Hence $\langle \mathbf{y}|\mathbf{y}\rangle = \langle \mathbf{x}|\mathbf{x}\rangle$, showing that the action of the linear operator represented by a unitary matrix does not change the norm of a complex vector. The action of a unitary matrix on a complex column matrix thus parallels that of an orthogonal matrix acting on a real column matrix.

8.12.7 Normal matrices

A final important set of special matrices consists of the *normal* matrices, for which

$$\mathsf{A}\mathsf{A}^\dagger = \mathsf{A}^\dagger\mathsf{A},$$

i.e. a normal matrix is one that commutes with its Hermitian conjugate.

We can easily show that Hermitian matrices and unitary matrices (or symmetric matrices and orthogonal matrices in the real case) are examples of normal matrices. For an Hermitian matrix, $\mathsf{A} = \mathsf{A}^\dagger$ and so

$$\mathsf{A}\mathsf{A}^\dagger = \mathsf{A}\mathsf{A} = \mathsf{A}^\dagger\mathsf{A}.$$

Similarly, for a unitary matrix, $\mathsf{A}^{-1} = \mathsf{A}^\dagger$ and so

$$\mathsf{A}\mathsf{A}^\dagger = \mathsf{A}\mathsf{A}^{-1} = \mathsf{A}^{-1}\mathsf{A} = \mathsf{A}^\dagger\mathsf{A}.$$

Finally, we note that, if A is normal then so too is its inverse A^{-1}, since

$$\mathsf{A}^{-1}(\mathsf{A}^{-1})^\dagger = \mathsf{A}^{-1}(\mathsf{A}^\dagger)^{-1} = (\mathsf{A}^\dagger\mathsf{A})^{-1} = (\mathsf{A}\mathsf{A}^\dagger)^{-1} = (\mathsf{A}^\dagger)^{-1}\mathsf{A}^{-1} = (\mathsf{A}^{-1})^\dagger\mathsf{A}^{-1}.$$

This broad class of matrices is important in the discussion of eigenvectors and eigenvalues in the next section.

8.13 Eigenvectors and eigenvalues

Suppose that a linear operator $\mathcal{A}$ transforms vectors $\mathbf{x}$ in an N-dimensional vector space into other vectors $\mathcal{A}\,\mathbf{x}$ in the same space. The possibility then arises that there exist vectors $\mathbf{x}$ each of which is transformed by $\mathcal{A}$ into a multiple of itself. Such vectors would have to satisfy

$$\mathcal{A}\,\mathbf{x} = \lambda\mathbf{x}. \tag{8.67}$$

Any non-zero vector $\mathbf{x}$ that satisfies (8.67) for some value of λ is called an *eigenvector* of the linear operator $\mathcal{A}$, and λ is called the corresponding *eigenvalue*. As will be discussed below, in general the operator $\mathcal{A}$ has N independent eigenvectors $\mathbf{x}^i$, with eigenvalues λ_i. The λ_i are not necessarily all distinct.

If we choose a particular basis in the vector space, we can write (8.67) in terms of the components of $\mathcal{A}$ and $\mathbf{x}$ with respect to this basis as the matrix equation

$$\mathsf{A}\mathsf{x} = \lambda\mathsf{x}, \tag{8.68}$$

where A is an $N \times N$ matrix. The column matrices x that satisfy (8.68) obviously represent the eigenvectors $\mathbf{x}$ of $\mathcal{A}$ in our chosen coordinate system. Conventionally, these column matrices are also referred to as the *eigenvectors of the matrix* A.[§] Clearly, if x is an eigenvector of A (with some eigenvalue λ) then any scalar multiple $\mu\mathsf{x}$ is also an eigenvector with the same eigenvalue. We therefore often use *normalised* eigenvectors, for which

$$\mathsf{x}^\dagger\mathsf{x} = 1$$

(note that $\mathsf{x}^\dagger\mathsf{x}$ corresponds to the inner product $\langle\mathbf{x}|\mathbf{x}\rangle$ in our basis). Any eigenvector x can be normalised by dividing all its components by the scalar $(\mathsf{x}^\dagger\mathsf{x})^{1/2}$.

