5 Linear Algebra 3: Matrix Algebra

Topics: Eigenvalues, eigenvectors, functions of matrices, summation notation

5.1 Motivation and approach

One of the major advantages of representing operators and transformations as matrices is that this allows us to take advantage of matrix algebra in order to perform computations. Many of the most important processes in matrix algebra have been streamlined for computational implementation; indeed, in some sense computers have literally been designed to do matrix algebra efficiently.

The fundamentals of this matrix algebra are the focus of this third and final module on linear algebra for the bootcamp. We'll discuss some of the theoretical underpinnings, briefly, and also illustrate how to work important types of problems.

5.2 Eigensystems

Conceptual and technical Video: Eigenvectors and eigenvalues by 3Blue1Brown (17:15)

5.2.1 Conceptual framework of an eigensystem

In this section will only discuss operators that act on a given N-dimensional vector space, or operators that can be represented by square N-by-N matrices. In general, the action of such an operator $\hat{A}$ on a vector $|v\rangle$ will yield a new vector $|w\rangle$ that is rotated relative to $|v\rangle$. However, operators and their representative matrices often have a set of vectors called **eigenvectors** for which the operator only acts to rescale the vector. If k is an eigenvector of $\hat{A}$, we have

$$\hat{A}\,|k\rangle = \lambda_k\,|k\rangle\,, \tag{5.1}$$

for some scalar λ_k, which we call the **eigenvalue** of $\hat{A}$ corresponding to the vector $|k\rangle$. Together, we refer to the eigenvalue-eigenvector pairs as the **eigensystem** of the matrix. Finding the eigensystem of an operator is an incredibly important operation in both linear algebra and quantum mechanics.

It is a fact, that we will not prove, that an Hermitian N-by-N matrix will have N eigenvector/eigenvalue pairs. These eigenvalues will be exclusively real, and the eigenvectors for different eigenvalues must be orthogonal. (You will prove these latter two statements in your quantum coursework, and you will employ all three of them repeatedly.)

5.2.2 Solving for an eigensystem

Technical video: Example solving for eigenvalues of a 2x2 matrix by Khan Academy (5:38)
Technical video: Finding eigenvectors and eigenspaces example by Khan Academy (14:33)
Technical video: 3 x 3 Determinant by Khan Academy (10:01)

We can follow a general procedure to solve for the eigensystem of a matrix according to Eq. (5.1). We can represent our operator and vector as a matrix and column vectors following the procedures above, such that we're ultimately interested in solving the matrix eigenvalue problem

$$\mathbf{A}\mathbf{k} = \lambda_k \mathbf{k} = \lambda_k \mathbb{1}_3 \mathbf{k}, \tag{5.2}$$

where $\mathbb{1}_3$ is the 3-by-3 identity matrix. This is equivalent to finding the eigenvalue λ_k and corresponding eigenvector $\mathbf{k}$ that satisfies

$$\left(\mathbf{A} - \lambda_k \mathbb{1}_N\right)\mathbf{k} = 0. \tag{5.3}$$

This equation is satisfied by values of λ_k that satisfy the **characteristic equation** of the matrix $\mathbf{A}$, which is given by

$$\det\left(\mathbf{A} - \lambda_k \mathbb{1}_3\right) = 0. \tag{5.4}$$

Once we solve for the values of λ_k that satisfy the characteristic equation, we can use them to find the eigenvectors they correspond to.

We won't belabor determinant theory here. We write the determinant of a matrix using vertical lines, and 2-by-2 and 3-by-3 matrices are obtained according to the following equations:

$$\det(\mathbf{M}_2) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc \tag{5.5}$$

$$\det(\mathbf{M}_3) = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + cdh - ceg - bdi - afh \tag{5.6}$$

There is a systematic procedure for obtaining higher-order determinants, but we do not include it here.

