

1 Eigenvalues and Diagonalization

Up to now we have concentrated primarily on finding eigenvalues and eigenvectors, and while we may have claimed that this is a particularly useful endeavour, we haven't offered much to back up this claim, aside from referring the reader to a Wikipedia page. While a complete treatment of the theory and applications of eigenvalues and eigenvectors goes well beyond the scope of this book, we can offer a brief taste of what's to come (should the reader choose to further their studies in linear algebra).

1.1 Similar matrices

Definition 1. Let A and B be $n \times n$ matrices. We say that A is **similar** to B , and write $A \sim B$, if there exists an invertible $n \times n$ matrix P such that

$$A = P^{-1}BP.$$

Although our definition above is not symmetric (it defines what it means for A to be similar to B , but not vice-versa), there is in fact no ambiguity in using a statement such as “ A and B are similar matrices,” as the next theorem demonstrates.

Theorem 2. Let A , B , and C be $n \times n$ matrices. Then:

1. $A \sim A$
2. If $A \sim B$, then $B \sim A$
3. If $A \sim B$ and $B \sim C$, then $A \sim C$.

Proof. 1. Setting $P = I$, the $n \times n$ identity matrix, we have $A = I^{-1}AI$, and thus $A \sim A$.

2. Suppose $A \sim B$. Then we know that $A = P^{-1}BP$ for some invertible matrix P . Setting $Q = P^{-1}$, (and noting $Q^{-1} = P$) we have

$$B = (PP^{-1})B(PP^{-1}) = P(P^{-1}BP)P^{-1} = Q^{-1}AQ,$$

so $B \sim A$.

3. Suppose that $A \sim B$ and $B \sim C$. Then there exist $n \times n$ matrices P and Q such that $A = P^{-1}BP$ and $B = Q^{-1}CQ$. But then

$$A = P^{-1}BP = P^{-1}(Q^{-1}CQ)P = (P^{-1}Q^{-1})C(QP) = (QP)^{-1}C(QP),$$

which shows that $A \sim C$.

□

The fact that the definition of matrix similarity satisfies the three properties in the theorem above tells us that matrix similarity is an example of what is called an *equivalence relation*. This equivalence relation breaks the set of all $n \times n$ matrices into sets of “equivalent” matrices. Similar matrices are so-called because, although they often have very different

entries, they share many of the same properties. (At the end of this section, we'll discuss the fact that from the point of view of linear transformations, similar matrices describe the same linear transformation, if we represent that linear transformation using different choices of basis.) The following theorem gives some of the properties similar matrices share.

Theorem 3. *Let A and B be $n \times n$ matrices. If $A \sim B$, then:*

1. $\text{tr}(A) = \text{tr}(B)$
2. $\det(A) = \det(B)$
3. A and B have the same eigenvalues.

Proof. 1. Recall that the trace satisfies $\text{tr}(AB) = \text{tr}(BA)$ for any $n \times n$ matrices A and B . Thus, if $A = P^{-1}BP$, we have

$$\text{tr}(A) = \text{tr}(P^{-1}BP) = \text{tr}(B(P^{-1}P)) = \text{tr}(BI) = \text{tr}(B).$$

2. Recall that the determinant satisfies $\det(AB) = \det(A)\det(B)$ for any $n \times n$ matrices A and B , and $\det(A^{-1}) = \frac{1}{\det(A)}$ for any invertible matrix A . Thus, if $A = P^{-1}BP$, we have

$$\det(A) = \det(P^{-1}BP) = \det(P^{-1})\det(B)\det(P) = \frac{1}{\det(P)}\det(B)\det(P) = \det(B).$$

3. Suppose that λ is an eigenvalue of A , and that $A \sim B$. Since λ is an eigenvalue of A , we have

$$A\vec{v} = \lambda\vec{v}$$

for some nonzero vector \vec{v} . But if $A \sim B$, then $A = P^{-1}BP$ for some invertible matrix P , so

$$P^{-1}BP\vec{v} = \lambda\vec{v}.$$

Multiplying both sides on the left by P , we have

$$B(P\vec{v}) = P(\lambda\vec{v}) = \lambda(P\vec{v}).$$

Since P is invertible and $\vec{v} \neq \vec{0}$, we know that $\vec{w} = P\vec{v}$ is nonzero, and thus λ is an eigenvalue of B with corresponding eigenvector $P\vec{v}$.