As will be seen, the problem of finding the eigenvalues and corresponding eigenvectors of a square matrix A plays an important role in many physical investigations. Throughout this chapter we denote the ith eigenvector of a square matrix A by x^i and the corresponding eigenvalue by λ_i. This superscript notation for eigenvectors is used to avoid any confusion with components.

►*A non-singular matrix* A *has eigenvalues* λ_i *and eigenvectors* x^i*. Find the eigenvalues and eigenvectors of the inverse matrix* A^{-1}.

The eigenvalues and eigenvectors of A satisfy

$$\mathsf{A}\mathsf{x}^i = \lambda_i\mathsf{x}^i.$$

Left-multiplying both sides of this equation by A^{-1}, we find

$$\mathsf{A}^{-1}\mathsf{A}\mathsf{x}^i = \lambda_i\mathsf{A}^{-1}\mathsf{x}^i.$$

Since $\mathsf{A}^{-1}\mathsf{A} = \mathsf{I}$, on rearranging we obtain

$$\mathsf{A}^{-1}\mathsf{x}^i = \frac{1}{\lambda_i}\mathsf{x}^i.$$

Thus, we see that A^{-1} has the *same* eigenvectors $\mathbf{x}^i$ as does A, but the corresponding eigenvalues are $1/\lambda_i$. ◄

In the remainder of this section we will discuss some useful results concerning the eigenvectors and eigenvalues of certain special (though commonly occurring) square matrices. The results will be established for matrices whose elements may be complex; the corresponding properties for real matrices may be obtained as special cases.

8.13.1 Eigenvectors and eigenvalues of a normal matrix

In subsection 8.12.7 we defined a normal matrix A as one that commutes with its Hermitian conjugate, so that

$$\mathsf{A}^\dagger\mathsf{A} = \mathsf{A}\mathsf{A}^\dagger.$$

[§] In this context, when referring to linear combinations of eigenvectors x we will normally use the term ‘vector’.

We also showed that both Hermitian and unitary matrices (or symmetric and orthogonal matrices in the real case) are examples of normal matrices. We now discuss the properties of the eigenvectors and eigenvalues of a normal matrix.

If x is an eigenvector of a normal matrix A with corresponding eigenvalue λ then $\mathsf{Ax} = \lambda\mathsf{x}$, or equivalently,

$$(\mathsf{A} - \lambda\mathsf{I})\mathsf{x} = \mathsf{0}. \tag{8.69}$$

Denoting $\mathsf{B} = \mathsf{A} - \lambda\mathsf{I}$, (8.69) becomes $\mathsf{Bx} = \mathsf{0}$ and, taking the Hermitian conjugate, we also have

$$(\mathsf{Bx})^\dagger = \mathsf{x}^\dagger\mathsf{B}^\dagger = \mathsf{0}. \tag{8.70}$$

From (8.69) and (8.70) we then have

$\langle x | B^\dagger B x \rangle = 0$

$$\mathsf{x}^\dagger\mathsf{B}^\dagger\mathsf{Bx} = 0. \tag{8.71}$$

However, the product $\mathsf{B}^\dagger\mathsf{B}$ is given by

$$\mathsf{B}^\dagger\mathsf{B} = (\mathsf{A} - \lambda\mathsf{I})^\dagger(\mathsf{A} - \lambda\mathsf{I}) = (\mathsf{A}^\dagger - \lambda^*\mathsf{I})(\mathsf{A} - \lambda\mathsf{I}) = \mathsf{A}^\dagger\mathsf{A} - \lambda^*\mathsf{A} - \lambda\mathsf{A}^\dagger + \lambda\lambda^*.$$

