St. Francis Xavier University
Department of Computer Science

CSCI 554: Matrix Computation
Lecture 8: Eigenvalues and Eigenvectors I
Winter 2022

1 Everything Eigen

Eigenvalues and *eigenvectors* are fundamental notions in not just matrix algebra and linear algebra, but mathematics in general. Although we will be studying eigenvalues and eigenvectors through our usual lens of matrix systems, these notions appear in many other areas. For instance, eigenvalues and eigenvectors play key roles in handling systems of differential equations. While we won't delve into the details of differential equations, we will briefly review them here in order to motivate our study of eigenvalues and eigenvectors.

We can model a system of n differential equations using a matrix of dimension $n \times n$ together with some vectors to obtain a linear system of the form $\dot{x} = Ax - b$. We can take this system and drop the term b to obtain the *homogeneous system* $\dot{x} = Ax$, where A is of dimension $n \times n$. This homogeneous system acts as a shorthand notation for the complete linear system

$$\begin{aligned}
\frac{\mathrm{d}x_1}{\mathrm{d}t} &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\
\frac{\mathrm{d}x_2}{\mathrm{d}t} &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\
&\vdots \\
\frac{\mathrm{d}x_n}{\mathrm{d}t} &= a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n
\end{aligned}$$

In order to solve linear differential equations like the ones found in this system, we often want to seek simple solutions of the form $x(t) = g(t)v$, where $g(t)$ is a nonzero scalar function of t and v is a nonzero constant vector. Taking these two components together, we get the "vector" part of $x(t)$ from v and the "differential" part of $x(t)$ from $g(t)$. Substituting this expression into our homogeneous system, we get the equation $\dot{g}(t)v = g(t)Av$, which can also be written as

$$\frac{\dot{g}(t)}{g(t)}v = Av.$$

Now, observe that both v and Av are constant vectors, so the value of $\dot{g}(t)/g(t)$ must also be constant. Let's denote this value by λ. Substituting λ into our earlier equation, we get

$$Av = \lambda v.$$

Now, we have drawn our connection to the topic at hand. Any nonzero vector v for which there exists a value λ such that $Av = \lambda v$ is called an *eigenvector* of A, and the value λ is called the *eigenvalue* of A associated with v.

Let's connect this back to our linear system. We can see that if $x(t)$ is a solution of $\dot{x} = Ax$, then v must be an eigenvector of A. Moreover, $g(t)$ must satisfy the equation $\dot{g}(t)/g(t) = \lambda$, where λ is the eigenvalue of A associated with v. Thus, with a little bit of work, we can show that if A has enough eigenvectors, then every solution of $\dot{x} = Ax$ can be expressed as a linear combination of eigenvectors and associated eigenvalues!

Let's now move away from differential equations and focus on the main topic of this lecture. Unlike in earlier lectures, where we assumed all of the matrices we were working with were real-valued, here we will assume that our matrices are complex-valued. Thus, we will consider matrices $A \in \mathbb{C}^{n \times n}$.

We have seen through our motivating example what an *eigenvalue* and an *eigenvector* are, so let's formalize these definitions.

Definition 1 (Eigenvector and eigenvalue)**.** Let $A \in \mathbb{C}^{n\times n}$. A vector $v \in \mathbb{C}^n$ is called an eigenvector of A if $v \neq 0$ and there exists a value $\lambda \in \mathbb{C}$ such that $Av = \lambda v$. The scalar value λ is called the eigenvalue of A associated with v.

Occasionally, we refer to the pair (v, λ) as an *eigenpair*.

Although every eigenvector is associated with a unique eigenvalue, the converse is not necessarily true. A single eigenvalue may be associated with potentially many eigenvectors. We can see this, for instance, by taking any nonzero multiple of an eigenvector v; all such multiples will have the same eigenvalue. The set of all eigenvalues of a matrix A is called the *spectrum* of A.

Remark. Following our observation that any nonzero multiple of an eigenvector v has the same eigenvalue, we can see that it's easy for us to ensure some eigenvector we're working with has unit norm if we need this property. We simply need to compute the eigenvector $\widehat{v} = (1/\|v\|)v$. Then, $\widehat{v}$ is still an eigenvector of A with the same eigenvalue, and $\|\widehat{v}\| = 1$.

Recall that the identity matrix acts exactly like an identity element: multiplying something by the identity matrix does not affect that thing. Thus, it should make no difference if we were to write $Av = \lambda Iv$, where I is the $n \times n$ identity matrix. However, observe that in doing so, we can rearrange this expression to obtain

$$(\lambda I - A)v = 0.$$

What this expression tells us is that the matrix $(\lambda I - A)$ transforms the vector v to the zero vector. Using our linear algebra knowledge, we then know that the matrix $(\lambda I - A)$ cannot have an inverse. As a result, the *determinant* of the matrix, $\det(\lambda I - A)$, must be equal to zero.