□

1.2 Multiplicity of an eigenvalue

To clarify the presentation in this section, we will shift our notation somewhat from what we've been using so far in the textbook. Previously, we used the definition

$$p_A(\lambda) = \det(A - \lambda I)$$

for the characteristic polynomial of a matrix A . We will adjust this in two ways: first, we will use x as the variable in our polynomial; second, we will shift the sign of the characteristic polynomial and make the following definition:

Definition 4. The **characteristic polynomial** of an $n \times n$ matrix A is the degree- n polynomial $p_A(x)$ defined by

$$p_A(x) = \det(xI - A).$$

Notice that we are using $xI - A$ rather than $A - xI$. This guarantees that the highest-degree term in $p_A(x)$ is always x^n (with coefficient 1), whereas before this term was $\pm x^n$, depending on whether n is even or odd. We are using x as our variable to make it easier to talk about factors of the characteristic polynomial: with this change, the statements “ λ is an eigenvalue of A ” and “ $(x - \lambda)$ is a factor of $p_A(x)$ ” are equivalent.

In the previous sections, we saw examples of matrices whose characteristic polynomial had a repeated root. In such cases, it is sometimes possible (but not guaranteed) that there is more than one independent eigenvector associated with the repeated eigenvalue. Indeed, consider the matrices

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}, \quad C = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}.$$

All three matrices are upper-triangular, so they have the same characteristic polynomial, namely:

$$p(x) = (x - 2)^3.$$

Notice that the eigenvalue $\lambda = 2$ is “repeated”, in the sense that it is a repeated root of the characteristic polynomial. This leads to the following definition:

Definition 5. We say that λ is an eigenvalue with **algebraic multiplicity** k if $(x - \lambda)^k$ is a factor of $p_A(x)$, but $(x - \lambda)^{k+1}$ is not.

In other words, the algebraic multiplicity of an eigenvalue λ is the (largest) power to which the corresponding factor $(x - \lambda)$ appears in the characteristic polynomial. In our example above, all three of the matrices A , B , and C have the eigenvalue 2 with algebraic multiplicity 3. However, consider the corresponding eigenvectors in each case:

For $A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$, we have $A - 2I = 0$, so *every* non-zero vector \vec{x} is an eigenvector.

Indeed, $A = 2I$, so

$$A\vec{x} = (2I)\vec{x} = I(2\vec{x}) = 2\vec{x}$$

for every vector. In particular, we can choose the standard unit basis vectors

$$\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \vec{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

as our basic eigenvectors, and we see that there are three independent eigenvectors corresponding to the eigenvalue $\lambda = 2$.

For $B = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}$, we have $B - 2I = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$, which is row-equivalent to $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

The general solution to the equation $(B - 2I)\vec{x} = \vec{0}$ is thus given by

$$\vec{x} = \begin{bmatrix} s \\ t \\ 0 \end{bmatrix} = s \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix},$$

since the reduced row-echelon form of $B - 2I$ tells us that we must have $z = 0$, while x and y are free. Thus, in this case we have two independent basic eigenvectors, given by the standard unit basis vectors \vec{e}_1 and \vec{e}_2 .

Finally, for $C = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}$ we have $C - 2I = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$, which is already in reduced-row echelon form. From this, we can see that the system $(C - 2I)\vec{x} = \vec{0}$ has general solution

$$\vec{x} = \begin{bmatrix} t \\ 0 \\ 0 \end{bmatrix} = t\vec{e}_1,$$

so we have only one independent eigenvector associated to $\lambda = 2$; namely, \vec{e}_1 .

The above shows us that when the algebraic multiplicity of an eigenvalue is greater than one, we might have multiple independent eigenvectors associated to that eigenvalue, but this is not guaranteed. To quantify this situation, we introduce further terminology. First, we make an observation: recall that the null space of an $n \times n$ matrix A is defined by

$$\text{null}(A) = \{\vec{x} \in \mathbb{R}^n \mid A\vec{x} = \vec{0}\}.$$

In this language, we note that every eigenvector associated to an eigenvalue λ belongs to the null space of $A - \lambda I$. Indeed, $\text{null}(A - \lambda I)$ is equal to the set of all eigenvectors associated to the eigenvalue λ , along with the zero vector. This leads to another definition:

Definition 6. Let A be an $n \times n$ matrix, and let λ be any real number. The λ -**eigenspace** of A , denoted by $E(A, \lambda)$, is defined by

$$E(A, \lambda) = \text{null}(A - \lambda I).$$

A couple of remarks on this definition are in order. First, notice that λ is not assumed to be an eigenvalue in the definition above. However, for any real number λ which is not an eigenvalue, we have

$$E(A, \lambda) = \{\vec{0}\}.$$

The eigenvalues of A can thus be described as those real numbers for which $E(A, \lambda) \neq \{\vec{0}\}$. When λ is an eigenvalue, the **basic** eigenvectors associated to λ (that is, the basic solutions to the system $(A - \lambda I)\vec{x} = \vec{0}$) form a **basis** for $E(A, \lambda)$. This tells us that the dimension of each eigenspace is given by the number of independent basic eigenvectors associated to the corresponding eigenvalue, and gives us one more definition:

Definition 7. Let A be an $n \times n$ matrix. The **geometric multiplicity** of an eigenvalue λ of A is defined to be the dimension of the eigenspace $E(A, \lambda)$.