Now since A is normal, $\mathsf{AA}^\dagger = \mathsf{A}^\dagger\mathsf{A}$ and so

$$\mathsf{B}^\dagger\mathsf{B} = \mathsf{AA}^\dagger - \lambda^*\mathsf{A} - \lambda\mathsf{A}^\dagger + \lambda\lambda^* = (\mathsf{A} - \lambda\mathsf{I})(\mathsf{A} - \lambda\mathsf{I})^\dagger = \mathsf{BB}^\dagger,$$

and hence B is also normal. From (8.71) we then find

$$\mathsf{x}^\dagger\mathsf{B}^\dagger\mathsf{Bx} = \mathsf{x}^\dagger\mathsf{BB}^\dagger\mathsf{x} = (\mathsf{B}^\dagger\mathsf{x})^\dagger\mathsf{B}^\dagger\mathsf{x} = 0,$$

$\langle B^\dagger x | B^\dagger x \rangle = 0$

from which we obtain

$\langle u | u \rangle = 0 \Rightarrow u = 0$

$$\mathsf{B}^\dagger\mathsf{x} = (\mathsf{A}^\dagger - \lambda^*\mathsf{I})\mathsf{x} = \mathsf{0}.$$

Therefore, for a normal matrix A, *the eigenvalues of* $\mathsf{A}^\dagger$ *are the complex conjugates of the eigenvalues of* A.

Let us now consider two eigenvectors x^i and x^j of a normal matrix A corresponding to two *different* eigenvalues λ_i and λ_j. We then have

$$\mathsf{Ax}^i = \lambda_i\mathsf{x}^i, \tag{8.72}$$

$$\mathsf{Ax}^j = \lambda_j\mathsf{x}^j. \tag{8.73}$$

Multiplying (8.73) on the left by $(\mathsf{x}^i)^\dagger$ we obtain

$\langle x^i | A x^j \rangle = \lambda_j \langle x^i | x^j \rangle$

$$(\mathsf{x}^i)^\dagger\mathsf{Ax}^j = \lambda_j(\mathsf{x}^i)^\dagger\mathsf{x}^j. \tag{8.74}$$

However, on the LHS of (8.74) we have

$$(\mathsf{x}^i)^\dagger\mathsf{A} = (\mathsf{A}^\dagger\mathsf{x}^i)^\dagger = (\lambda_i^*\mathsf{x}^i)^\dagger = \lambda_i(\mathsf{x}^i)^\dagger, \tag{8.75}$$

where we have used (8.40) and the property just proved for a normal matrix to write $\mathsf{A}^\dagger \mathsf{x}^i = \lambda_i^* \mathsf{x}^i$. From (8.74) and (8.75) we have

$(\lambda_i - \lambda_j)\langle x^i | x^j \rangle = 0$

$$(\lambda_i - \lambda_j)(\mathsf{x}^i)^\dagger \mathsf{x}^j = 0. \qquad (8.76)$$

Thus, *if* $\lambda_i \neq \lambda_j$ *the eigenvectors* x^i *and* x^j *must be orthogonal*, i.e. $(\mathsf{x}^i)^\dagger \mathsf{x}^j = 0$. $\langle x^i | x^j \rangle = 0$

It follows immediately from (8.76) that if all N eigenvalues of a normal matrix A are distinct then all N eigenvectors of A are mutually orthogonal. If, however, two or more eigenvalues are the same then further consideration is required. An eigenvalue corresponding to two or more different eigenvectors (i.e. they are not simply multiples of one another) is said to be *degenerate*. Suppose that λ_1 is k-fold degenerate, i.e.