Let's illustrate this by solving for the eigensystem of the following 3-by-3 matrix:

$$\mathbf{A} = \begin{bmatrix} 5 & 2i & 0 \\ -2i & 1 & 0 \\ 0 & 0 & -3 \end{bmatrix}. \tag{5.7}$$

We'll break the process of solving the eigensystem into a few steps:

1. Determine the characteristic equation of the matrix $\mathbf{A}$. We do this using Eq. (5.4),

which gives

$$\begin{aligned}
\det(\mathbf{A} - \lambda_k \mathbb{1}_3) &= \det\left(\begin{bmatrix} 5 & 2i & 0 \\ -2i & 1 & 0 \\ 0 & 0 & -3 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right) \\
&= \begin{vmatrix} 5-\lambda_k & 2i & 0 \\ -2i & 1-\lambda_k & 0 \\ 0 & 0 & -3-\lambda_k \end{vmatrix} \\
&= (5-\lambda_k)(1-\lambda_k)(-3-\lambda_k) - (2i)(-2i)(-3-\lambda_k) \\
&= (-3-\lambda_k)\left(\lambda_k^2 - 6\lambda_k + 1\right)
\end{aligned} \tag{5.8}$$

2. Find the roots of the characteristic equation. The roots of the equation are the values of λ_k that make the equation equal to zero. We can further factorize the equation determined above to write

$$(-3-\lambda_k)(\lambda_k - 2)(\lambda_k + 3) = 0. \tag{5.9}$$

In this form, it is clear that the roots of the equation are

$$\lambda_k = 2, -3, -3. \tag{5.10}$$

Make note of the fact that the value of -3 appears twice. We call the redundancy of this eigenvalue **degeneracy**, saying things like "the eigenvalue -3 is two-fold degenerate."

3. Solve for the eigenvectors for each eigenvalue. We do this by plugging each of the eigenvalues into the characteristic equation in turn, labeling the basis vectors $|e_1\rangle$, $|e_2\rangle$, and $|e_3\rangle$ and labeling the eigenvectors $|v_1\rangle$, $|v_2\rangle$, and $|v_3\rangle$.

$\lambda_k = 2$

The eigenvalue equation for this eigenvalue is

$$\mathbf{A}\,|e_1\rangle = 2\,|e_1\rangle \rightarrow (\mathbf{A} - 2)\,|e_1\rangle = 0. \tag{5.11}$$

In expanded form, this is written

$$\begin{aligned}
\begin{bmatrix} 5-2 & 2i & 0 \\ -2i & 1-2 & 0 \\ 0 & 0 & -3-2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \\
\begin{bmatrix} 3 & 2i & 0 \\ -2i & -1 & 0 \\ 0 & 0 & -5 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.
\end{aligned} \tag{5.12}$$

If we perform the matrix multiplication defined from this equation, this results in the following system of equations

$$\begin{cases} 3c_1 + 2ic_2 &= 0 \\ -2ic_1 - c_2 &= 0 \\ -5c_3 &= 0. \end{cases} \tag{5.13}$$

From these equations, we must have $c_3 = 0$. We can obtain the other two coefficients by solving

$$\begin{aligned} 3c_1 + 2ic_2 &= -2ic_1 - c_2 \\ (3+2i)\,c_1 &= (-1-2i)\,c_2 \\ c_1 &= -\frac{1+2i}{3+2i}c_2 \\ &= -\frac{7+4i}{13}c_2. \end{aligned} \tag{5.14}$$

This equation defines the proportionality of c_1 and c_2. Their exact values are arbitrary, and we will choose them according to the requirement that the resulting eigenvector is normalized. Assuming our initial basis is orthonormal, the normalization condition is

$$\begin{aligned} |c_1|^2 + |c_2|^2 &= 1 \\ \left|-\frac{7+4i}{13}c_2\right|^2 + |c_2|^2 &= 1 \\ \left(\frac{65}{169}+1\right)c_2^2 &= 1 \\ c_2 &= \pm\frac{13}{3\sqrt{26}} \end{aligned} \tag{5.15}$$