Example 8. Determine the eigenvalues of the matrix $A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$, along with their algebraic and geometric multiplicities.

Solution: We first compute the eigenvalues of A . The characteristic polynomial of A is given by

$$\begin{aligned} c_A(x) &= \begin{vmatrix} x & -1 & -1 \\ -1 & x & -1 \\ -1 & -1 & x \end{vmatrix} \\ &= x \begin{vmatrix} x & -1 \\ -1 & x \end{vmatrix} + \begin{vmatrix} -1 & -1 \\ -1 & x \end{vmatrix} - \begin{vmatrix} -1 & x \\ -1 & -1 \end{vmatrix} \\ &= x(x^2 - 1) - (x + 1) - (x + 1) \\ &= x(x - 1)(x + 1) - 2(x + 1) \\ &= (x^2 - x - 2)(x + 1) \\ &= (x - 2)(x + 1)^2. \end{aligned}$$

We see that the roots of $c_A(x)$ are $\lambda_1 = 2$, which has algebraic multiplicity 1, and $\lambda_2 = -1$, which has algebraic multiplicity 2. To determine the geometric multiplicities, we compute the corresponding eigenvectors. For $\lambda = 2$, we have

$$A - 2I = \begin{bmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus, $(A - 2I)\vec{x} = \vec{0}$ for $\vec{x} = \begin{bmatrix} t \\ t \\ t \end{bmatrix} = t \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, so $\vec{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ is the (single) basic eigenvector associated to $\lambda = 2$. This gives us

$$E(A, 2) = \text{span} \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\},$$

so the geometric multiplicity of $\lambda = 2$ is $\dim E(A, 2) = 1$. For $\lambda = -1$, we find

$$A - (-1)I = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

This tells us that the general solution to $(A + I)\vec{x} = \vec{0}$ is

$$\vec{x} = \begin{bmatrix} -s - t \\ s \\ t \end{bmatrix} = s \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix},$$

so

$$E(A, -1) = \text{span} \left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\},$$

and the geometric multiplicity of $\lambda = -1$ is $\dim E(A, -1) = 2$.

1.3 Diagonalization

One common application of matrix algebra is to a class of dynamical systems known as *linear recurrences*. A linear recurrence is a process that occurs in discrete steps, such that the state of the system at any step k depends linearly on the state of the system after the previous step. An example might be a population model for an ecosystem involving several species. For a simple two-species “predator-prey” model, we might have two populations P_1 (of rabbits, perhaps?) and P_2 (let’s say these are foxes) where the populations at a time t_k satisfy a relationship such as

$$\begin{aligned} P_1(t_k) &= aP_1(t_{k-1}) + bP_2(t_{k-1}) \\ P_2(t_k) &= cP_1(t_{k-1}) + dP_2(t_{k-1}). \end{aligned}$$

A reasonable model would probably have positive values for the coefficients a and d (absent competition, predators, etc. we expect the populations to increase), while b would be negative (more foxes mean fewer rabbits) and c would be positive (more rabbits provide more fox food). If we introduce a “population vector” $\vec{P}(t) = \begin{bmatrix} P_1(t) \\ P_2(t) \end{bmatrix}$ to describe the situation at a time t , then the above system can be written in the form $\vec{P}(t_k) = A\vec{P}(t_{k-1})$, where $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

Now, given an initial population state $\vec{P}(t_0)$, we might want to know what to expect after several generations (or seasons, or some other reasonable unit of time). Notice that we have

$$\begin{aligned} \vec{P}(t_1) &= A\vec{P}(t_0) \\ \vec{P}(t_2) &= A\vec{P}(t_1) = A(A\vec{P}(t_0)) = A^2\vec{P}(t_0) \\ \vec{P}(t_3) &= A\vec{P}(t_2) = A(A^2\vec{P}(t_0)) = A^3\vec{P}(t_0) \\ &\vdots \\ \vec{P}(t_k) &= A^k\vec{P}(t_0). \end{aligned}$$

Thus, to model the time evolution of our system, it suffices to compute powers of our matrix A . If we want to study a system over long periods of time, we need to be able to compute large powers of our matrix, and for complex systems, being able to do so efficiently will become increasingly important. (For example, we may be interested in whether or not the system settles into an equilibrium state. This would be the case, for example, if we were able to show that A^k was approximately equal to the identity matrix for large values of k .)

