Math 217: Eigen Everything[1]

1. Eigenvectors and Eigenvalues.

Eigenvectors and eigenvalues are objects associated to a *fixed linear transformation*

$$V \xrightarrow{T} V$$

where the source and target are the *same.*[2] Eigenvectors and eigenvalues *always go together:* each eigenvector of T has some associated eigenvalue, and each eigenvalue has associated eigenvectors.

Definition 1.1. An **eigenvector** of a linear transformation $V \xrightarrow{T} V$ is any non-zero vector $v \in V$ such that $T(v) = \lambda v$ for some scalar λ. The scalar λ is called the **eigenvalue** of the eigenvector v.

It bears repeating:

- An eigenvector is a VECTOR (in V) and an eigenvalue is a SCALAR (in $\mathbb{R}$).
- These are objects associated to a given linear transformation from a space *to itself.*

Eigenvectors are important geometrically because they help us understand the map better—if T is a transformation which has eigenvector v with eigenvalue 2, then we know that T is scaling vectors in the direction of v by two. That is, T is pulling/stretching vectors in the direction of v to twice their length.

CAUTION: Not every linear transformation has an eigenvalue!

Example 1.2. Consider the map $\mathbb{R}^2 \xrightarrow{\rho} \mathbb{R}^2$ given by rotation counterclockwise through $\pi/2$. Since no non-zero vector is taken to a scalar multiple of itself, ρ has no eigenvectors.

Example 1.3. Consider the differentiation map of $\mathcal{C}^\infty$, the vector space of all infinitely differentiable functions. Since $\frac{d}{dx}e^{\lambda x} = \lambda e^{\lambda x}$ for any real number λ, we see that every real number is an eigenvalue of the differentiation map, with corresponding eigenvectors $f_\lambda(x) = e^{\lambda x}$.

Eigenvectors are important computationally because they often provide convenient coordinate systems in which it is simpler to perform computations. This is especially true when we can find a basis consisting of eigenvectors:

Definition 1.4. Let $V \xrightarrow{T} V$ be a linear transformation. An **eigenbasis** is a basis for V consisting of eigenvectors for T.

[1]Thanks to Harrison Centner, Joey DiCresce, Aelita Klausmeier, Brendan Nell and Shiqi Sheng for valuable feedback on an earlier draft.

[2]In German, *eigen* means "same" or "itself".

CAUTION: Not every linear transformation has an eigenbasis! For example, the rotation transformation in Example 1.2 does not have an eigenbasis (because it has no eigenvectors!)

Example 1.5. Let's look at an example of a transformation that *does* have an eigenbasis. Consider the map

$$\mathbb{R}^2 \xrightarrow{p} \mathbb{R}^2$$

given by projection onto a line L through the origin.

Since every point of L is taken to itself, we have $p(\vec{v}) = \vec{v}$ for all $\vec{v} \in L$. That is, *every non-zero vector in L is an eigenvector with eigenvalue 1.*

Since every vector $w \in L^{\perp}$ is crushed to $\vec{0}$ by the transformation p, we have $p(\vec{w}) = \vec{0} = 0\,\vec{w}$ for all $\vec{w} \in L^{\perp}$. Thus *every non-zero vector in $L^{\perp}$ is an eigenvector with eigenvalue* 0.

If we let $\vec{v}$ be a basis for L and $\vec{w}$ be a basis for $L^{\perp}$, then $(\vec{v}, \vec{w})$ is an eigenbasis for p.

An eigenbasis for a transformation T is important algebraically because the coordinate system it determines is especially helpful for dealing with T. For example, in Example 1.5, the matrix of p in the eigenbasis $\mathcal{B} = (\vec{v}, \vec{w})$ is

$$[p]_{\mathcal{B}} = \big[[p(\vec{v})]_{\mathcal{B}} \quad [p(\vec{w})]_{\mathcal{B}}\big] = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Imagine how much easier it is to work with this $\mathcal{B}$-matrix—which is mostly zeros—instead of the standard matrix or some other matrix of p!

This is a general phenomenon: a linear transformation will be represented by a *diagonal matrix* if and only if we model it using an *eigenbasis:*

Theorem 1.6. *Let $V \xrightarrow{T} V$ be a linear transformation of a finite dimensional vector space.*

- *If T has an eigenbasis $\mathcal{B}$, then the matrix $[T]_{\mathcal{B}}$ is a diagonal matrix.*
- *Conversely, if $\mathcal{B}$ is a basis for V such that $[T]_{\mathcal{B}}$ is diagonal, then the basis $\mathcal{B}$ is an eigenbasis.*

In this case, the elements on the diagonal of the matrix $[T]_{\mathcal{B}}$ are precisely the eigenvalues of the vectors in $\mathcal{B}$, in the same order.

Proof. Suppose $\mathcal{B} = (v_1, \dots, v_n)$ is an eigenbasis for T. Let us compute the $\mathcal{B}$-matrix for T. Its i-th column is $[T(v_i)]_{\mathcal{B}}$. But because v_i is an eigenvector, we know that $T(v_i) = \lambda_i v_i$ for some scalar $\lambda_i \in \mathbb{R}$. So the $\mathcal{B}$-coordinate vector of $T(v_i)$ is $[T(v_i)]_{\mathcal{B}} = [\lambda_i v_i]_{\mathcal{B}} = \lambda_i \vec{e}_i$. Since the ith column of $[T]_{\mathcal{B}}$ is $\lambda_i \vec{e}_i$, we see that

$$[T]_{\mathcal{B}} = \begin{bmatrix} \lambda_1 & 0 & 0 & \dots & 0 & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & \dots & 0 & \lambda_n \end{bmatrix}$$

Conversely, suppose $\mathcal{B} = (v_1, v_2, \dots, v_n)$ is a basis for V such that $[T]_{\mathcal{B}}$ has the diagonal form above. The i-th column is $T(v_i)$ expressed in $\mathcal{B}$-coordinates as $0\,v_1 + 0v_2 + \cdots + \lambda_i\,v_i + 0\;v_{i+1} + \cdots + 0\,v_n$. That is, $T(v_i) = \lambda_i\;v_i$ for each $i = 1, \dots n$. This exactly says that each element of the basis $\mathcal{B}$ is an eigenvector with eigenvalue λ_i. $\square$

This general phenomenon—that a transformation has an eigenbasis if and only if it can be represented by a diagonal matrix—justifies the next definition:

Definition 1.7. A linear transformation $V \xrightarrow{T} V$ is **diagonalizable** if V admits an eigenbasis for T. Equivalently, a linear transformation $V \xrightarrow{T} V$ is **diagonalizable** if it can be represented by a diagonal matrix $[T]_{\mathcal{B}}$ for some basis $\mathcal{B}$.