$$\mathsf{A}\mathsf{x}^i = \lambda_1 \mathsf{x}^i \qquad \text{for } i = 1, 2, \ldots, k, \qquad (8.77)$$

but that it is different from any of λ_{k+1}, λ_{k+2}, etc. Then any linear combination of these x^i is also an eigenvector with eigenvalue λ_1, since, for $\mathsf{z} = \sum_{i=1}^k c_i \mathsf{x}^i$,

$$\mathsf{A}\mathsf{z} \equiv \mathsf{A}\sum_{i=1}^k c_i \mathsf{x}^i = \sum_{i=1}^k c_i \mathsf{A}\mathsf{x}^i = \sum_{i=1}^k c_i \lambda_1 \mathsf{x}^i = \lambda_1 \mathsf{z}. \qquad (8.78)$$

If the x^i defined in (8.77) are not already mutually orthogonal then we can construct new eigenvectors z^i that are orthogonal by the following procedure:

$$\begin{aligned}
\mathsf{z}^1 &= \mathsf{x}^1,\\
\mathsf{z}^2 &= \mathsf{x}^2 - \left[(\hat{\mathsf{z}}^1)^\dagger \mathsf{x}^2\right]\hat{\mathsf{z}}^1,\\
\mathsf{z}^3 &= \mathsf{x}^3 - \left[(\hat{\mathsf{z}}^2)^\dagger \mathsf{x}^3\right]\hat{\mathsf{z}}^2 - \left[(\hat{\mathsf{z}}^1)^\dagger \mathsf{x}^3\right]\hat{\mathsf{z}}^1,\\
&\vdots\\
\mathsf{z}^k &= \mathsf{x}^k - \left[(\hat{\mathsf{z}}^{k-1})^\dagger \mathsf{x}^k\right]\hat{\mathsf{z}}^{k-1} - \cdots - \left[(\hat{\mathsf{z}}^1)^\dagger \mathsf{x}^k\right]\hat{\mathsf{z}}^1.
\end{aligned}$$

In this procedure, known as *Gram–Schmidt orthogonalisation*, each new eigenvector z^i is normalised to give the unit vector $\hat{\mathsf{z}}^i$ before proceeding to the construction of the next one (the normalisation is carried out by dividing each element of the vector z^i by $[(\mathsf{z}^i)^\dagger \mathsf{z}^i]^{1/2}$). Note that each factor in brackets $(\hat{\mathsf{z}}^m)^\dagger \mathsf{x}^n$ is a scalar product and thus only a number. It follows that, as shown in (8.78), each vector z^i so constructed is an eigenvector of A with eigenvalue λ_1 and will remain so on normalisation. It is straightforward to check that, provided the previous new eigenvectors have been normalised as prescribed, each z^i is orthogonal to all its predecessors. (In practice, however, the method is laborious and the example in subsection 8.14.1 gives a less rigorous but considerably quicker way.)

Therefore, even if A has some degenerate eigenvalues we can *by construction* obtain a set of N mutually orthogonal eigenvectors. Moreover, it may be shown (although the proof is beyond the scope of this book) that these eigenvectors are *complete* in that they form a basis for the N-dimensional vector space. As a result any arbitrary vector y can be expressed as a linear combination of the eigenvectors x^i:

$$\mathsf{y} = \sum_{i=1}^{N} a_i \mathsf{x}^i, \tag{8.79}$$

where $a_i = (\mathsf{x}^i)^\dagger \mathsf{y}$. Thus, the eigenvectors form an orthogonal basis for the vector space. By normalising the eigenvectors so that $(\mathsf{x}^i)^\dagger \mathsf{x}^i = 1$ this basis is made orthonormal.