One case where this is easily done is when the initial state $\vec{P}_0 = \vec{P}(t_0)$ is an eigenvector for our matrix: if $A\vec{P}_0 = \lambda\vec{P}_0$, then we have

$$\begin{aligned} A^2\vec{P}_0 &= A(A\vec{P}_0) = A(\lambda\vec{P}_0) = \lambda(A\vec{P}_0) = \lambda(\lambda\vec{P}_0) = \lambda^2\vec{P}_0 \\ A^3\vec{P}_0 &= A(A^2\vec{P}_0) = A(\lambda^2\vec{P}_0) = \lambda^2(A\vec{P}_0) = \lambda^2(\lambda\vec{P}_0) = \lambda^3\vec{P}_0, \end{aligned}$$

and so on: in general, $A^k\vec{P}_0 = \lambda^k\vec{P}_0$. This is useful, since computing powers of a number is much simpler than computing powers of a matrix.

What if the initial state is not an eigenvector? It turns out that the same trick works as long as \vec{P}_0 can be written as a *linear combination* of eigenvectors: suppose that

$$\vec{P}_0 = c_1\vec{x}_1 + c_2\vec{x}_2 + \cdots + c_k\vec{x}_k,$$

where $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k$ are eigenvectors of A with eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_k$, respectively. (Note that at this point we've moved away from the original premise of a 2×2 matrix A to consider a more general situation.) In this case, for any natural number n , we have

$$\begin{aligned} A^n\vec{P}_0 &= A^n(c_1\vec{x}_1 + c_2\vec{x}_2 + \cdots + c_k\vec{x}_k) \\ &= c_1(A^n\vec{x}_1) + c_2(A^n\vec{x}_2) + \cdots + c_k(A^n\vec{x}_k) \\ &= c_1(\lambda_1^n\vec{x}_1) + c_2(\lambda_2^n\vec{x}_2) + \cdots + c_k(\lambda_k^n\vec{x}_k), \end{aligned}$$

using our previous result. Again, we only need to be able to compute powers of the eigenvalues, and not of the original matrix.

With the above in mind, we will clearly be in an advantageous situation if we can guarantee that **every** vector in our space can be written as a linear combination of eigenvectors. We know from our earlier work that in order to guarantee this, we would need to be able to construct a **basis of eigenvectors**. We will see how such a basis can be used after we make a definition.

Definition 9. We say that an $n \times n$ matrix A is **diagonalizable** (can be diagonalized) if $A \sim D$ for some diagonal matrix D . In other words, A is diagonalizable if there exists an invertible matrix P such that $D = P^{-1}AP$ is a diagonal matrix.

Theorem 10. An $n \times n$ matrix A is diagonalizable if and only if there exists a basis for \mathbb{R}^n consisting of eigenvectors of A .

Proof. First, suppose that A can be diagonalized. Then there exists an invertible matrix P such that $P^{-1}AP = D$, where

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix},$$

where $\lambda_1, \lambda_2, \dots, \lambda_n$ are the (not necessarily distinct) eigenvalues of A . Letting $\vec{e}_1, \dots, \vec{e}_n$ denote the standard basis for \mathbb{R}^n we have, for each $j = 1, 2, \dots, n$:

$$\begin{aligned} D\vec{e}_j &= \lambda_j \vec{e}_j \\ (P^{-1}AP)\vec{e}_j &= \lambda_j \vec{e}_j \\ (AP)\vec{e}_j &= P(\lambda_j \vec{e}_j) \\ A(P\vec{e}_j) &= \lambda_j (P\vec{e}_j), \end{aligned}$$

so $\vec{x}_j = P\vec{e}_j$ is an eigenvector for A with eigenvalue λ_j . Moreover, since P is invertible, we know that the vectors $\vec{x}_1, \dots, \vec{x}_n$ are linearly independent,¹ and thus form our desired basis of eigenvectors.

Conversely, suppose that we have a basis $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n$ of eigenvectors for A , with corresponding eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$. Let P be the matrix whose columns are given by these vectors. Then P is invertible, since its columns are linearly independent. Let D be defined as above. Then

$$PD = [\vec{x}_1 \ \vec{x}_2 \ \cdots \ \vec{x}_n] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = [\lambda_1 \vec{x}_1 \ \lambda_2 \vec{x}_2 \ \cdots \ \lambda_n \vec{x}_n],$$

while

$$AP = A[\vec{x}_1 \ \vec{x}_2 \ \cdots \ \vec{x}_n] = [A\vec{x}_1 \ A\vec{x}_2 \ \cdots \ A\vec{x}_n] = [\lambda_1 \vec{x}_1 \ \lambda_2 \vec{x}_2 \ \cdots \ \lambda_n \vec{x}_n].$$