Recall that, by definition, two $n \times n$ matrices A and B are **similar** if there exists an invertible matrix S such that $A = SBS^{-1}$. Another way to think about similarity is this: two $n \times n$ matrices A and B are similar if and only if they represent the *same linear transformation in two different coordinate systems.*

To see why, say $T : V \to V$ be a linear transformation of an n-dimensional vector space. Let $\mathcal{A}$ and $\mathcal{B}$ be two bases for V. We can model T with either basis, of course. Remember how the matrices representing T in these two basis are related:

$$[T]_{\mathcal{A}} = S_{\mathcal{B}\to\mathcal{A}}\ [T]_{\mathcal{B}}\ S_{\mathcal{A}\to\mathcal{B}},$$

or

$$[T]_{\mathcal{A}} = S\ [T]_{\mathcal{B}}\ S^{-1},$$

where $S = S_{\mathcal{B}\to\mathcal{A}}$ is the change of basis matrix from $\mathcal{B}$-coordinates to $\mathcal{A}$-coordinates. That is, $[T]_{\mathcal{A}}$ and $[T]_{\mathcal{B}}$ are similar.

This leads to the following important idea:

Theorem 1.8. *Let $T : V \to V$ be a linear transformation of an n-dimensional vector space. Let $\mathcal{A}$ be an arbitrary basis for V. Then T is diagonalizable if and only if $[T]_{\mathcal{A}}$ is similar to a diagonal matrix.*

Proof. By definition, T is diagonalizable if and only if it has an eigenbasis—call it $\mathcal{B}$— and in this case, $[T]_{\mathcal{B}}$ is diagonal. Since $[T]_{\mathcal{B}}$ and $[T]_{\mathcal{A}}$ are similar, we see that $[T]_{\mathcal{A}}$ is similar to a diagonal matrix.

Conversely, suppose $[T]_{\mathcal{A}}$ is similar to a diagonal matrix D. This means we can write $[T]_{\mathcal{A}} = SDS^{-1}$ for some invertible S. Construct a new basis $\mathcal{B}$ whose elements are the vectors in V whose $\mathcal{A}$-coordinate columns are the columns of S. By construction $S = S_{\mathcal{B}\to\mathcal{A}}$, and $[T]_{\mathcal{B}} = S_{\mathcal{A}\to\mathcal{B}}[T]_{\mathcal{A}}S_{\mathcal{B}\to\mathcal{A}} = S^{-1}[T]_{\mathcal{A}}S = D$. So $[T]_{\mathcal{B}}$ is diagonal, and $\mathcal{B}$ is an eigenbasis. So if $[T]_{\mathcal{A}}$ is similar to a diagonal matrix, then T is diagonalizable. $\square$

1.9. **Matrices.** A common (and useful) abuse of notation is to speak about matrices as though they are transformations. So, for example, we are led to the following:

Definition 1.10. An $n \times n$ matrix A is **diagonalizable** if the linear transformation

$$\mathbb{R}^n \xrightarrow{T_A} \mathbb{R}^n \quad \text{defined by} \quad T_A(\vec{x}) = A\vec{x}$$

is diagonalizable.

In the language of matrices, Theorems 1.6 and 1.8 can be combined into the following:

Theorem 1.11. *A square matrix A is diagonalizable if and only if A is similar to a diagonal matrix. In this case, we can write $A = SDS^{-1}$ where*

- *S is an $n \times n$ matrix whose columns are an eigenbasis for A and*
- *D is a diagonal matrix whose entries are the eigenvalues of the elements in the eigenbasis (in the same order).*

Proof. Suppose that A is diagonalizable in the sense that T_A has an eigenbasis $\mathcal{B} = (\vec{v}_1, \ldots, \vec{v}_n)$. Let $D = [T]_{\mathcal{B}}$. We know that D is diagonal by Theorem 1.6. On the other hand, we know that any two bases in which we write a transformation T will have corresponding matrices related by the formula

$$[T]_{\mathcal{A}} = S_{\mathcal{B}\to\mathcal{A}}\ [T]_{\mathcal{B}}\ S_{\mathcal{A}\to\mathcal{B}}.$$

If we take T to be the transformation T_A and write it in the standard basis $\mathcal{E}$, we know that $[T_A]_{\mathcal{E}} = A$. So this formula becomes

$$A = S_{\mathcal{B}\to\mathcal{E}}\ [T_A]_{\mathcal{B}}\ S^{-1}_{\mathcal{B}\to\mathcal{E}}.$$

Finally, note that the change of basis $S_{\mathcal{B}\to\mathcal{E}}$ is easy to write: its columns are the elements of the basis $\mathcal{B}$ written in standard coordinates—that is, its columns are just $\vec{v}_1, \ldots, \vec{v}_n$. So taking $\mathcal{B}$ to be an eigenbasis for T_A, we have that $[T_A]_{\mathcal{B}} = D$ is diagonal with the eigenvalue of the i-th element of $\mathcal{B}$ in spot ii, and we can write

$$A = S\ D\ S^{-1}$$

where S and D satisfy the desired conditions. □

Example 1.12. Consider the map $\mathbb{R}^2 \xrightarrow{T} \mathbb{R}^2$ given by left multiplication by $\begin{bmatrix} 2 & 3 \\ 0 & -1 \end{bmatrix}$. Note that $T(\vec{e}_1) = 2\vec{e}_1$. Also, for $\vec{v} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$, we have $T(\vec{v}) = -\vec{v}$. So

$$\mathcal{B} = \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right)$$

is an eigenbasis and the matrix of T in this eigenbasis is

$$[T]_{\mathcal{B}} = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix}.$$

Note that the elements on the diagonal are exactly the eigenvalues $\{2, -1\}$.