►*Show that a normal matrix* A *can be written in terms of its eigenvalues* λ_i *and orthonormal eigenvectors* x^i *as*

$$\mathsf{A} = \sum_{i=1}^{N} \lambda_i \mathsf{x}^i (\mathsf{x}^i)^\dagger. \tag{8.80}$$

The key to proving the validity of (8.80) is to show that both sides of the expression give the same result when acting on an arbitary vector y. Since A is normal, we may expand y in terms of the eigenvectors x^i, as shown in (8.79). Thus, we have

$$\mathsf{A}\mathsf{y} = \mathsf{A} \sum_{i=1}^{N} a_i \mathsf{x}^i = \sum_{i=1}^{N} a_i \lambda_i \mathsf{x}^i.$$

Alternatively, the action of the RHS of (8.80) on y is given by

$$\sum_{i=1}^{N} \lambda_i \mathsf{x}^i (\mathsf{x}^i)^\dagger \mathsf{y} = \sum_{i=1}^{N} a_i \lambda_i \mathsf{x}^i,$$

since $a_i = (\mathsf{x}^i)^\dagger \mathsf{y}$. We see that the two expressions for the action of each side of (8.80) on y are identical, which implies that this relationship is indeed correct. ◄

8.13.2 Eigenvectors and eigenvalues of Hermitian and anti-Hermitian matrices

For a normal matrix we showed that if $\mathsf{A}\mathsf{x} = \lambda\mathsf{x}$ then $\mathsf{A}^\dagger \mathsf{x} = \lambda^* \mathsf{x}$. However, if A is also Hermitian, $\mathsf{A} = \mathsf{A}^\dagger$, it follows necessarily that $\lambda = \lambda^*$. Thus, the eigenvalues of an Hermitian matrix are real, a result which may be proved directly.

►*Prove that the eigenvalues of an Hermitian matrix are real.*

For any particular eigenvector x^i, we take the Hermitian conjugate of $\mathsf{A}\mathsf{x}^i = \lambda_i \mathsf{x}^i$ to give

$$(\mathsf{x}^i)^\dagger \mathsf{A}^\dagger = \lambda_i^* (\mathsf{x}^i)^\dagger. \tag{8.81}$$

Using $\mathsf{A}^\dagger = \mathsf{A}$, since A is Hermitian, and multiplying on the right by x^i, we obtain

$\langle Ax^i | x^i \rangle = \lambda_i^* \langle x^i | x^i \rangle$

$$(\mathsf{x}^i)^\dagger \mathsf{A} \mathsf{x}^i = \lambda_i^* (\mathsf{x}^i)^\dagger \mathsf{x}^i. \tag{8.82}$$

But multiplying $\mathsf{A}\mathsf{x}^i = \lambda_i \mathsf{x}^i$ through on the left by $(\mathsf{x}^i)^\dagger$ gives

$$(\mathsf{x}^i)^\dagger \mathsf{A} \mathsf{x}^i = \lambda_i (\mathsf{x}^i)^\dagger \mathsf{x}^i.$$

$\langle x^i | A x^i \rangle = \lambda_i \langle x^i | x^i \rangle$

Subtracting this from (8.82) yields

$$0 = (\lambda_i^* - \lambda_i)(\mathsf{x}^i)^\dagger \mathsf{x}^i.$$

$(\lambda_i^* - \lambda_i) \langle x^i | x^i \rangle = 0$

But $(x^i)^\dagger x^i$ is the modulus squared of the non-zero vector x^i and is thus non-zero. Hence λ_i^* must equal λ_i and thus be real. The same argument can be used to show that the eigenvalues of a real symmetric matrix are themselves real. ◄

The importance of the above result will be apparent to any student of quantum mechanics. In quantum mechanics the eigenvalues of operators correspond to measured values of observable quantities, e.g. energy, angular momentum, parity and so on, and these clearly must be real. If we use Hermitian operators to formulate the theories of quantum mechanics, the above property guarantees physically meaningful results.

Since an Hermitian matrix is also a normal matrix, its eigenvectors are orthogonal (or can be made so using the Gram–Schmidt orthogonalisation procedure). Alternatively we can prove the orthogonality of the eigenvectors directly.