Thus, $PD = AP$, and since P is invertible, this gives us $D = P^{-1}AP$, as required. \square

As we discussed above, the diagonalizability of A makes it easy to compute powers of A . Indeed, suppose $P^{-1}AP = D$ is diagonal, for some invertible matrix P . Then $A = PDP^{-1}$, and

$$A^n = (PDP^{-1})^n = (PDP^{-1}) \cdots (PDP^{-1}) = PD(P^{-1}P)D \cdots (P^{-1}P)DP^{-1} = PD^nP^{-1}.$$

This makes it easy to compute A^n , since it's easy to compute D^n : if $D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$,

$$\text{then } D^n = \begin{bmatrix} \lambda_1^n & 0 & \cdots & 0 \\ 0 & \lambda_2^n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n^n \end{bmatrix}.$$

¹Suppose that $c_1\vec{x}_1 + c_2\vec{x}_2 + \cdots + c_k\vec{x}_k = \vec{0}$ for some scalars c_1, c_2, \dots, c_k . Then we have

$$\vec{0} = c_1P\vec{e}_1 + \cdots + c_nP\vec{e}_n = P(c_1\vec{e}_1 + \cdots + c_n\vec{e}_n).$$

Since P is invertible, we can conclude that $c_1\vec{e}_1 + \cdots + c_n\vec{e}_n = \vec{0}$, and thus $c_1 = c_2 = \cdots = c_n = 0$, as required.

The proof Theorem 10 is quite useful, since it tells us two things: first, we now know what it takes to diagonalize a matrix: for an $n \times n$ matrix, we need to be able to find n independent eigenvectors. Second, the proof of the theorem tells us **how** to diagonalize: once we've found our eigenvectors, we construct the matrix P whose columns are given by the eigenvectors of A ; the matrix $P^{-1}AP$ will then be a diagonal matrix with the eigenvalues of A on the main diagonal. This leaves us with the question: when can we be sure we have enough eigenvectors? To answer this, we begin with a theorem:

Theorem 11. *The eigenvectors $\vec{x}_1, \dots, \vec{x}_k$ corresponding to **distinct** eigenvalues of A are linearly independent.*

Corollary 12. *If an $n \times n$ matrix A has n distinct eigenvalues, then A is diagonalizable.*

Let us first see how the above corollary follows from our theorem. We know that every eigenvalue has at least one associated eigenvector. If there are n distinct eigenvalues, then we have n associated eigenvectors which, by the theorem above, are linearly independent, and we know that any set of n linearly independent vectors in \mathbb{R}^n forms a basis for \mathbb{R}^n .

As for the proof of the theorem, we use an argument by contradiction: suppose, to the contrary, that the eigenvectors $\vec{x}_1, \dots, \vec{x}_k$ are linearly *dependent*. Then there is some m , $2 \leq m \leq k-1$, such that the vectors $\vec{x}_1, \dots, \vec{x}_m$ are linearly independent, but \vec{x}_{m+1} can be written as a linear combination of the vectors $\vec{x}_1, \dots, \vec{x}_m$; that is,

$$\vec{x}_{m+1} = c_1\vec{x}_1 + c_2\vec{x}_2 + \dots + c_m\vec{x}_m, \quad (1)$$

for some scalars c_1, \dots, c_m . Then, multiplying both sides of (1) on the left by A , we must have

$$A\vec{x}_{m+1} = A(c_1\vec{x}_1 + c_2\vec{x}_2 + \dots + c_m\vec{x}_m) = c_1(A\vec{x}_1) + c_2(A\vec{x}_2) + \dots + c_m(A\vec{x}_m),$$

but each vector \vec{x}_i is an eigenvector of A with eigenvalue λ_i . Thus,

$$\lambda_{m+1}\vec{x}_{m+1} = c_1\lambda_1\vec{x}_1 + c_2\lambda_2\vec{x}_2 + \dots + c_m\lambda_m\vec{x}_m.$$

On the other hand, if we multiply both sides of (1) by the scalar λ_{m+1} , we have

$$\lambda_{m+1}\vec{x}_{m+1} = c_1\lambda_{m+1}\vec{x}_1 + c_2\lambda_{m+1}\vec{x}_2 + \dots + c_m\lambda_{m+1}\vec{x}_m.$$

Subtracting these last two equations, we find:

$$\vec{0} = c_1(\lambda_{m+1} - \lambda_1)\vec{x}_1 + c_2(\lambda_{m+1} - \lambda_2)\vec{x}_2 + \dots + c_m(\lambda_{m+1} - \lambda_m)\vec{x}_m.$$