This observation gives us an alternative way to characterize eigenvalues of A in terms of the so-called *characteristic equation* of A.

Proposition 2. *A value λ is an eigenvalue of a matrix A if and only if*

$$\det(\lambda I - A) = 0.$$

We can furthermore verify that $\det(\lambda I - A)$ is a polynomial of degree n in λ, and we refer to this as the *characteristic polynomial*.

Note that this is the first time we've dealt with determinants in this course. This is not to suggest that determinants are useless; in fact, they have great theoretical use when we're discussing matrices and their properties. Determinants simply aren't as commonly used in actual computations, and since this course places such an emphasis on computation, we haven't had a need for determinants until now.

In the following example, we assume a familiarity with determinants. If you aren't familiar with this notion, then you can consult any textbook on linear algebra or use computer software to calculate the determinant of a given matrix.

Example 3. Consider the matrix

$$A = \begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix}.$$

We will find the eigenvalues of A by solving the characteristic equation $\det(\lambda I - A) = 0$.

Plugging in I and A to the characteristic equation, we get

$$\det\left(\lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix}\right) = 0, \text{ or}$$
$$\det\left(\begin{bmatrix} \lambda + 5 & -2 \\ 7 & \lambda - 4 \end{bmatrix}\right) = 0.$$

Computing the determinant of this matrix, we obtain the equation $(\lambda+5)(\lambda-4)-(-2)(7) = \lambda^2+\lambda-6 = 0$. We then solve this equation to get the eigenvalues $\lambda_1 = 2$ and $\lambda_2 = -3$.

1.1 Characterizing Eigenvalues of a Matrix

Recall that, when we defined eigenvalues and eigenvectors, we chose to work in the complex number space instead of the real number space. Having discussed characteristic polynomials, we can now reason about why we made such a choice.

It is a well-known fact that a polynomial equation of degree n has n complex roots, and this implies that the associated matrix A has n eigenvalues (some of which may not be unique). If we took A to be a real-valued matrix, then its characteristic polynomial would have real-valued coefficients, but the zeroes of the polynomial are not guaranteed to be real-valued. Thus, a real-valued matrix may, in some cases, have *complex eigenvalues*.

Let's now flip things around and consider what happens if we take a real eigenvalue of a real-valued matrix. Suppose we have a matrix $A \in \mathbb{R}^{n\times n}$ with a real eigenvalue λ, and consider the equation $(\lambda I - A)v = 0$. In this case, the coefficient matrix $(\lambda I - A)$ is also real-valued, and the fact that $(\lambda I - A)$ is not invertible implies that this equation has real-valued solutions. Thus, for every *real eigenvalue* of a real-valued matrix, there exists an associated real-valued eigenvector.

Example 4. Recall the matrix A from our previous example. We found that A had two eigenvalues: $\lambda_1 = 2$ and $\lambda_2 = -3$. Let's now find the eigenvectors associated with these eigenvalues.

Starting with $\lambda_1 = 2$, we wish to find all vectors $v \neq 0$ such that $Av = 2v$. We can do so by finding the solutions to the equation

$$\left(2\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} - \begin{bmatrix}-5 & 2\\ -7 & 4\end{bmatrix}\right)\begin{bmatrix}x_1\\ x_2\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix}, \text{ or}$$
$$\begin{bmatrix}7 & -2\\ 7 & -2\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix}.$$

If we convert this matrix to its augmented form and perform row reduction, we obtain the matrix

$$\left[\begin{array}{cc|c}1 & -\frac{2}{7} & 0\\ 0 & 0 & 0\end{array}\right],$$

and the solution to the corresponding linear system is a vector of the form $s\begin{bmatrix}2/7\\ 1\end{bmatrix}$.

We can multiply this solution by 7 to obtain the nicer vector $s'\begin{bmatrix}2\\ 7\end{bmatrix}$, and this leads us to conclude that the eigenvector corresponding to $\lambda_1 = 2$ is $v_1 = \begin{bmatrix}2\\ 7\end{bmatrix}$. We can verify this simply by checking that $Av_1 = 2v_1$.

Following a similar process for the eigenvalue $\lambda_2 = -3$, we find that the associated eigenvector is $v_2 = \begin{bmatrix}1\\ 1\end{bmatrix}$.

Once we've computed the eigenvalues and eigenvectors of a matrix, we can use the following well-known result from linear algebra to establish an important property of these eigenvectors.

