec{\boldsymbol{c}}_j = \sum_{j=1}^{n} v'_j \sum_{i=1}^{n} P_{ij} \vec{\boldsymbol{b}}_i = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} P_{ij} v'_j \right) \vec{\boldsymbol{b}}_i = \sum_{i=1}^{n} v_i \vec{\boldsymbol{b}}_i \,, \tag{8}$$

where we have used eq. (6) to express the $\vec{\boldsymbol{c}}_j$ in terms of the $\vec{\boldsymbol{b}}_i$. The last step in eq. (8) can be rewritten as:

$$\sum_{i=1}^{n} \left(v_i - \sum_{j=1}^{n} P_{ij} v'_j \right) \vec{\boldsymbol{b}}_i = 0 \,. \tag{9}$$

Since the $\vec{\boldsymbol{b}}_i$ are linearly independent, the coefficient inside the parentheses in eq. (9) must vanish. Hence,

$$v_i = \sum_{j=1}^{n} P_{ij} v'_j \,, \qquad \text{or equivalently} \qquad \boxed{[\vec{\boldsymbol{v}}]_{\mathcal{B}} = P[\vec{\boldsymbol{v}}]_{\mathcal{C}}} \,. \tag{10}$$

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

$$v'_j = \sum_{k=1}^{n} (P^{-1})_{jk} v_k \,, \qquad \text{or equivalently} \qquad \boxed{[\vec{\boldsymbol{v}}]_{\mathcal{C}} = P^{-1}[\vec{\boldsymbol{v}}]_{\mathcal{B}}} \,. \tag{11}$$

Thus, we have determined the relation between the components 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. (5)], and the matrix elements of A with respect to the basis $\mathcal{C}$, which we denote by a'_{ij}:

$$A\vec{\boldsymbol{c}}_j = \sum_{i=1}^{n} a'_{ij} \vec{\boldsymbol{c}}_i \,. \tag{12}$$

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. (6) and (7) and the definition of the matrix elements of A with respect to the basis $\mathcal{C}$ [eq. (12)]. Comparing this result with eq. (5), 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:

$$
\boxed{[A]_{\mathcal{B}} = P[A]_{\mathcal{C}}P^{-1}\,.} \tag{13}
$$

The meaning of this equation is that 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. (13). We can invert eq. (13) to obtain:

$$
\boxed{[A]_{\mathcal{C}} = P^{-1}[A]_{\mathcal{B}}P\,.} \tag{14}
$$

In fact, there is a much faster method to derive eqs. (13) and (14). 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. (10), $[\vec{\boldsymbol{w}}]_{\mathcal{B}} = [A]_{\mathcal{B}}[\vec{\boldsymbol{v}}]_{\mathcal{B}}$ 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

$$
\bigl\{[A]_{\mathcal{C}} - P^{-1}[A]_{\mathcal{B}}P\bigr\}[\vec{\boldsymbol{v}}]_{\mathcal{C}} = 0\,. \tag{15}
$$

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. (15) must vanish. This yields eq. (14).

The significance of eq. (14) 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. (6). 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.

Example: Let $\mathcal{B}$ be the standard basis and let $\mathcal{C} = \big\{(1,0,0)\,,\,(1,1,0)\,,\,(1,1,1)\big\}$. Given a linear operator A whose matrix elements with respect to the basis $\mathcal{B}$ are:

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

we shall determine $[A]_{\mathcal{C}}$. First, we need to work out P. Noting that:

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

it follows from eq. (6) that

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

Inverting, $\vec{\boldsymbol{b}}_1 = \vec{\boldsymbol{c}}_1\,,\ \vec{\boldsymbol{b}}_2 = \vec{\boldsymbol{c}}_2 - \vec{\boldsymbol{c}}_1\,,$ and $\vec{\boldsymbol{b}}_3 = \vec{\boldsymbol{c}}_3 - \vec{\boldsymbol{c}}_2$, so that eq. (7) yields:

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

Thus, using eq. (15), we obtain:

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

3. Application to matrix diagonalization

Consider the eigenvalue problem for a matrix A:

$$A\vec{\boldsymbol{v}}_j = \lambda_j \vec{\boldsymbol{v}}_j\,, \qquad \vec{\boldsymbol{v}}_j \neq 0 \quad \text{for} \quad j = 1, 2, \ldots, n\,. \tag{16}$$

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 non-degenerate, then A is called *simple*. In this case, the n eigenvectors are linearly independent and span the vector space V.‡ 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 *semi-simple*.§ 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 semi-simple.