►*Prove that the eigenvectors corresponding to different eigenvalues of an Hermitian matrix are orthogonal.*

Consider two unequal eigenvalues λ_i and λ_j and their corresponding eigenvectors satisfying

$$Ax^i = \lambda_i x^i, \tag{8.83}$$
$$Ax^j = \lambda_j x^j. \tag{8.84}$$

Taking the Hermitian conjugate of (8.83) we find $(x^i)^\dagger A^\dagger = \lambda_i^*(x^i)^\dagger$. Multiplying this on the right by x^j we obtain

$$(x^i)^\dagger A^\dagger x^j = \lambda_i^*(x^i)^\dagger x^j,$$

$\langle x^i | A x^j \rangle = \langle A x^i | x^j \rangle =$

$= \lambda_i^* \langle x^i | x^j \rangle$

and similarly multiplying (8.84) through on the left by $(x^i)^\dagger$ we find

$$(x^i)^\dagger A x^j = \lambda_j (x^i)^\dagger x^j.$$

$\langle x^i | A x^j \rangle = \lambda_j \langle x^i | x^j \rangle$

Then, since $A^\dagger = A$, the two left-hand sides are equal and, because the λ_i are real, on subtraction we obtain

$$0 = (\lambda_i - \lambda_j)(x^i)^\dagger x^j.$$

$(\lambda_i - \lambda_j) \langle x^i | x^j \rangle = 0$

Finally we note that $\lambda_i \neq \lambda_j$ and so $(x^i)^\dagger x^j = 0$, i.e. the eigenvectors x^i and x^j are orthogonal. ◄

$\langle x^i | x^j \rangle = 0$

In the case where some of the eigenvalues are equal, further justification of the orthogonality of the eigenvectors is needed. The Gram–Schmidt orthogonalisation procedure discussed above provides a proof of, and a means of achieving, orthogonality. The general method has already been described and we will not repeat it here.

We may also consider the properties of the eigenvalues and eigenvectors of an anti-Hermitian matrix, for which $A^\dagger = -A$ and thus

$$AA^\dagger = A(-A) = (-A)A = A^\dagger A.$$

Therefore matrices that are anti-Hermitian are also normal and so have mutually orthogonal eigenvectors. The properties of the eigenvalues are also simply deduced, since if $Ax = \lambda x$ then

$$\lambda^* x = A^\dagger x = -Ax = -\lambda x.$$

Hence $\lambda^* = -\lambda$ and so λ must be *pure imaginary* (or *zero*). In a similar manner to that used for Hermitian matrices, these properties may be proved directly.

8.13.3 Eigenvectors and eigenvalues of a unitary matrix

A unitary matrix satisfies $\mathsf{A}^\dagger = \mathsf{A}^{-1}$ and is also a normal matrix, with mutually orthogonal eigenvectors. To investigate the eigenvalues of a unitary matrix, we note that if $\mathsf{Ax} = \lambda\mathsf{x}$ then

$$\mathsf{x}^\dagger\mathsf{x} = \mathsf{x}^\dagger\mathsf{A}^\dagger\mathsf{Ax} = \lambda^*\lambda\mathsf{x}^\dagger\mathsf{x},$$

$\langle x|x\rangle = \langle x|A^\dagger A x\rangle =$
$= \langle Ax|Ax\rangle = \lambda^*\lambda\,\langle x|x\rangle$

and we deduce that $\lambda\lambda^* = |\lambda|^2 = 1$. Thus, the eigenvalues of a unitary matrix have unit modulus.

8.13.4 Eigenvectors and eigenvalues of a general square matrix

When an $N \times N$ matrix is not normal there are no general properties of its eigenvalues and eigenvectors; in general it is not possible to find any orthogonal set of N eigenvectors or even to find *pairs* of orthogonal eigenvectors (except by chance in some cases). While the N non-orthogonal eigenvectors are usually linearly independent and hence form a basis for the N-dimensional vector space, this is not necessarily so. It may be shown (although we will not prove it) that any $N \times N$ matrix with *distinct* eigenvalues has N linearly independent eigenvectors, which therefore form a basis for the N-dimensional vector space. If a general square matrix has degenerate eigenvalues, however, then it may or may not have N linearly independent eigenvectors. A matrix whose eigenvectors are not linearly independent is said to be *defective*.