Proposition 5. *Let $v_1, v_2, \ldots, v_k$ be the eigenvectors associated with the distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$ of a matrix A. Then each of v_1 through v_k are linearly independent.*

The number of linearly independent eigenvectors of a matrix A provides a characterization of the matrix. If A is of dimension $n \times n$ and A has n linearly independent eigenvectors, then we say that A is *semisimple* or *nondefective*. Naturally, if A has fewer than n linearly independent eigenvectors, then it is *defective*.

Note that, while a defective matrix is guaranteed to have at least one repeated eigenvalue, the existence of repeated eigenvalues alone is not sufficient to guarantee that a matrix is defective. This is a consequence of our earlier observation that single eigenvalues may be associated with potentially many eigenvectors.

Example 6. Consider the matrix

$$A = \begin{bmatrix} 1 & -3 \\ 3 & 7 \end{bmatrix}.$$

Solving the characteristic equation of A gives us two eigenvalues: $\lambda_1 = 4$ and $\lambda_2 = 4$. Note that $\lambda_1 = \lambda_2$, so A has one repeated eigenvalue.

Using these eigenvalues, we find that A has one eigenvector: $v = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. Therefore, A is a defective matrix.

1.2 Eigenvalues of Special Matrices

In some cases, we don't need to go through the entire process to calculate the eigenvalues of a matrix A. If A is a special matrix, then we can sometimes get this information directly from A itself.

We begin with one of our familiar special matrix classes: the triangular matrices.

Theorem 7. *Let $A \in \mathbb{C}^{n \times n}$ be a triangular matrix. Then the eigenvalues of A are exactly the entries $a_{11}, a_{22}, \ldots, a_{nn}$ along the main diagonal of A.*

Proof. It is straightforward to show that the determinant of the matrix A (and, indeed, of any triangular matrix) is $\det(A) = a_{11}a_{22}\ldots a_{nn}$. Since A is a triangular matrix, so too is $\lambda I - A$. Therefore, the characteristic equation of A is

$$\begin{aligned} 0 &= \det(\lambda I - A) \\ &= (\lambda - a_{11})(\lambda - a_{22})\ldots(\lambda - a_{nn}), \end{aligned}$$

and the roots of this equation are the diagonal entries a_{11} through a_{nn} of A. □

Next, we introduce a class of special matrices that we haven't studied previously: so-called *similar matrices*. We say that two $n \times n$ matrices A and B are similar if there exists an invertible matrix P such that $A = P^{-1}BP$.

With this notion of similar matrices, if we know the eigenvalues of one of the two matrices A and B, and if we know that these two matrices are similar, then we can automatically obtain the eigenvalues of the other matrix.

Theorem 8. *Let A and B be similar matrices, each of dimension $n \times n$; that is, $A = P^{-1}BP$ for some invertible matrix P. Then A and B have the same eigenvalues.*

Proof. Suppose that $A = P^{-1}BP$ and λ is an eigenvalue of A; that is, $Av = \lambda v$ for some vector $v \neq 0$. Then we have that $P^{-1}BPv = \lambda v$, and so $BPv = \lambda Pv$. Since $v \neq 0$, it follows that $Pv \neq 0$, and thus Pv acts as an eigenvector of B. Therefore, λ must also be an eigenvalue of B. An analogous argument applies in the other direction. □

2 The Power Method

Now that we're familiar with eigenvalues and eigenvectors, how can we find the eigenvalues of a given matrix $A \in \mathbb{C}^{n \times n}$? Interestingly, in some applications, we don't need to know every eigenvalue of A; it suffices for us to know only the largest eigenvalue. Thus, our first method for computing eigenvalues will give us a way to obtain the largest eigenvalue from a given matrix A.

To make things simple, let's assume that A is a semisimple matrix, so that it has n linearly independent eigenvectors. Let's further suppose that the n eigenvalues of A are ordered by magnitude, so that $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$. If $|\lambda_1| > |\lambda_2|$, then we say that λ_1 is the *dominant eigenvalue* of A, and the associated eigenvector v_1 is the *dominant eigenvector* of A.

We can compute the dominant eigenvalue of A using a technique known as the *power method*. As the name suggests, this method involves us taking repeated powers; namely, we take powers of the matrix A multiplied by an arbitrary vector v. In doing so, we compute a sequence

$$v,\ Av,\ A^2v,\ A^3v,\ \ldots$$

and each term of this sequence is easily obtained by multiplying the previous term by A directly. This iterative method eventually converges to a dominant eigenvector of A, no matter what initial vector v we choose (with very limited exceptions).