Since the eigenvectors of a semi-simple matrix A span the vector space V, we may choose the eigenvectors of A as a basis. Suppose we are given the matrix elements of A with respect to the standard basis $\mathcal{B}_s = \left\{\widehat{\boldsymbol{x}}_1\,, \widehat{\boldsymbol{x}}_2\,, \widehat{\boldsymbol{x}}_3\,, \ldots\,, \widehat{\boldsymbol{x}}_n\right\}$. We shall then compute the matrix elements of A with respect to a new basis $\mathcal{C} = \left\{\vec{\boldsymbol{v}}_1\,, \vec{\boldsymbol{v}}_2\,, \vec{\boldsymbol{v}}_3\,, \ldots\,, \vec{\boldsymbol{v}}_n\right\}$, where the $\vec{\boldsymbol{v}}_i$ are the eigenvectors of A. To determine $[A]_{\mathcal{C}}$, we use eq. (12):

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

But, eq. (16) implies that $a'_{ij} = \lambda_j \delta_{ij}$. That is,

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

Finally, we shall determine the matrix P that governs the relation between $\mathcal{B}_s$ and $\mathcal{C}$ [eq. (6)]. Consider the coordinates of $\vec{\boldsymbol{v}}_j$ with respect to the standard basis $\mathcal{B}_s$:

$$\vec{\boldsymbol{v}}_j = \sum_{i=1}^{n} (\vec{\boldsymbol{v}}_j)_i\, \widehat{\boldsymbol{x}}_i = \sum_{i=1}^{n} P_{ij}\widehat{\boldsymbol{x}}_i\,, \tag{17}$$

where $(\vec{\boldsymbol{v}}_j)_i$ is the ith coordinate of the jth eigenvector. Using eq. (17), we identify $P_{ij} = (\vec{\boldsymbol{v}}_j)_i$. In matrix form,

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

‡This result is proved in the appendix to these notes.

§Note that if A is semi-simple, then A is also simple only if the eigenvalues of A are distinct.

Finally, we use eq. (14) to conclude that $[A]_{\mathcal{C}} = P^{-1}AP$. If we denote the diagonalized matrix by $D \equiv [A]_{\mathcal{C}}$, then

$$P^{-1}AP = D\,, \tag{18}$$

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 semi-simple matrix.

If the eigenvectors of A do *not* span the vector space V (*i.e.*, A is defective), then A is not diagonalizable.[¶] That is, there does not exist a matrix P and a diagonal matrix D such that eq. (18) is satisfied.

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. In this case, we can define the concepts of orthogonality and orthonormality. In particular, given an arbitrary basis $\mathcal{B}$, we can use the Gram-Schmidt process to construct an orthonormal basis. Thus, when considering inner product spaces, it is convenient to always choose an orthonormal basis.

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.[‖] Namely, the transformation between any two orthonormal bases is always unitary.

The following question naturally arises—which 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. if the matrix commutes with its hermitian conjugate.[**]

Then, following the arguments of section 3, it follows that for any normal matrix A (which satisfies $AA^\dagger = A^\dagger A$), there exists a diagonalizing matrix U such that

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

where U is the unitary matrix ($U^\dagger = U^{-1}$) whose columns are the orthonormal eigenvectors of A and D is the diagonal matrix whose diagonal elements are the eigenvalues of A.

[¶] The simplest example of a defective matrix is $B = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$. One can quickly check that the eigenvalues of B are given by the double root $\lambda = 0$ of the characteristic equation. However, solving the eigenvalue equation, $B\vec{\boldsymbol{v}} = 0$, yields only one linearly independent eigenvector, $\left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$. One can verify explicitly that no matrix P exists such that $P^{-1}BP$ is diagonal.