Furthermore, since the change of basis matrix $S_{\mathcal{B}\to\mathcal{E}}$ is $\big[[\vec{e}_1]_{\mathcal{E}} \quad [\vec{v}]_{\mathcal{E}}\big] = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}$, we also have

$$\begin{bmatrix} 2 & 3 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}^{-1}.$$

2. Eigenspaces

Proposition 2.1. *Let λ be an eigenvalue of a linear transformation $V \xrightarrow{T} V$. The set*

$$E_\lambda = \{v \in V \mid T(v) = \lambda v\} \subset V$$

consisting of all λ-eigenvectors (together with 0_V) is a **subspace** *of V.*

Proof. You should prove this, using the standard technique for checking a subset of V is a subspace. □

Definition 2.2. Let $V \xrightarrow{T} V$ be a linear transformation with eigenvalue λ. The **λ-eigenspace** is the subspace

$$E_\lambda = \{v \in V \mid T(v) = \lambda v\}$$

of V.

Definition 2.3. Let $V \xrightarrow{T} V$ be a linear transformation with eigenvalue λ. The **geometric multiplicity** of λ is the dimension of the λ-eigenspace

$$E_\lambda = \{v \in V \mid T(v) = \lambda v\}.$$

Example 2.4. Eigenspaces are subspaces of the *source* of a linear transformation, just like the kernel is. In fact, the zero-eigenspace of a transformation $V \xrightarrow{T} V$ is *precisely* the kernel of T:

$$E_0 = \{v \in V \mid T(v) = 0\ v\} = \{v \in V \mid T(v) = 0_V\}.$$

The geometric multiplicity of 0 is therefore the nullity of T. More generally, you should think of eigenspaces as subspaces similar to the kernel of T, or as a generalization of the kernel. We will explore this in more depth in the next section.

Remark 2.5. Even if λ is not an eigenvalue of T, we can define $E_\lambda = \{v \in V \mid T(v) = \lambda v\}$. What is E_λ when λ is not an eigenvalue? Since there are no non-zero vectors such that $T(v) = \lambda v$. The *only* vector in V satisfying $T(v) = \lambda v$ is the zero vector. Thus $E_\lambda = \{0_V\}$ if and only if λ is *not* an eigenvalue for T.

Example 2.6. Consider the linear transformation $\mathbb{R}^{2\times 2} \xrightarrow{T} \mathbb{R}^{2\times 2}$ sending A to $A + A^\top$.

Let us compute the 2-eigenspace for this transformation. We claim that the 2-eigenspace of T is the subspace of symmetric matrices. To see this, first say A is symmetric. This means that $A = A^\top$, so that $T(A) = A + A^\top = 2A$. This shows that every symmetric matrix is a 2-eigenvector for T, or equivalently, the set of symmetric matrices is contained in the 2-eigenspace. In the other direction, suppose that A is in the 2-eigenspace of T. This means that $T(A) = 2A$, which in turn means that $A + A^\top = 2A$. Subtracting A from both sides, we see that $A = A^\top$, which means A is symmetric. This proves that E_2 is precisely the subspace of symmetric matrices.

How can we compute the geometric multiplicity of the eigenvalue 2? Well, this is the dimension of E_2, which we just observed is the subspace of all symmetric 2×2 matrices. A basis for the subspace of symmetric matrices is

$$\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right).$$

So the 2-eigenspace E_2 has dimension three, or equivalently, the geometric multiplicity of the eigenvalue 2 is three.

Are there any other eigenvalues and vectors for the map T? Well, the zero-eigenspace is simply the kernel of T. Observe that $\ker T = \{A \mid A = -A^\top\}$, the set of so-called *skew-symmetric* matrices. The matrix $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ can be easily seen to span the kernel, hence it is a basis for the zero-eigenspace E_0.

To summarize, the transformation $\mathbb{R}^{2\times 2} \xrightarrow{T} \mathbb{R}^{2\times 2}$ sending A to $A+A^\top$ has two eigenspaces:

- The 2-eigenspace is the space of symmetric matrices; it has dimension three.
- The 0-eigenspace is the space of skew-symmetric matrices; it has dimension one.

Putting together bases for each of these eigenspaces gives us an eigenbasis

$$\mathfrak{B} = \left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \right).$$

for T. Indeed, we know that the vector space $\mathbb{R}^{2\times 2}$ is four-dimensional, so any set of four linearly independent vectors in $\mathbb{R}^{2\times 2}$—including these four eigenvectors— will be a basis.

Now, if we represent T in this eigenbasis $\mathfrak{B}$, we get

$$[T]_{\mathfrak{B}} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

which is diagonal. This is the beauty of eigenvectors! Imagine how much easier it will be to work with this matrix full of many zeros!

Although it is not completely obvious, there are no other eigenvalues for the transformation T in Example 2.6. One important fact in understanding why is the following theorem:

Theorem 2.7. *Let $V \xrightarrow{T} V$ be a linear transformation. Let $\mathcal{B}_1, \mathcal{B}_2 \ldots, \mathcal{B}_r$ be bases for eigenspaces $E_{\lambda_1}, \ldots, E_{\lambda_r}$ of distinct eigenvalues $\lambda_1, \ldots, \lambda_r$. Then the set*

$$\mathcal{B}_1 \cup \cdots \cup \mathcal{B}_r$$

is linearly independent.

Do you see why Theorem 2.7 implies that the linear transformation in Example 2.6 has no further eigenvalues beyond 0 and 2? If it did, there would be a further eigenvector B, and the set

$$\left\{ \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \right\} \cup \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right\} \cup \{B\}$$

would be a set of five linearly independent elements in $\mathbb{R}^{2\times 2}$. This would contradict the fact that $\mathbb{R}^{2\times 2}$ is dimension four.