8.13.5 Simultaneous eigenvectors

We may now ask under what conditions two different normal matrices can have a common set of eigenvectors. The result – that they do so if, and only if, they commute – has profound significance for the foundations of quantum mechanics.

To prove this important result let A and B be two $N \times N$ normal matrices and x^i be the ith eigenvector of A corresponding to eigenvalue λ_i, i.e.

$$\mathsf{Ax}^i = \lambda_i\mathsf{x}^i \quad \text{for} \quad i = 1, 2, \ldots, N.$$

For the present we assume that the eigenvalues are all different.

(i) First suppose that A and B commute. Now consider $AB = BA$

$$\mathsf{ABx}^i = \mathsf{BAx}^i = \mathsf{B}\lambda_i\mathsf{x}^i = \lambda_i\mathsf{Bx}^i,$$

where we have used the commutativity for the first equality and the eigenvector property for the second. It follows that $\mathsf{A}(\mathsf{Bx}^i) = \lambda_i(\mathsf{Bx}^i)$ and thus that Bx^i is an eigenvector of A corresponding to eigenvalue λ_i. But the eigenvector solutions of $(\mathsf{A}-\lambda_i\mathsf{I})\mathsf{x}^i = \mathsf{0}$ are unique to within a scale factor, and we therefore conclude that

$$\mathsf{B}\mathsf{x}^i = \mu_i \mathsf{x}^i$$

for some scale factor μ_i. However, this is just an eigenvector equation for B and shows that x^i is an eigenvector of B, in addition to being an eigenvector of A. By reversing the roles of A and B, it also follows that every eigenvector of B is an eigenvector of A. Thus the two sets of eigenvectors are identical.

(ii) Now suppose that A and B have all their eigenvectors in common, a typical one x^i satisfying both

$$\mathsf{A}\mathsf{x}^i = \lambda_i \mathsf{x}^i \quad \text{and} \quad \mathsf{B}\mathsf{x}^i = \mu_i \mathsf{x}^i.$$

As the eigenvectors span the N-dimensional vector space, any arbitrary vector x in the space can be written as a linear combination of the eigenvectors,

$$\mathsf{x} = \sum_{i=1}^{N} c_i \mathsf{x}^i.$$

Now consider both

$$\mathsf{ABx} = \mathsf{AB}\sum_{i=1}^{N} c_i \mathsf{x}^i = \mathsf{A}\sum_{i=1}^{N} c_i \mu_i \mathsf{x}^i = \sum_{i=1}^{N} c_i \lambda_i \mu_i \mathsf{x}^i,$$

and

$$\mathsf{BAx} = \mathsf{BA}\sum_{i=1}^{N} c_i \mathsf{x}^i = \mathsf{B}\sum_{i=1}^{N} c_i \lambda_i \mathsf{x}^i = \sum_{i=1}^{N} c_i \mu_i \lambda_i \mathsf{x}^i.$$

It follows that ABx and BAx are the same for any arbitrary x and hence that

$$(\mathsf{AB}-\mathsf{BA})\mathsf{x} = \mathsf{0}$$

for all x. That is, A and B *commute*.

This completes the proof that a necessary and sufficient condition for two normal matrices to have a set of eigenvectors in common is that they commute. It should be noted that if an eigenvalue of A, say, is degenerate then not all of its possible sets of eigenvectors will also constitute a set of eigenvectors of B. However, provided that by taking linear combinations one set of joint eigenvectors can be found, the proof is still valid and the result still holds.

When extended to the case of Hermitian operators and continuous eigenfunctions (sections 17.2 and 17.3) the connection between commuting matrices and a set of common eigenvectors plays a fundamental role in the postulatory basis of quantum mechanics. It draws the distinction between commuting and non-commuting observables and sets limits on how much information about a system can be known, even in principle, at any one time.