[‖] In a real inner product space, a unitary transformation is real and hence an orthogonal transformation.

[**] 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 VIII of the course website.

Appendix: Proof that the eigenvectors corresponding to distinct eigenvalues are linearly independent

The statement that $\vec{\boldsymbol{v}}_i$ is an eigenvector of A with eigenvalue λ_i means that

$$A\vec{\boldsymbol{x}}_i = \lambda_i \vec{\boldsymbol{x}}_i \,.$$

We can rewrite this condition as:

$$(A - \lambda_i \mathbf{I})\vec{\boldsymbol{x}}_i = 0 \,.$$

If $\vec{\boldsymbol{x}}_1, \vec{\boldsymbol{x}}_2, \ldots, \vec{\boldsymbol{x}}_n$ are linearly independent, then

$$c_1\vec{\boldsymbol{x}}_1 + c_2\vec{\boldsymbol{x}}_2 + \cdots + c_n\vec{\boldsymbol{x}}_n = 0 \qquad \Longleftrightarrow \qquad c_i = 0 \text{ for all } i = 1, 2, \ldots, n \,. \tag{19}$$

We prove this result by assuming the contrary and arrive at a contradiction. That is, we will assume that one of the coefficients is nonzero. Without loss of generality, we shall assume that $c_1 \neq 0$ [this can always be arranged by reordering the $\{\vec{\boldsymbol{x}}_i\}$]. Multiplying both sides of eq. (19) by $A - \lambda_2\mathbf{I}$, and using the fact that

$$(A - \lambda_2\mathbf{I})\vec{\boldsymbol{x}}_i = (\lambda_i - \lambda_2)\vec{\boldsymbol{x}}_i \,,$$

we obtain:

$$c_1(\lambda_1 - \lambda_2)\vec{\boldsymbol{x}}_1 + c_3(\lambda_3 - \lambda_2)\vec{\boldsymbol{x}}_3 + \cdots + c_n(\lambda_n - \lambda_2)\vec{\boldsymbol{x}}_n = 0 \,. \tag{20}$$

Note that the term $c_2\vec{\boldsymbol{x}}_2$ that appears in eq. (19) has been removed from the sum. Next, multiply both sides of eq. (20) by $A - \lambda_3\mathbf{I}$. A similar computation yields:

$$c_1(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3)\vec{\boldsymbol{x}}_1 + c_4(\lambda_4 - \lambda_2)(\lambda_4 - \lambda_3)\vec{\boldsymbol{x}}_4 + \cdots + c_n(\lambda_n - \lambda_2)(\lambda_n - \lambda_3)\vec{\boldsymbol{x}}_n = 0 \,.$$

Note that the term $c_3\vec{\boldsymbol{x}}_3$ that appears in eq. (19) has been removed from the sum. Continuing the process of multiplying on the left successively by $A - \lambda_4\mathbf{I}$, $A - \lambda_5\mathbf{I}$,... ,$A - \lambda_n\mathbf{I}$, all the terms involving $c_i\vec{\boldsymbol{x}}_i$ will be removed with the exception of one term proportional to c_1. The end result is:

$$c_1(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3)(\lambda_1 - \lambda_4)\cdots(\lambda_1 - \lambda_n) = 0 \,. \tag{21}$$

By assumption, all the eigenvalues are distinct. Thus, eq. (21) implies that $c_1 = 0$, which contradicts our original assumption. We conclude that our assumption that at least one of the c_i is nonzero is incorrect. Hence, if all the eigenvalues are distinct, then $c_i = 0$ for all $i = 1, 2, \ldots, n$. That is, the n eigenvectors $\vec{\boldsymbol{x}}_i$ are linearly independent.