To prove Theorem 2.7, we need the following theorem:

Theorem 2.8. *Let $V \xrightarrow{T} V$ be a linear transformation. If $v_1\, v_2, \ldots, v_n$ are eigenvectors with distinct eigenvalues, then the set $\{v_1\, v_2, \ldots, v_n\}$ is linearly independent.*

Proof. We induce on n. If $n = 1$, suppose that $a_1 v_1 = 0$ is a relation. But since $v_1 \neq 0$, this implies $a_1 = 0$, so the relation is trivial.

Now suppose, inductively, that any set of $n-1$ eigenvectors for T with *distinct* eigenvalues is linearly independent. We will show that the same is true for a set $\{v_1\, v_2, \ldots, v_n\}$ of n eigenvectors with distinct eigenvalues.

Suppose we have a relation

$$a_1 v_1 + \cdots + a_n v_n = 0 \tag{2.8.1}$$

To show $\{v_1\, v_2, \ldots, v_n\}$, we need to show the relation is trivial. Apply T to the relation (2.8.1). Because each v_i is an eigenvector, this produces a new relation

$$a_1 \lambda_1 v_1 + \cdots + a_n \lambda_n v_n = 0, \tag{2.8.2}$$

where λ_i is the eigenvalue of the eigenvector v_i. Therefore, multiplying the relation (2.8.1) by λ_n and subtracting it from the relation (2.8.2), we get a new relation

$$(\lambda_1 - \lambda_n) a_1 v_1 + \cdots + (\lambda_{n-1} - \lambda_n) a_{n-1} v_{n-1} = 0. \tag{2.8.3}$$

This is a relation on the eigenvectors $\{v_1, v_2, \ldots, v_{n-1}\}$, which are linearly independent by induction. This means that each coefficient

$$(\lambda_i - \lambda_n) a_i = 0$$

for $i = 1, \ldots, n-1$. Since the λ_i are distinct, we have $\lambda_i - \lambda_n \neq 0$, and can conclude that $a_1 = a_2 = \cdots = a_{n-1} = 0$. Looking back at our original relation (2.8.1), we are left only with

$$a_n v_n = 0.$$

But again, since the eigenvector v_n is non-zero, we can also conclude that $a_n = 0$. That is, our relation is trivial and so the eigenvectors $\{v_1, \ldots, v_n\}$ are linearly independent. QED. $\square$

Finally, we use Theorem 2.8 to prove Theorem 2.7. This is easier to understand than it looks—only the unwieldy notation makes it look complicated.

Proof of Theorem 2.7. Fix i, and let $\mathcal{B}_i = (v_{i1}, \ldots, v_{im_i})$. Suppose we have a relation on $\mathcal{B}_1 \cup \cdots \cup \mathcal{B}_r$:

$$\underbrace{a_{11} v_{11} + \cdots + a_{1m_1} v_{1m_1}}_{\text{vector in } \lambda_1\text{-eigenspace}} + \underbrace{a_{21} v_{11} + \cdots + a_{1m_2} v_{2m_2}}_{\text{vector in } \lambda_2\text{-eigenspace}} + \cdots + \underbrace{a_{r1} v_{r1} + \cdots + a_{1m_r} v_{1m_r}}_{\text{vector in } \lambda_r\text{-eigenspace}} = 0.$$

Let $w_i = a_{i1} v_{i1} + \cdots + a_{im_i} v_{im_i}$ be the vector in the λ_i-eigenspace indicated above. We can thus rewrite the relation as

$$w_1 + w_2 + \cdots + w_r = 0, \tag{2.8.4}$$

where each w_i is the eigenspace for λ_i. Now, because the λ_i are distinct, we know that any set of its eigenvectors must be linearly independent by Theorem 2.8. The relation (2.8.4) would appear to contradict Theorem 2.8! The only way out of this conundrum is that each w_i is the zero vector. This gives us the relations

$$w_i = a_{i1} v_{i1} + \cdots + a_{im_i} v_{im_i} = 0$$

for each $i = 1, \ldots r$. But this is now a relation on the elements of the basis $\mathcal{B}_i$! So it must be trivial, and we can conclude all the a_{ij} are zero. This means our original relation on $\mathcal{B}_1 \cup \cdots \cup \mathcal{B}_r$ is trivial as well. Theorem 2.7 is proved. $\square$

3. Finding Eigenvalues and Eigenvectors.

Consider an $n \times n$ matrix A, and the linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ defined by $T(\vec{x}) = A\vec{x}$. Suppose that λ is an eigenvalue. Then the corresponding eigenspace is

$$\begin{aligned} E_\lambda =& \{\vec{x} \in \mathbb{R}^n \,|\, T(\vec{x}) = \lambda \vec{x}\} \\ =& \{\vec{x} \in \mathbb{R}^n \,|\, A\vec{x} = \lambda \vec{x}\} \\ =& \{\vec{x} \in \mathbb{R}^n \,|\, A\vec{x} - \lambda \vec{x} = 0\} \\ =& \{\vec{x} \in \mathbb{R}^n \,|\, (A - \lambda I_n)\vec{x} = 0\} \\ =& \ker(A - \lambda I_n). \end{aligned}$$

This gives a useful way to compute eigenvectors, which we state as a theorem in the language of both matrices and transformations:

Theorem 3.1. (1) *Let A be an $n \times n$ matrix. The λ-eigenspace of A is* $\ker(A - \lambda I_n)$.

(2) *Let $V \xrightarrow{T} V$ be a linear transformation. The λ-eigenspace of T is* $\ker(T - \lambda)$*, where $T - \lambda$ denotes the transformation of V sending each v to $T(v) - \lambda v$.*

In practice, we often compute eigenvectors for T by choosing a basis $\mathcal{B}$, then working in $\mathcal{B}$-coordinates with the matrix $[T]_\mathcal{B}$. The matrix $[T]_\mathcal{B}$ has the same eigenvalues as T, as you should check. The eigenspaces of the matrix $[T]_\mathcal{B}$ will be the $\mathcal{B}$-coordinates of the corresponding eigenspaces of T.

This idea also leads to a useful way to find eigenvalues of a transformation.