Now, we are assuming that our eigenvalues are all *distinct*. Thus, $\lambda_{m+1} - \lambda_i \neq 0$ for all $i = 1, 2, \dots, m$. On the other hand, the vectors $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_m$ are assumed to be linearly independent, so we must have

$$c_1(\lambda_{m+1} - \lambda_1) = 0, c_2(\lambda_{m+1} - \lambda_2) = 0, \dots, c_m(\lambda_{m+1} - \lambda_m) = 0.$$

Since none of the $\lambda_{m+1} - \lambda_i$ terms vanish, it must be the case that $c_1 = c_2 = \dots = c_m = 0$. But if that is true, then

$$\vec{x}_{m+1} = 0\vec{x}_1 + 0\vec{x}_2 + \dots + 0\vec{x}_m = \vec{0},$$

which is impossible, since \vec{x}_{m+1} was assumed to be an eigenvector, and eigenvectors are nonzero. Thus, it must be the case that all of our eigenvectors are linearly independent.

The above tells us that any $n \times n$ matrix A with n distinct eigenvalues is automatically diagonalizable. There are two other possibilities. One is that the characteristic polynomial of A cannot be completely factored over the real numbers. In this case, diagonalization is impossible, *unless* we are willing to consider **complex** eigenvalues. Recall (from way back in Chapter 2) that over the complex numbers, *every* polynomial can be completely factored. (This is the Fundamental Theorem of Algebra.) In this case, we may still be able to diagonalize, but the eigenvectors corresponding to the complex eigenvalues may well be complex themselves, so the matrix P used to diagonalize A could have complex entries.

The other possibility is that we have repeated eigenvalues. Here, it is possible that we encounter truly insurmountable obstacles. First, we have the following theorem, which we will state without proof. The interested reader can easily find the details in other linear algebra textbooks or online.

Theorem 13. *Let A be an $n \times n$ matrix, and for each real number λ , let $m_\lambda(A)$ denote the algebraic multiplicity of λ . (If λ is not an eigenvalue of A , we set $m_\lambda(A) = 0$.) Then for all numbers λ , we have*

$$\dim \text{null}(A - \lambda I) \leq m_\lambda(A).$$

In other words, the geometric multiplicity of an eigenvalue is always less than or equal to the algebraic multiplicity. Note that the sum of the algebraic multiplicities is always equal to the degree of the characteristic polynomial, which, in turn, is equal to n , for an $n \times n$ matrix A . On the other hand, the total number of independent eigenvectors for A is equal to the sum of the geometric multiplicities. (We know that eigenvectors for *different* eigenvalues are independent, and the geometric multiplicity tells us how many independent eigenvectors we'll have for a single eigenvalue.) As a result, we have the following:

Theorem 14. *An $n \times n$ matrix A is diagonalizable if and only if $\dim \text{null}(A - \lambda I) = m_\lambda(A)$ for each eigenvalue λ of A ; that is, if the geometric multiplicity of each eigenvalue is equal to its algebraic multiplicity.*

This also tells us exactly when we can expect a matrix to fail to be diagonalizable: this will be the case if A has an eigenvalue of algebraic multiplicity greater than 1 for which the geometric multiplicity is less than the algebraic multiplicity. (That is, if we don't get "enough" basic eigenvectors corresponding to a repeated eigenvalue.)

The reader may wonder if there is anything further that can be said in the case that a matrix A is not diagonalizable, and indeed, there is. As long as we are willing to work over the complex numbers (to avoid situations where the characteristic polynomial cannot be factored), one can prove that in the worst-case scenario, there exists an invertible matrix P such that the matrix $P^{-1}AP$ is triangular. (If we can't get a diagonal matrix, upper or lower-triangular is the next best thing.) The standard form of such a matrix is called the **Jordan Canonical Form**; it is obtained using what are called *generalized eigenvectors*. The Jordan Canonical Form of a matrix is studied in more advanced courses in linear algebra.

One last important result that we do not have time to study concerns the case of symmetric matrices. Recall that an $n \times n$ matrix A is *symmetric* if $A^T = A$. An important result

called the **Spectral Theorem** guarantees that every symmetric matrix A is diagonalizable. In fact, the spectral theorem goes one step further. Recall that two vectors \vec{u} and \vec{v} are *orthogonal* if $\vec{u} \cdot \vec{v} = 0$. We say that a *set* of vectors $\{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k\}$ is **orthogonal** if $\vec{x}_i \neq \vec{0}$ for each i , and $\vec{x}_i \cdot \vec{x}_j = 0$ for all $i \neq j$. Orthogonality is a “stronger” condition than linear independence. Indeed, if the vectors $\vec{x}_1, \dots, \vec{x}_k$ are orthogonal and

$$c_1\vec{x}_1 + c_2\vec{x}_2 + \dots + c_k\vec{x}_k = \vec{0}$$

for some scalars c_1, c_2, \dots, c_k , taking the dot product of both sides of the above equation with \vec{x}_1 gives us

$$c_1(\vec{x}_1 \cdot \vec{x}_1) + c_2(0) + \dots + c_k(0) = 0,$$

and since $\vec{x}_1 \cdot \vec{x}_1 = \|\vec{x}_1\|^2 \neq 0$, it must be that $c_1 = 0$, and similarly, all the other scalars must be zero as well.