This means that we have

$$c_1 = \mp\frac{7+4i}{3\sqrt{26}}. \tag{5.16}$$

After all of that (I know it was a lot), we have determined that the normalized eigenvector that has eigenvalue 2 is

$$|v_1\rangle = \mp\frac{7+4i}{3\sqrt{26}}\,|e_1\rangle \pm \frac{13}{3\sqrt{26}}\,|e_2\rangle + 0\,|e_3\rangle\,. \tag{5.17}$$

The absolute sign of the coefficients is undefined by the matrix algebra. In physics, this absolute sign is referred to as a "phase," and it is normally the case that phase is arbitrary, physically.

$\lambda_k = -3$

Repeating this procedure for the two-fold degenerate eigenvalue $\lambda_k = 3$, we obtain the following matrix equation:

$$\begin{bmatrix} 8 & 2i & 0 \\ -2i & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{5.18}$$

The matrix multiplication for this system gives

$$\begin{cases} 8c_1 + 2ic_2 & = 0 \\ -2ic_1 + 4c_2 & = 0 \\ 0 & = 0 \end{cases} \tag{5.19}$$

It may seem odd that we wrote the final equation this way, but this is what the matrix algebra gives us, and it indicates something important: the vector $(c_1, c_2, c_3) = (0, 0, c_3)$ will satisfy this equation for *any* finite value of c_3. If we want the vector to be normalized, then we'll choose $c_3 = 1$. Hence, one of the eigenvectors with eigenvalue -3 is simply

$$|v_3\rangle = 0\,|e_1\rangle + 0\,|e_2\rangle + 1\,|e_3\rangle\,. \tag{5.20}$$

We can then solve for the other eigenvector, where we have

$$\begin{aligned} 8c_1 + 2ic_2 &= -2ic_1 + 4c_2 \\ (8+2i)\,c_1 &= (4-2i)\,c_2 \\ c_1 &= \frac{4-2i}{8+2i}c_2 \\ &= \frac{7-6i}{17}c_2 \end{aligned} \tag{5.21}$$

Then the normalization condition gives

$$\begin{aligned} \left|\frac{7-6i}{17}\right|^2 c_2^2 + c_2^2 &= 1 \\ \left(\frac{85}{289}+1\right)c_2^2 &= 1 \\ c_2 &= \pm\frac{289}{\sqrt{374}}. \end{aligned} \tag{5.22}$$

Hence, the other eigenvector with eigenvalue -3 is

$$|v_2\rangle = \pm\frac{119-102i}{\sqrt{374}}\,|e_1\rangle \pm \frac{289}{\sqrt{374}}\,|e_2\rangle + 0\,|e_3\rangle \tag{5.23}$$

The Final Results:

The eigensystem of **A** is as follows:

Eigenvalue	Eigenvector(s)
2	$\|v_1\rangle = \mp\frac{7+4i}{3\sqrt{26}}\,\|e_1\rangle \pm \frac{13}{3\sqrt{26}}\,\|e_2\rangle$
-3	$\|v_2\rangle = \pm\frac{119-102i}{\sqrt{374}}\,\|e_1\rangle \pm \frac{289}{\sqrt{374}}\,\|e_2\rangle$ $\|v_3\rangle = \|e_3\rangle$

Because the matrix **A** that we started with was Hermitian, these eigenvectors must be mutually orthogonal. The final orthogonal must be orthogonal to the first two because the basis vectors form an orthogonal set. It can be shown that the first two are orthogonal. (If they aren't, then I made a mistake in the diagonalization process.)