Definition 3.2. Consider an $n \times n$ matrix A. The **characteristic polynomial** of A is the degree n polynomial (in the variable x)

$$\chi_A(x) = \det(A - xI_n).$$

Here, we think of the x as an indeterminate scalar—plugging in any scalar for x, the expression $(A - xI_n)$ is an $n \times n$ matrix, so it makes sense to compute its determinant. Thus we can think of the characteristic polynomial as a function which takes in a scalar (plugging in for x) and spits out another scalar (the determinant of a matrix).

The computation we began this section with proves the following theorem:

Theorem 3.3. *The eigenvalues of a matrix A are precisely the roots of its characteristic polynomial.*

Example 3.4. Let $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. Then the characteristic polynomial of A is the determinant of the matrix $\begin{bmatrix} -x & 1 \\ -1 & -x \end{bmatrix}$. Thus $\chi_A(x) = x^2 + 1$. For example, plugging in $x = 0$, the function $\chi_A(0) = \det \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = 1$. This characteristic polynomial has no (real) roots! And indeed, because the matrix A represents a 90 degree clockwise rotation, A has no real eigenvalues.

You should convince yourself that $\chi_A(x)$ really is a *polynomial* in x by checking it for a few small sized matrices and thinking about how it generalizes to the general case. The proof is a bit of pain to write down in detail (see Worksheet 24).

Remark 3.5. A somewhat annoying feature of the characteristic polynomial, as defined above, is that its leading coefficient is not necessarily 1. That is, the characteristic polynomial does *not* always have the form $x^n + a_1x^{n-1} + \cdots + a_n$ for some scalars a_i. However, we *can* say that its leading coefficient is $(-1)^n$, hence 1 if n is even and -1 if n is odd. To avoid this annoying quirk, some authors define the characteristic polynomial as $\det(xI_n - A)$ instead, which is the same as ours up to sign: $\det(xI_n - A) = (-1)^n \det(A - xI_n)$. In particular, it has the same roots as the characteristic polynomial.

Lemma 3.6. *Similar matrices have the same characteristic polynomial. That is, if A and B are $n \times n$ matrices for which there exists an $n \times n$ matrix S with*

$$B = S^{-1}AS,$$

then $\chi_A(x) = \chi_B(x)$.

Proof. Suppose $B = S^{-1}AS$. Recall that a scalar matrix xI_n commutes with every matrix: $MxI_n = xI_nM$ for any scalar x and any $n \times n$ matrix M. So compute:

$$\begin{aligned} \det(B - xI_n) = \det(S^{-1}AS - xI_n) &= \det(S^{-1}AS - S^{-1}\ S\ xI_n) \\ &= \det(S^{-1}AS - S^{-1}xI_nS) \\ &= \det\left[S^{-1}(A - xI_n)\ S\right] \\ &= \det S^{-1}\ \det(A - xI_n)\ \det S \\ &= \det(A - xI_n). \end{aligned}$$

□

Lemma 3.6 ensures that the following definition makes sense:

Definition 3.7. Let $V \xrightarrow{T} V$ be a linear transformation of a vector space V of finite dimension n. The **characteristic polynomial** of T is the degree n polynomial (in the variable x)

$$\chi_T(x) = \det(A - xI_n)$$

where $A = [T]_{\mathcal{A}}$ is the matrix of T in *any* basis $\mathcal{A}$ for V.

Theorem 3.8. *Let $V \xrightarrow{T} V$ be a linear transformation of a finite dimensional vector space. The eigenvalues of T are precisely the roots of the characteristic polynomial of T.*

Corollary 3.9. *A linear transformation of an n-dimensional space has at most n eigenvalues.*

Proof. The characteristic polynomial has degree n, so it can have at most n roots. That is, we can factor the characteristic polynomial as

$$\chi_T(x) = \pm(x - \lambda_1)^{m_1}(x - \lambda_2)^{m_2} \cdots (x - \lambda_t)^{m_t} g(x)$$

where g is some polynomial (such as $x^2 + 5$ or $(x^4 + 5)(x^8 + 9)$) which has *no roots*) and each m_i is a natural number. The eigenvalues in this case will be precisely $\lambda_1, \lambda_2, \cdots, \lambda_t$. Note that this is less than the degree of f which is $m_1 + m_2 + \cdots + m_t + \deg g \geq t$. □

Example 3.10. Suppose that $V \xrightarrow{T} V$ is diagonalizable. This means there exists a basis $\mathcal{B}$ (an eigenbasis) such that

$$[T]_{\mathcal{B}} = \begin{bmatrix} a_{11} & 0 & 0 & \dots & 0 \\ 0 & a_{22} & 0 & \dots & 0 \\ 0 & \vdots & \ddots & \vdots & 0 \\ 0 & 0 & \dots & a_{n-1n-1} & 0 \\ 0 & 0 & \dots & 0 & a_{nn} \end{bmatrix}.$$

The diagonal elements a_{ii} are the eigenvalues of T. Computing the characteristic polynomial, we see that

$$\begin{aligned} \chi_T(x) &= (a_{11} - x)(a_{22} - x) \cdots (a_{nn} - x) \\ &= (-1)^n (x - a_{11})(x - a_{22}) \cdots (x - a_{nn}), \end{aligned}$$

where the a_{ii} are the eigenvalues of T (possibly repeated multiple times).

Remark 3.11. The characteristic polynomial may or may not have any real roots at all! Example 3.8, however, shows that the characteristic polynomial of a *diagonalizable* transformation always has real roots; indeed, it factors into n linear factors.

Remember, however, that over the complex numbers, every polynomial factors completely into linear factors

$$a(x - \lambda_1)^{a_1}(x - \lambda_2)^{a_2} \cdots (x - \lambda_t)^{a_t}$$

where the roots $\lambda_i \in \mathbb{C}$. [The scalar a here is the leading coefficient of the polynomial, which in the case of the characteristic polynomial is ± 1.] Mathematicians (and anyone who uses linear algebra) will call the roots of the characteristic polynomial eigenvalues *even when they are complex.* To avoid confusion, we will call such roots **complex eigenvalues**. It turns out that if λ is a complex eigenvalue of a matrix A, then there is a column vector with *complex entries* $\vec{v}$ such that $A\vec{v} = \lambda\vec{v}$. More on this soon.