For a symmetric matrix, eigenvectors corresponding to distinct eigenvalues are not just independent, they’re orthogonal. Indeed, since $\vec{x}_1 \cdot \vec{x}_2$ can be written as the matrix product $\vec{x}_1^T \vec{x}_2$, we have (using the fact that $A^T = A$):

$$\lambda_1(\vec{x}_1 \cdot \vec{x}_2) = (\lambda_1 \vec{x}_1)^T \vec{x}_2 = (A\vec{x}_1)^T \vec{x}_2 = (\vec{x}_1^T A^T) \vec{x}_2 = \vec{x}_1^T (A\vec{x}_2) = \vec{x}_1^T (\lambda_2 \vec{x}_2) = \lambda_2(\vec{x}_1 \cdot \vec{x}_2),$$

Thus, $(\lambda_1 - \lambda_2)(\vec{x}_1 \cdot \vec{x}_2) = 0$, and since $\lambda_1 \neq \lambda_2$, we must have $\vec{x}_1 \cdot \vec{x}_2 = 0$.

It is possible to prove that for a symmetric matrix, one can find an **orthogonal basis** of eigenvectors. This observation has a number of important applications. (Try an online search for “orthogonal diagonalization” and you should be able to find several examples.)

1.4 Matrix similarity and change of basis for transformations

We end this section with the promised explanation of how similar matrices can be viewed as different representations of the same linear transformation. If we think of matrices in terms of the matrix transformations they define, then two matrix transformations obtained from similar matrices should be viewed as two descriptions of the same linear transformation using different “coordinate systems”. Let us explain what we mean here. Suppose

$$T : \mathbb{R}^n \rightarrow \mathbb{R}^n$$

is a matrix transformation defined by $T(\vec{x}) = A\vec{x}$ for some $n \times n$ matrix A . Let us consider this to be a transformation defined in terms of the standard basis vectors $\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n$ for \mathbb{R}^n . We can then write

$$\vec{x} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}^T = x_1\vec{e}_1 + x_2\vec{e}_2 + \dots + x_n\vec{e}_n$$

and think of T as a function of the variables x_1, x_2, \dots, x_n . To put this another way, when we speak of a matrix transformation with standard matrix A , that matrix A is defined with respect to the standard basis. In general, let $\mathcal{A} = \{\vec{e}_1, \dots, \vec{e}_n\}$ denote our standard basis, and suppose

$$\mathcal{B} = \{\vec{b}_1, \vec{b}_2, \dots, \vec{b}_n\}$$

is another basis for \mathbb{R}^n . Recall that when we say that the set \mathcal{B} is a *basis*, we mean that $\text{span } \mathcal{B} = \mathbb{R}^n$, so that every vector in \mathbb{R}^n can be written as a linear combination of the vectors

in \mathcal{B} , and \mathcal{B} is linearly independent, which means that every vector in \mathbb{R}^n can be written *uniquely* as a linear combination of the vectors in \mathcal{B} .

Let $\vec{v} = y_1\vec{b}_1 + y_2\vec{b}_2 + \cdots + y_n\vec{b}_n$ any vector in \mathbb{R}^n . Using the same linear transformation T as above, we have

$$T(\vec{v}) = T(y_1\vec{b}_1 + y_2\vec{b}_2 + \cdots + y_n\vec{b}_n) = y_1T(\vec{b}_1) + y_2T(\vec{b}_2) + \cdots + y_nT(\vec{b}_n).$$

Now, each of the vectors $T(\vec{b}_i)$ can, in turn, be written in terms of the basis \mathcal{B} : we have

$$\begin{aligned} T(\vec{b}_1) &= b_{11}\vec{b}_1 + b_{21}\vec{b}_2 + \cdots + b_{n1}\vec{b}_n \\ T(\vec{b}_2) &= b_{12}\vec{b}_1 + b_{22}\vec{b}_2 + \cdots + b_{n2}\vec{b}_n \\ &\vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\ T(\vec{b}_n) &= b_{1n}\vec{b}_1 + b_{2n}\vec{b}_2 + \cdots + b_{nn}\vec{b}_n. \end{aligned}$$