5.2.3 Expansion in a basis of eigenvectors

Solving for the eigensystem of an operator or matrix is sometimes referred to as **diagonalizing** the operator or matrix because an operator is diagonal in the basis of its eigenvectors. Suppose we have a 3-by-3 matrix $\mathbf{A}$ that has 3 eigenvectors with eigenvalues λ_1, λ_2, and λ_3 with corresponding eigenvectors $|\lambda_1\rangle$, $|\lambda_2\rangle$, and $|\lambda_3\rangle$, respectively. In the basis of its eigenvectors, we know we can write the matrix representation of $\mathbf{A}$ as

$$\mathbf{A} = \begin{bmatrix} | & | & | \\ \hat{A}\,|\lambda_1\rangle & \hat{A}\,|\lambda_2\rangle & \hat{A}\,|\lambda_3\rangle \\ | & | & | \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}, \tag{5.24}$$

where the second equality follows from the definition of an eigenvector, given in Eq. (5.1). Hence, we see the sense in which solving for the eigensystem diagonalizes the operator $\hat{A}$.

The logic of this goes both ways. If you are given (or determine) the matrix representation of some operator, and you see that it is a diagonal matrix, then the eigenvalues are the diagonal elements of the matrix and the eigenvectors are the basis vectors that define the matrix representation.

We can also make powerful statements about matrices that are **block diagonal**. Specifically, the eigenvalues (eigenvectors) of a block diagonal matrix can be obtained as the eigenvalues (eigenvectors) of each of the blocks individually.

In the case worked above, notice that the matrix takes the form of a 2-by-2 block followed by a 1-by-1 block. This form means that we can diagonalize each block individually. In particular, the diagonalizing the 1-by-1 block is trivial, and it has the eigenvalue -3 and the eigenvector $|v_3\rangle$, which, again, is what we saw above.

5.3 Functions of matrices

The function of an operator (or matrix) is defined in terms of the Taylor series expansion of the function:

$$f(\mathbf{A}) = \sum_{n=0}^{\infty} f^{(n)}(0)\frac{\mathbf{A}^n}{n!}. \tag{5.25}$$

The power of a matrix $\mathbf{A}$ is not a trivial thing to compute by hand unless the matrix $\mathbf{A}$ is a diagonal matrix. Powers of a diagonal matrix

$$\mathbf{D} = \begin{bmatrix} D_1 & 0 & 0 \\ 0 & D_2 & 0 \\ 0 & 0 & D_3 \end{bmatrix} \tag{5.26}$$

are given by the power of the elements of the matrix, *viz.*

$$\mathbf{D}^n = \begin{bmatrix} (D_1)^n & 0 & 0 \\ 0 & (D_2)^n & 0 \\ 0 & 0 & (D_3)^n \end{bmatrix}. \tag{5.27}$$

You can convince yourself of this by evaluating the case of a smaller matrix, like $\mathbf{D}^2$ for a 3-by-3 matrix, or by rigorously proving it with methods we discuss below.

In any case, the importance of diagonal representations of matrices highlights the utility of the eigenbasis of an operator, because we have shown above that an operator is diagonal in its eigenbasis.

We can obtain the same result a different way by expanding a vector of interest $|\psi\rangle$ in the eigenbasis of the operator $\mathbf{A}$ as we have done for other bases previously. In the general case of an N-dimensional vector space, we can write

$$|\psi\rangle = \sum_k \lambda_k |k\rangle , \tag{5.28}$$

where the kets $\{|k\rangle\}$ are eigenvectors of the operator $\hat{A}$ with eigenvalues $\{\lambda_k\}$. Then the action of $\hat{A}$ on $|\psi\rangle$ is given by

$$\hat{A}^n |\psi\rangle = \sum_k c_k \hat{A}^n |k\rangle . \tag{5.29}$$

Because $|k\rangle$ is an eigenvector of $\hat{A}$,

$$\hat{A}^n |k\rangle = \lambda_k \hat{A}^{n-1} |k\rangle = \lambda_k^2 \hat{A}^{n-2} |k\rangle = \cdots = \lambda_k^n |k\rangle . \tag{5.30}$$