In case the details remain unclear, we formally prove the statements on how to find eigenspaces and eigenvalues in the case of a transformation of an abstract vector space, even though the method is exactly the same as in the matrix case:

Proof of Theorems 3.1 and 3.8. Fix a scalar c. Consider the linear transformation[3]

$$V \xrightarrow{\Phi} V \quad v \mapsto T(v) - c\, v.$$

Note that c is an eigenvalue if and only if this transformation Φ has a non-zero kernel. In this case, the kernel is precisely the c-eigenspace. Since the source and target of T have the same dimension, Φ has a non-zero kernel if and only if it is *not* invertible (using Rank-Nullity). In turn, this happens if and only if the determinant of Φ is **zero**.

We can compute the determinant of Φ by computing the determinant of its $\mathcal{B}$-matrix for *any basis* $\mathcal{B}$. First notice that for all $v \in V$,

$$\Phi(v) = T(v) - c\, v = T(v) - cI_V(v) = (T - cI_V)(v)$$

where I_V denotes the identity transformation of V. Now fix some basis $\mathcal{B}$ for V. We compute the $\mathcal{B}$-matrix for Φ as follows:

$$[\Phi]_{\mathcal{B}} = [T - cI_V]_{\mathcal{B}} = [T]_{\mathcal{B}} - [cI_V]_{\mathcal{B}}.$$

[3]As a good reader of mathematics, you realize you should check yourself that the given map is really a linear transformation

If we let A be the matrix $[T]_{\mathcal{B}}$ and observe that the matrix of cI_V in *any basis* is cI_n, this shows that

$$[\Phi]_{\mathcal{B}} = [T]_{\mathcal{B}} - [cI_V]_{\mathcal{B}} = A - cI_n.$$

Its determinant is precisely the result of plugging c into the expression $\det(A - xI_n)$. Of course, by definition $\det(A - xI_n)$ is the characteristic polynomial $\chi_T(x)$. This means that c is an eigenvalue if and only if c is a root of the characteristic polynomial. $\square$

Example 3.12. Let us find an eigenbasis for the transformation $\mathbb{R}^2 \xrightarrow{T_A} \mathbb{R}^2$ given by multiplication by $A = \begin{bmatrix} 5 & -7 \\ -6 & 6 \end{bmatrix}$. We can then find a diagonal matrix similar to A and a *diagonalization* of A.

Step 1: Find the eigenvalues. The characteristic polynomial is

$$\det[A - xI_2] = \det \begin{bmatrix} 5-x & -7 \\ -6 & 6-x \end{bmatrix} = (x-5)(x-6) - 42 = x^2 - 11x - 12 = (x+1)(x-12).$$

This tells us the eigenvalues are 12 and -1. We will find their eigenspaces one at a time.

Step 2: For each eigenvalue, compute the eigenspace using Theorem **??**. Make note of its dimension.

Here, we first find 12-eigenspace. This is the kernel of $A - 12I_2 = \begin{bmatrix} 6 & -7 \\ -6 & 7 \end{bmatrix}$. This matrix has rank 1, so by the Rank-nullity theorem, the kernel is one dimensional. The kernel contains $\begin{bmatrix} 7 \\ 6 \end{bmatrix}$, so this 12-eigenvector must be a basis for the 12-eigenspace.

Next, we find the -1-eigenspace. This is the kernel of $A + I_2 = \begin{bmatrix} -7 & -7 \\ -6 & -6 \end{bmatrix}$, which is clearly spanned by $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

Step 3: Find bases for each eigenspace. See if putting them together gives you a basis for V. In this small case, this is easy: the vectors $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and $\begin{bmatrix} 7 \\ 6 \end{bmatrix}$ span the two-dimensional space $\mathbb{R}^2$ and both are eigenvectors. So $(\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 7 \\ 6 \end{bmatrix})$ is an eigenbasis for T_A.

Since T_A has an eigenbasis $\mathcal{B}$, we know the $\mathcal{B}$-matrix of T_A will be diagonal. It is easy to find:

$$[T]_{\mathcal{B}} = \begin{bmatrix} -1 & 0 \\ 0 & 12 \end{bmatrix}.$$

Finally, to **diagonalize** A, we want to explicity exhibit A as similar to a diagonal matrix. The usual change of basis formula for matrices in different basis tells us thats

$$[T]_{\mathcal{B}} = S_{\mathcal{E}\to\mathcal{B}}[T]_{\mathcal{E}} S_{\mathcal{B}\to\mathcal{E}}.$$

But we compute $[T]_{\mathcal{E}} = A$ and $S_{\mathcal{B}\to\mathcal{E}} = \begin{bmatrix} 1 & 7 \\ -1 & 6 \end{bmatrix}$. So we have

$$\begin{bmatrix} -1 & 0 \\ 0 & 12 \end{bmatrix} = S^{-1}AS$$

for $S = \begin{bmatrix} 1 & 7 \\ -1 & 6 \end{bmatrix}$.

4. Multiplicities of Eigenvalues

Definition 4.1. Let $V \xrightarrow{T} V$ be a linear transformation of a finite dimensional vector space V. The **algebraic multiplicity** of an eigenvalue λ is the largest power r such that $(x-\lambda)^r$ divides the characteristic polynomial.

The algebraic multiplicity should not be confused with the geometric multilplicity, whose definition we repeat here:

Definition 4.2. The **geometric multiplicity** of λ is the dimension of the λ-eigenspace, or equivalently, the maximal number of linearly independent eigenvectors with eigenvalue λ.