Why does the power method work for (almost) any initial vector? If we assume that there exists a dominant eigenvalue of A, and if the initial vector we choose has some nonzero component pointing in the same direction as a dominant eigenvector of A, then the repeated multiplication by A "magnifies" that dominant component and results in the vector approximating the dominant eigenvector more closely on each iteration.

More precisely, we know that the eigenvectors v_1 through v_n form a basis for $\mathbb{C}^n$, and so we can write our initial vector v as

$$v = c_1v_1 + c_2v_2 + \cdots + c_nv_n,$$

where c_1 through c_n are some constants. Since we don't know v_1 through v_n in advance, we also don't know the values of these constants. However, for (almost) any choice of v, the first constant c_1 will be nonzero. Then, multiplying by A, we obtain

$$\begin{aligned} Av &= c_1Av_1 + c_2Av_2 + \cdots + c_nAv_n \\ &= c_1\lambda_1v_1 + c_2\lambda_2v_2 + \cdots + c_n\lambda_nv_n \end{aligned}$$

and, on the ith iteration, we have

$$A^iv = c_1\lambda_1^iv_1 + c_2\lambda_2^iv_2 + \cdots + c_n\lambda_n^iv_n.$$

By our assumption that λ_1 was the dominant eigenvalue, the component $c_1\lambda_1^iv_1$ dominates the sum as the number of iterations increases.

Since this term grows to dominate the sum, we must be careful to avoid issues such as integer overflow. Earlier, we observed that every multiple of an eigenvector is itself an eigenvector, so in principle, we should care only about the direction of an eigenvector and not the magnitude. As a consequence, we will multiply our vector by some scaling factor on each iteration to keep its magnitude reasonable. While we can use any scaling factor we want, it's easiest for us to take the scaling factor to be the largest component of the vector on that iteration. Indeed, doing so is doubly beneficial for us: not only does this allow us to maintain a reasonably-sized vector, but this particular scaling factor also gives us the dominant eigenvalue, with the added guarantee that the largest component of the dominant eigenvector is 1.

Example 9. Let's use the power method to compute the dominant eigenvalue and associated eigenvector of the matrix

$$A = \begin{bmatrix} 5 & 2 \\ 2 & 4 \end{bmatrix}.$$

Suppose we start with the initial vector $v = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, and we wish to iterate 10 times.

On the first step, we compute $Av = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$. We normalize by dividing by 7 to get $v_1 = \begin{bmatrix} 1 \\ 0.85714 \end{bmatrix}$.

On the second step, we compute $Av_1 = \begin{bmatrix} 6.714 \\ 5.428 \end{bmatrix}$. We normalize by dividing by 6.714 to get $v_2 = \begin{bmatrix} 1 \\ 0.80851 \end{bmatrix}$.

We repeat this process another 8 times, with the results of all 10 iterations summarized in the following table:

i	Scaling Factor	Normalized v_i
1	7	$\begin{bmatrix}1 & 0.85714\end{bmatrix}^T$
2	6.714	$\begin{bmatrix}1 & 0.80851\end{bmatrix}^T$
3	6.617	$\begin{bmatrix}1 & 0.79099\end{bmatrix}^T$
4	6.581	$\begin{bmatrix}1 & 0.78456\end{bmatrix}^T$
5	6.569	$\begin{bmatrix}1 & 0.78218\end{bmatrix}^T$
6	6.564	$\begin{bmatrix}1 & 0.78129\end{bmatrix}^T$
7	6.562	$\begin{bmatrix}1 & 0.78097\end{bmatrix}^T$
8	6.5619	$\begin{bmatrix}1 & 0.78084\end{bmatrix}^T$
9	6.56169	$\begin{bmatrix}1 & 0.78080\end{bmatrix}^T$
10	6.56160	$\begin{bmatrix}1 & 0.78078\end{bmatrix}^T$

As we can see, this sequence converges quite close to the dominant eigenvalue, $\lambda_1 = 6.56155$.

In terms of pseudocode, we can implement the power method in a rather straightforward manner.

Algorithm 1: Power method

$v_0 \leftarrow$ arbitrary vector of same size as matrix A
for $1 \leq i \leq$ number of iterations **do**
  $v_i \leftarrow Av_{i-1} \,/\, \|Av_{i-1}\|_2$
  $\lambda_i \leftarrow \max\{|v_i|\}$
return λ, v

Performance-wise, the power method is reasonably efficient. Given an $n \times n$ matrix A, the cost of multiplying A by the vector v is approximately $2n^2$ flops, and the process of normalizing the result via the scaling factor is $O(n)$. Iterating the power method a total of m times, therefore, takes approximately $2n^2m$ flops. We can also easily adapt the power method to work for defective matrices, thus removing the need for our assumption that A is semisimple.