Hence, we have

$$\hat{A}^n |\psi\rangle = \sum_k \lambda_k^n |k\rangle . \tag{5.31}$$

In terms of evaluating a function of an operator, then, we have

$$\begin{aligned} f(\mathbf{A}) |\psi\rangle &= \sum_k^N c_k \sum_{n=0}^{\infty} f^{(n)}(0) \frac{1}{n!} \mathbf{A}^n |k\rangle \\ &= \sum_k^N c_k \left(\sum_{n=0}^{\infty} f^{(n)}(0) \frac{1}{n!} \lambda_k^n \right) |k\rangle \\ &= \sum_k^N c_k f(\lambda_k) |k\rangle . \end{aligned} \tag{5.32}$$

This is a compact expression that we can evaluate for all sorts of functions of operators, like $\exp(\mathbf{A})$, $\mathbf{A}^{-1}$, and others. This expression *only* holds for the eigenbasis of $\hat{A}$. Because of this, **it is incredibly common that you will need to change to the eigenbasis of the operator of interest in order to make any progress.** This is particularly true for problems involving time evolution (how a system changes with time), where the operator of interest is the Hamiltonian.

5.4 Summation notation

Our final topic in linear algebra is a little bit more technical, and it develops the tools necessary to do general proofs in linear algebra. We have, by and large, steered clear of proofs in the bootcamp curriculum, choosing to focus our time on underlying concepts and the mechanics of problem solving. Still, you will need to use linear algebra prove things in your quantum mechanics coursework, and summation notation is the way to prove things in the general case in linear algebra.

This section takes the algorithms we commonly use to compute matrix/vector operations and formalizes them into **summation notation.** It can be tricky to get used to at first, and I recommend working out each of the operations below for small (2-dimensions is fine!) matrices and vectors to get the hang of what the expressions are saying.

The Kronecker delta. The Kronecker delta (δ_{ij}) is defined according to

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} . \tag{5.33}$$

This term comes up a lot when we are working with diagonal matrices, where we might write something like

$$D_{ij} = d_i \delta_{ij}, \tag{5.34}$$

to indicate that the element at the ith row and the jth column of the diagonal matrix $\mathbf{D}$ is zero unless the row number and the column number are the same (hence we only need one index, i in this case, to specify the value of the elements $\{d_i\}$.

It also comes up when we work with in an orthonormal basis, which we do often. In fact, it is actually common to write the condition of orthonormality with a Kronecker delta. If a set of basis vectors $\{|\phi_i\rangle\}$ is orthonormal, then the inner product

$$\langle \phi_i | \phi_j \rangle = \delta_{ij} \tag{5.35}$$

must hold for all possible values of i and j in the basis. In essence, this equation says that the value of the inner product is 1 if the two basis vectors in the inner product are the same (the basis is normalized) and that the value of the inner product is 0 if the two basis vectors are different (all pairs of vectors in the basis are orthogonal).

Mechanically, we can use the Kronecker delta to simplify and eliminate various sums. As a contrived example, suppose we have the sum over two indices i and j given below:

$$\sum_{ij} a_i b_j \delta_{ij}. \tag{5.36}$$

Because of the definition of the Kronecker delta, we will multiply the expression being summed by zero for all of the terms where $i \neq j$. The only terms that "survive" the Kronecker delta are those where $i = j$, and we can simplify the sum as

$$\sum_{ij} a_i b_j \delta_{ij} = \sum_i a_i b_i. \tag{5.37}$$

The upshot is this: when we sum over both indices of a Kronecker delta, we can replace the sum over two indices with a sum over one index, replacing the label of one of the indices with the other.