Theorem 4.3. *Let $V \xrightarrow{T} V$ be a linear transformation of a finite dimensional vector space. Then for each eigenvalue λ,*

$$\textit{geometric multiplicity of } \lambda \leq \textit{algebraic multiplicity of } \lambda.$$

Proof. Suppose that m is the geometric multiplicity of λ. Let $v_1, \ldots, v_m$ be a basis for V_λ, the λ-Eigenspace of T. We can extend $v_1, \ldots, v_m$ to a full basis $\mathfrak{B} = (v_1, \ldots, v_m, v_{m+1}, \ldots, v_n$ for V. The $\mathfrak{B}$-matrix of T has the form

$$[T]_{\mathfrak{B}} = \begin{bmatrix} \lambda & 0 & 0 & \ldots & a_{1\,m+1} & a_{1\,m+2} & \ldots & a_{1\,n} \\ 0 & \lambda & 0 & \ldots & a_{2\,m+1} & a_{2\,m+2} & \ldots & a_{2\,n} \\ \vdots & & & \vdots & a_{2\,m+1} & a_{2\,m+2} & \ldots & a_{2\,n} \\ 0 & 0 & \ldots & \lambda & a_{m\,m+1} & a_{m\,m+2} & \ldots & a_{m\,n} \\ 0 & 0 & \ldots & 0 & a_{m+1\,m+1} & a_{m+1\,m+2} & \ldots & a_{m+1\,n} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & \ldots & 0 & a_{n\,m+1} & a_{n\,m+2} & \ldots & a_{n\,n} \end{bmatrix}$$

Computing the characteristic polynomial $\det(A - xI_n)$ by Laplace expansion along the first column, we have

$$\det\left([T]_{\mathfrak{B}} - xI_n\right) = (\lambda - x)\det \begin{bmatrix} \lambda - x & 0 & 0 & \ldots & a_{2\,m+1} & a_{2\,m+2} & \ldots & a_{2\,n} \\ 0 & \lambda - x & 0 & \ldots & a_{3\,m+1} & a_{3\,m+2} & \ldots & a_{3\,n} \\ \vdots & & & \vdots & a_{2\,m+1} & a_{2\,m+2} & \ldots & a_{2\,n} \\ 0 & & \ldots & \lambda - x & a_{m\,m+1} & a_{m\,m+2} & \ldots & a_{m\,n} \\ 0 & 0 & \ldots & 0 & a_{m+1\,m+1} - x & a_{m+1\,m+2} & \ldots & a_{m+1\,n} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & \ldots & 0 & a_{n\,m+1} & a_{n\,m+2} & \ldots & a_{n\,n} - x \end{bmatrix}$$

and continuing to compute along the next columns, iteratively, we eventually get

$$\det\left([T]_{\mathfrak{B}} - xI_n\right) = (\lambda - x)^m \det \begin{bmatrix} a_{m+1\,m+1} - x & a_{m+1\,m+2} & \ldots & a_{m+1\,n} \\ a_{m+1\,m+2} & a_{m+2\,m+2} - x & \ldots & a_{m+1\,n} \\ \vdots & \ddots & \vdots & \vdots \\ a_{n\,m+1} & a_{n\,m+2} & \ldots & a_{n\,n} - x \end{bmatrix}$$

We see that this will produce a polynomial divisible by $(x-\lambda)^m$. So the algebraic multiplicity of λ is at least m. $\square$

Example 4.4. Consider the three different linear transformations of $\mathbb{R}^3 \to \mathbb{R}^3$ given by the matrices

$$A_1 = \begin{bmatrix} \pi & 0 & 0 \\ 0 & \pi & 0 \\ 0 & 0 & \pi \end{bmatrix}, \qquad A_2 = \begin{bmatrix} \pi & 1 & 0 \\ 0 & \pi & 0 \\ 0 & 0 & \pi \end{bmatrix}, \qquad A_3 = \begin{bmatrix} \pi & 1 & 0 \\ 0 & \pi & 1 \\ 0 & 0 & \pi \end{bmatrix}.$$

In all three cases, the characteristic polynomial is $(x - \pi)^3$. So π is the only eigenvalue, and its algebraic multiplicity is 3. The geometric multiplicities of π can be computed by examining the kernels of the matrices $(A_i - \pi I_3)$. These are, respectively,

$$A_1 - \pi I_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad A_2 - \pi I_3 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad A_3 - \pi I_3 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

We conclude that the geometric multiplicity of π is three for A_1, two for A_2 and one for A_3 (remember rank-nullity!). These multiplicities represent the maximal number of linearly independent eigenvectors (of the eigenvalue π, but again, π is the only eigenvalue). Thus only A_1 is diagonalizable because it is the only matrix for which there exists an eigenbasis.

Theorem 4.5. *Let V be a vector space of dimension n. Let $V \xrightarrow{T} V$ be a linear transformation. The following are equivalent:*

(1) *T is diagonalizable;*
(2) *T has an eigenbasis;*
(3) *The sum of the geometric multiplicities of the (real) eigenvalues of T is n.*
(4) *All eigenvalues are real and the geometric multiplicity of each equals its algebraic multiplicity.*

Corollary 4.6. *Let $V \xrightarrow{T} V$ be a linear transformation of an n-dimensional vector space. If T has n distinct eigenvalues, then T is is diagonalizable.*

Example 4.7. The converse to Corollary 4.6 is false: consider the map of $\mathbb{R}^2$ that scales every vector by 5.

5. The Spectral Theorem

Given an $n \times n$ matrix A, is there a way to tell at a glance whether or not it is diagonalizable? Sometimes it's easy. For example, any *diagonal* matrix is diagonalizable (what is an eigenbasis?). And a matrix such as

$$\begin{bmatrix} 1 & \sqrt{17} & \pi & \frac{3}{4} \\ 0 & \sqrt{13} & \frac{7}{4} & e \\ 0 & 0 & \frac{\pi}{4} & -7 \\ 0 & 0 & 0 & \sqrt[3]{7} \end{bmatrix}$$

is easily seen to be diagonalizable because it has four distinct eigenvalues $1, \sqrt{13}, \frac{\pi}{4}, \sqrt[3]{7}$ each of algebraic (and geometric) multiplicity one. But what about the matrices

$$\begin{bmatrix} 1 & \sqrt{17} & \pi & \frac{1}{4} \\ 5 & \sqrt{13} & \frac{11}{4} & e \\ -7 & 6 & \frac{\pi}{9} & -7 \\ 0 & \frac{2}{9} & -33 & \sqrt[7]{7} \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 1 & \sqrt{17} & \pi & \frac{1}{4} \\ \sqrt{17} & \sqrt{13} & 6 & \frac{2}{9} \\ \pi & 6 & \frac{\pi}{9} & -7 \\ \frac{1}{4} & \frac{2}{9} & -7 & \sqrt[7]{7} \end{bmatrix} ??$$