The coefficients b_{ij} obtained above determine a new matrix B , which we define to be the **matrix of T with respect to the basis \mathcal{B}** . Notice that if we write $\vec{y} = [y_1 \ y_2 \ \cdots \ y_n]^T$, we have

$$B\vec{y} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} b_{11}y_1 + b_{12}y_2 + \cdots + b_{1n}y_n \\ b_{21}y_1 + b_{22}y_2 + \cdots + b_{2n}y_n \\ \vdots \\ b_{n1}y_1 + b_{n2}y_2 + \cdots + b_{nn}y_n \end{bmatrix}.$$

Let's introduce some notation: for any vector \vec{w} in \mathbb{R}^n , we will let $[\vec{w}]_{\mathcal{B}}$ denote the $n \times 1$ column vector defined as follows:

$$\text{If } \vec{w} = w_1\vec{b}_1 + w_2\vec{b}_2 + \cdots + w_n\vec{b}_n, \text{ then } [\vec{w}]_{\mathcal{B}} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}.$$

In this notation, we have $\vec{y} = [\vec{v}]_{\mathcal{B}}$, and $[T(\vec{v})]_{\mathcal{B}} = \begin{bmatrix} b_{11}y_1 + b_{12}y_2 + \cdots + b_{1n}y_n \\ b_{21}y_1 + b_{22}y_2 + \cdots + b_{2n}y_n \\ \vdots \\ b_{n1}y_1 + b_{n2}y_2 + \cdots + b_{nn}y_n \end{bmatrix}$, since

$$\begin{aligned} &(b_{11}y_1 + \cdots + b_{1n}y_n)\vec{b}_1 + \cdots + (b_{n1}y_1 + \cdots + b_{nn}y_n)\vec{b}_n \\ &= y_1(b_{11}\vec{b}_1 + \cdots + b_{n1}\vec{b}_n) + \cdots + y_n(b_{1n}\vec{b}_1 + \cdots + b_{nn}\vec{b}_n) \\ &= y_1T(\vec{b}_1) + \cdots + y_nT(\vec{b}_n) \\ &= T(y_1\vec{b}_1 + \cdots + y_n\vec{b}_n) \\ &= T(\vec{v}). \end{aligned}$$

Let's tie everything together. First, let P be the $n \times n$ matrix whose columns are the vectors in \mathcal{B} . Then we know that

$$\vec{b}_1 = P\vec{e}_1, \vec{b}_2 = P\vec{e}_2, \dots, \vec{b}_n = P\vec{e}_n,$$

and conversely,

$$\vec{e}_1 = P^{-1}\vec{b}_1, \vec{e}_2 = P^{-1}\vec{b}_2, \dots, \vec{e}_n = P^{-1}\vec{b}_n.$$

If $[\vec{v}]_{\mathcal{B}} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$, then we have

$$\vec{v} = v_1\vec{b}_1 + \dots + v_n\vec{b}_n = v_1P\vec{e}_1 + \dots + v_nP\vec{e}_n = P(v_1\vec{e}_1 + \dots + v_n\vec{e}_n) = P[\vec{v}]_{\mathcal{B}}.$$

Note that we can also write $\vec{v} = [\vec{v}]_{\mathcal{A}}$, where \mathcal{A} is the standard basis, since by default, all of our vectors are written in terms of the standard basis. Thus, we can conclude that

$$[\vec{v}]_{\mathcal{A}} = P[\vec{v}]_{\mathcal{B}} \quad (2)$$

defines the relationship between the column vector representations of our vector with respect to the two bases. Similarly, we can write

$$[T(\vec{v})]_{\mathcal{A}} = P[T(\vec{v})]_{\mathcal{B}} \quad (3)$$

for the column representations of the output of our linear transformation T . However, we also know that our transformation was originally defined using the matrix A according to

$$[T(\vec{v})]_{\mathcal{A}} = A[\vec{v}]_{\mathcal{A}} \quad (4)$$

and that the matrix B above was defined such that

$$[T(\vec{v})]_{\mathcal{B}} = B[\vec{v}]_{\mathcal{B}}. \quad (5)$$

Equating the two expressions above for $[T(\vec{v})]_{\mathcal{A}}$, in Equations (3) and (4) we have

$$P[T(\vec{v})]_{\mathcal{B}} = A[\vec{v}]_{\mathcal{A}}.$$

Plugging in Equations (5) on the left and (2) on the right, we have

$$PB[\vec{v}]_{\mathcal{B}} = AP[\vec{v}]_{\mathcal{B}}.$$

Since this must be true for any vector, we conclude that $PB = AP$, and thus

$$B = P^{-1}AP,$$

so B is similar to A .