Inner product. The inner product of two N-dimensional vectors $\mathbf{v}$ and $\mathbf{w}$ is given by

$$\langle v|w\rangle = \sum_i^N v_i^* w_i. \tag{5.38}$$

Matrix–vector multiplication. For the matrix equation

$$\mathbf{Av} = \mathbf{w},$$

where $\mathbf{A}$ is a matrix with elements A_{ij} at the ith row and the jth column and $\mathbf{v}$ and $\mathbf{w}$ are vectors with elements v_i and w_i, the ith element of $\mathbf{w}$ can be obtained as

$$w_i = \sum_j A_{ij} v_j. \tag{5.39}$$

Matrix–matrix multiplication. Likewise, if we want to multiply two matrices $\mathbf{A}$ and $\mathbf{B}$ together to obtain a third matrix $\mathbf{C}$ according to

$$\mathbf{AB} = \mathbf{C}, \tag{5.40}$$

we can obtain each of the elements C_{ij} as

$$C_{ij} = \sum_k A_{ik} B_{kj}. \tag{5.41}$$

It can be difficult to remember which index should be summed over while you are getting used to this notation. It can be helpful to remember that indices that are adjacent to each other in the expression are usually summed over, while those on the outside of expressions are not.

Matrix trace. The trace of an N-by-N square matrix $\mathbf{A}$ is the sum of its diagonal elements, which we write as

$$\mathrm{Tr}(\mathbf{A}) = \sum_i A_{ii}. \tag{5.42}$$

Example 5.1: Prove that the kth power of an N-by-N diagonal matrix $\mathbf{D}$ can be obtained by raising each of the elements of $\mathbf{D}$ to the kth power.

5.5 Connections to physical chemistry

Eigensystems are relevant to physical chemistry and quantum mechanics because in the latter we postulate that the measurement of some observable will always take on the value of one of the eigenvalues of operator that corresponds to this observable. Furthermore, we postulate that the state of the system after measurement corresponds to the eigenvector/eigenspace associated with the eigenvalue that was measured. In this way, eigensystems show up any time we are interested in the properties of a quantum mechanical system—which is all the time! We will also commonly transform a vector into the eigenbasis of a particular operator in order to make our mathematical manipulations more straightforward.

Functions of matrices become an important topic when we start to use the Schrödinger equation to consider the time evolution of a quantum mechanical system. We will need to exponentiate the Hamiltonian matrix in order to consider this, and this is done using the Taylor series expansion that was described in Section. 5.3 above.

As indicated in the notes, summation notation is used to prove things in linear algebra, and by extension in quantum mechanics as well as statistical mechanics, although to a more limited extent. Facility with these concepts will enable you to better understand and work through important derivations in your coursework and in research.

5.6 Example problem solutions

Example 5.1: *Prove that the kth power of an N-by-N diagonal matrix $\mathbf{D}$ can be obtained by raising each of the elements of $\mathbf{D}$ to the kth power.*

We start with the example where $k = 2$ and then generalize from there. The ij'th element of the square of the matrix $\mathbf{D}$ is given by

$$(\mathbf{D}^2)_{ij} = \sum_k D_{ik} D_{kj}. \tag{5.43}$$

Because $\mathbf{D}$ is a diagonal matrix, we must have

$$D_{ij} = \delta_{ij} D_{ii}. \tag{5.44}$$

Hence, we can simplify the preceding expression to

$$(\mathbf{D}^2)_{ij} = \sum_k \delta_{ij}\delta_{kj} D_{ii} D_{kk} = D_{ii} D_{ii} = D_{ii}^2, \tag{5.45}$$

because the two Kroenecker delta's make every time in the sum where $k \neq i$ vanish. Based on induction, we can see that this result generalizes to any integral value of k. To see this, we can write

$$\begin{aligned}
(\mathbf{D}^3)_{ij} &= (\mathbf{D}^2 \cdot \mathbf{D})_{ij} \\
&= \sum_k (\mathbf{D}^2)_{ik} D_{kj} \\
&= \sum_k D_{ii}^2 D_{kj} \\
&= \sum_k D_{ii}^2 \delta_{kj} D_{kk} \\
&= D_{ii}^3
\end{aligned} \tag{5.46}$$

In words, we have shown that the kth power of an N-by-N diagonal matrix can be obtained by raising each of the elements of the matrix to the kth power for $k = 2$, and that the fact that this is true for k implies that it is true for $k + 1$. Hence, it must be true for all $k > 2$.