The Spectral Theorem is an amazing theorem that tells us, among other things, that the second matrix is diagonalizable because it is *symmetric*. Here is a simple corollary of the Spectral Theorem:

Theorem 5.1. *A symmetric matrix is diagonalizable. That is, a symmetric matrix has an eigenbasis.*

Put differently, Theorem 5.1 tells us that every symmetric matrix A is similar to a diagonal matrix D. The diagonal matrix D will have the eigenvalues of A along its diagonal, where each eigenvalue is repeated as many times as its multiplicity. By multiplicity here, we mean either algebraic or geometric multiplicity: they are the same for each eigenvalue of a diagonalizable transformation by Theorem 4.5. So a symmetric matrix, like any diagonalizable matrix, has the property that *all roots of its characteristic polynomial are real.* That is, *all eigenvalues of a symmetric matrix are real.*

Theorem 5.1 is really just a simple corollary of the glorious Spectral theorem, which give us a necessary and sufficient condition for matrices to have *especially nice* eigenbases:

Theorem 5.2. The Spectral Theorem.
Let A be an $n \times n$ symmetric matrix, then A has an orthonormal eigenbasis.
Conversely, any matrix with an orthonormal eigenbasis must be symmetric.

This means that there is an *orthonormal basis* $\mathcal{U} = (\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n)$ for $\mathbb{R}^n$ such that, writing the transformation T_A in the basis $\mathcal{U}$ we have

$$[T_A]_{\mathcal{U}} = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & \cdots & & \lambda_n \end{bmatrix}$$

where the elements λ_i are the eigenvalues of the transformation T_A. So the transformation T_A pulls the space $\mathbb{R}^n$ by λ_i in the direction $\vec{u}_i$: it would take a unit hypercube formed by $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n$ to an n-dimensional rectangular hypersolid (or **right** parallelopiped) with side lengths $\lambda_1, \ldots, \lambda_n$.

Another way to state the Spectral Theorem:

Theorem 5.3. The Spectral Theorem II.
An $n \times n$ matrix A is symmetric if and only if its eigenspaces are orthogonal and their dimensions sum to n.

Example 5.4. Let $\mathbb{R}^2 \xrightarrow{T_A} \mathbb{R}^2$ by the transformation given by multiplication by $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$. We can find an eigenbasis for this transformation as we have done before: by computing the characteristic polynomial, finding its roots, then computing the eigenspace. Or, in this very nice case, it is possible to find an eigenbasis by inspection: since

$$A \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix} \quad \text{and} \quad A \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 2 \end{bmatrix},$$

we see that

$$\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right)$$

is an eigenbasis with corresponding eigenvalues $4, 2$. Note that these eigenvectors are orthogonal. We can scale them to find an orthonormal eigenbasis

$$\left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right)$$

for the matrix A. The fact that A has an orthonormal eigenbasis is *equivalent* to the fact that A is symmetric—this is exactly what the Spectral Theorem says!

What is so great about an orthonormal eigenbasis? We have seen in Chapter 5 that orthonormal bases are especially nice. Algebraically, it is easy to find coordinates of a vector $\vec{v} \in \mathbb{R}^n$ in an orthonormal basis $\mathcal{U} = (\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n)$: recall that

$$[\vec{v}]_{\mathcal{B}} = \begin{bmatrix} \vec{v} \cdot \vec{u}_1 \\ \vec{v} \cdot \vec{u}_2 \\ \vdots \\ \vec{v} \cdot \vec{u}_n \end{bmatrix}.$$

Geometrically, orthonormal bases are nice because they produce coordinate systems which have the traditional "grid paper" look: the basis elements are all length one and perpendicular to each other. Now, if $\mathcal{U}$ is also an eigenbasis for some transformation T_A, then the "unit hypercube" in this coordinate system will be pulled into a hyper-rectangle by the transformation, so the transformation is especially nice to visualize.

We can write the Spectral Theorem in the following equivalent way involving matrices:

Corollary 5.5. *Let A be an $n \times n$ matrix. Then A is symmetric if and only if there exists an orthogonal matrix P such that $P^\top AP$ is diagonal.*

Proof. Suppose A is symmetric. Let $\mathcal{U} = (\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n)$ be an orthonormal eigenbasis. We have

$$A = [T_A]_{\mathcal{E}} = S_{\mathcal{U}\to\mathcal{E}}\ [T_A]_{\mathcal{U}}\ S_{\mathcal{E}\to\mathcal{U}} = S\ D\ S^{-1}$$

where $S = S_{\mathcal{U}\to\mathcal{E}} = [\vec{u}_1\ \vec{u}_2\ \ldots\ \vec{u}_n]$ and $D = [T_A]_{\mathcal{U}}$ is the diagonal matrix with the eigenvalue λ_i of $\vec{u}_i$ in position ii. Multiplying by S and S^{-1} appropriately, this can be rewritten as $D = S^{-1}\ A\ S$. Finally, since the columns of S are orthonormal, the matrix S is orthogonal, so that $S^{-1} = S^\top$. Taking P to be S, we see that $P^\top AP$ is diagonal, as needed.

Conversely, say $P^\top AP$ is diagonal, where P is some orthogonal matrix. Equivalently, this means $P^{-1}AP = D$ is diagonal. From here, it is easy to see that the columns of P are an orthonormal eigenbasis. We already know they are a basis, since P is invertible, and we only need to check that each column of P is an eigenvector. The j-th column of P can be written $P\vec{e}_j$ (the Unusually Useful Lemma). So we need to check that $A\ P\vec{e}_j$ is a scalar multiple of $P\vec{e}_j$. We compute: since $AP = PD$,

$$A\ P\vec{e}_j = PD\vec{e}_j = P\lambda_j\vec{e}_j = \lambda_j P\vec{e}_j.$$

Here the second equality again uses the Unusually Useful Lemma, applied to the diagonal matrix D. The result of this computation is that the j-th column of P is an eigenvector with eigenvalue λ_j (the jj-th entry of D). QED. □

The proof of the Spectral Theorem is a bit involved. See Worksheet 27. You do not need to know this proof for the Final exam in Math 217.