BONUS MATERIAL:

There is a more elegant way to prove that if $\vec{\boldsymbol{x}}_1, \vec{\boldsymbol{x}}_2, \ldots, \vec{\boldsymbol{x}}_n$ are eigenvectors corresponding to the distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of A, then $\vec{\boldsymbol{x}}_1, \vec{\boldsymbol{x}}_2, \ldots, \vec{\boldsymbol{x}}_n$

are linearly independent. Starting from $A\vec{\boldsymbol{x}} = \lambda\vec{\boldsymbol{x}}$, we multiply on the left by A to get

$$A^2\vec{\boldsymbol{x}} = A\cdot A\vec{\boldsymbol{x}} = A(\lambda\vec{\boldsymbol{x}}) = \lambda A\vec{\boldsymbol{x}} = \lambda^2\vec{\boldsymbol{x}}\,.$$

Continuing this process of multiplication on the left by A, we conclude that:

$$A^k\vec{\boldsymbol{x}} = A\left(A^{k-1}\vec{\boldsymbol{x}}\right) = A\left(\lambda^{k-1}\vec{\boldsymbol{x}}\right) = \lambda^{k-1}A\vec{\boldsymbol{x}} = \lambda^k\vec{\boldsymbol{x}}\,, \tag{22}$$

for $k = 2, 3, \ldots, n$. Thus, if we multiply eq. (19) on the left by A^k, then we obtain n separate equations by choosing $k = 0, 1, 2, \ldots, n-1$ given by:

$$c_1\lambda_1^k\vec{\boldsymbol{x}}_1 + c_2\lambda_2^k\vec{\boldsymbol{x}}_2 + \cdots + c_n\lambda_n^k\vec{\boldsymbol{x}}_n = 0\,, \qquad k = 0, 1, 2, \ldots, n-1\,.$$

In matrix form,

$$\begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \lambda_3 & \cdots & \lambda_n \\ \lambda_1^2 & \lambda_2^2 & \lambda_3^2 & \cdots & \lambda_n^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{n-1} & \lambda_2^{n-1} & \lambda_3^{n-1} & \cdots & \lambda_n^{n-1} \end{pmatrix} \begin{pmatrix} c_1\vec{\boldsymbol{x}}_1 \\ c_2\vec{\boldsymbol{x}}_2 \\ c_3\vec{\boldsymbol{x}}_3 \\ \vdots \\ c_n\vec{\boldsymbol{x}}_n \end{pmatrix} = 0\,. \tag{23}$$

The matrix appearing above is equal to the transpose of a well known matrix called the Vandermonde matrix. There is a beautiful formula for its determinant:

$$\begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \lambda_3 & \cdots & \lambda_n \\ \lambda_1^2 & \lambda_2^2 & \lambda_3^2 & \cdots & \lambda_n^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{n-1} & \lambda_2^{n-1} & \lambda_3^{n-1} & \cdots & \lambda_n^{n-1} \end{vmatrix} = \prod_{i<j}(\lambda_i - \lambda_j)\,. \tag{24}$$

Here, I am using the $\prod$ to symbolize multiplication, so that

$$\prod_{i<j}(\lambda_i - \lambda_j) = (\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3)\cdots(\lambda_1 - \lambda_n)(\lambda_2 - \lambda_3)\cdots(\lambda_2 - \lambda_n)\cdots(\lambda_{n-1} - \lambda_n)\,.$$

I leave it as a challenge to the reader for providing a proof of eq. (24). This result implies that if all the eigenvalues λ_i are distinct, then the determinant of the Vandermonde matrix is nonzero. In this case, the Vandermonde matrix is invertible. Multiplying eq. (23) by the inverse of the Vandermonde matrix then yields $c_i\vec{\boldsymbol{x}}_i = 0$ for all $i = 1, 2, \ldots, n$. Since the eigenvectors are nonzero by definition, it follows that $c_i = 0$ for all $i = 1, 2, \ldots, n$. Hence the $\{\vec{\boldsymbol{x}}_i\}$ are linearly independent.

Note that we can work backwards. That is, using the first proof above to conclude that the $\{\vec{\boldsymbol{x}}_i\}$ are linearly independent, it then follows that the determinant of the Vandermonde matrix must be nonzero if the λ_i are distinct.