Introduction to diagonalization: how a little transformation $T$ that thought of itself as ugly came to see that it was really very beautiful.

Suppose $D$ is an $n \times n$ diagonal matrix with diagonal entries $\lambda_1, \lambda_2, \ldots, \lambda_n$. Let $T(x) = Dx$, so $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a linear transformation. Such transformations are very easy to understand and work with. A quick illustration about why they're useful shows up later in these notes.

First look at what $T$ does to the standard basis vectors $e_1, \ldots, e_n$:

$$T(e_i) = De_i = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \lambda_n \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \lambda_1 e_i \\ \vdots \\ \lambda_i e_i \\ \vdots \\ \lambda_n e_i \end{bmatrix}$$

Multiplying $e_i$ by $D$ just rescales $e_i$, multiplying by the scalar $\lambda_i$ from the diagonal of $D$.

Then, each $x \in \mathbb{R}^n$ can be written $x = x_1 e_1 + \ldots + x_n e_n$, so

$$T(x) = Dx = D(x_1 e_1 + \ldots + x_n e_n) = x_1 D(e_1) + \ldots + x_n D(e_n)$$

For example, look $T(x) = Dx = \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & -4 \end{bmatrix} x$. Writing out the formula in terms of $x_1, x_2, x$ shows the rescaling: $T(x_1, x_2, x_3) = (2x_1, \frac{1}{2}x_2, -4x_3)$. \ldots
Example Suppose $T : \mathbb{R}^2 \to \mathbb{R}^2$ is given by $T(x) = Dx = \begin{bmatrix} 3 & 0 \\ 0 & 6 \end{bmatrix} x$. Then

$$D e_1 = \begin{bmatrix} 3 & 0 \\ 0 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix} = 3 e_1, \quad D e_2 = \begin{bmatrix} 3 & 0 \\ 0 & 6 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \end{bmatrix} = 6 e_2,$$

and

$$D x = \begin{bmatrix} 3 & 0 \\ 0 & 6 \end{bmatrix} x = \begin{bmatrix} 3 & 0 \\ 0 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 x_1 \\ 6 x_2 \end{bmatrix}.$$ Multiplication by $D$ simply rescales the coordinates of $x$ by a factor of 3 along the coordinate axis determined by $e_1$ (horizontal) and by a factor of 6 in the $e_2$ direction (vertical). In this particular example, the diagonal entries of $D$ are both bigger than 1, so multiplying $x$ by $D$ “stretches” the coordinates of $x$ in each of the coordinate directions, $e_1$ and $e_2$. You can see this in the figure below where

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} \mapsto D \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$$

Since $D$ is a diagonal matrix, the image of a point in $\text{Span}\{e_1\}$ ($= \text{the } x\text{-axis}$) is another point in $\text{Span}\{e_1\} : T$ maps $\text{Span}\{e_1\}$ into itself. The same is true for $\text{Span}\{e_2\}$ ($= \text{the } y\text{ axis}$).

When a linear transformation $T$ maps a subspace $H$ into itself ($T(H) \subseteq H$), we call $H$ an invariant subspace for $T$. For $T(x) = Dx$, each coordinate axis is an invariant subspace.

How would things change (slightly) if $D = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & -6 \end{bmatrix}$?
Example Let \( T(x) = Dx = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} x \) and consider the difference equation

\[
x_{k+1} = Dx_k
\]

The long term behavior of the \( x_k \)'s can be very important in applications, and it depends completely on the diagonal entries of \( D \).

For an initial value \( x_0 = \begin{bmatrix} a \\ b \end{bmatrix} \), we get \( x_1 = D \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 \cdot a \\ \frac{1}{2} \cdot b \end{bmatrix} \), \( x_2 = Dx_1 = D \begin{bmatrix} 1 \cdot a \\ \frac{1}{2} \cdot b \end{bmatrix} = \begin{bmatrix} a \\ (\frac{1}{2})^2 \cdot b \end{bmatrix} \), and in general, \( \ldots x_k = \begin{bmatrix} 1^k \cdot a \\ (\frac{1}{2})^k \cdot b \end{bmatrix} = \begin{bmatrix} a \\ (\frac{1}{2})^k \cdot b \end{bmatrix} \).

Since \( (\frac{1}{2})^k \to 0 \) as \( k \to \infty \), we immediately see that \( x_k \to \begin{bmatrix} a \\ 0 \end{bmatrix} \) as \( k \to \infty \).

More generally: if \( D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \) and \( x_0 = \begin{bmatrix} a \\ b \end{bmatrix} \), we get \( x_k = \begin{bmatrix} \lambda_1^k \cdot a \\ \lambda_2^k \cdot b \end{bmatrix} \).

**Note** If \( A \) is not diagonal, then sometimes we can pick a new basis \( \mathcal{B} \) and set up a new coordinate system in which the transformation \( T \) is just a rescaling along the new coordinate axes. This process (when we can do it) is called “diagonalization.”

Now suppose that \( T : \mathbb{R}^n \to \mathbb{R}^n \), where \( T(x) = Ax \).

(**) If \( A \) is not diagonal, then sometimes we can pick a new basis \( \mathcal{B} \) and set up a new coordinate system in which the transformation \( T \) is just a rescaling along the new coordinate axes. This process (when we can do it) is called “diagonalization.”

Imagine this: suppose \( Q \) is a geometric point in \( \mathbb{R}^n \). No coordinate system given, but we have some geometric description of a mapping \( T \): geometrically, we can see that \( Q \xrightarrow{T} S \) = some other point in \( \mathbb{R}^n \). Now introduce the standard coordinate system with basis \( \mathcal{E} = \{ e_1, \ldots, e_n \} \). The point \( Q \) now get names using standard coordinates: \( Q \) is the point \( x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \). If \( T \) is a linear mapping, then we know from Chapter 1 that if we figure out (geometrically) the coordinates for each point \( T(e_1), \ldots, T(e_n) \), then we can write \( T \) using a matrix formula: \( T(x) = Ax \), where the columns of \( A \) are the standard coordinate vectors \( T(e_1), \ldots, T(e_n) \). In a diagram, \( T \) is described as the matrix mapping

\[
x \mapsto Ax
\]

If we choose a different basis \( \mathcal{B} = \{ b_1, \ldots, b_n \} \), then \( Q \) and \( S \) get new names in the \( \mathcal{B} \)-coordinate system: \( Q \) now has coordinates \( [x]_\mathcal{B} \). Geometrically, this does not change the mapping \( T \) at all: \( Q \xrightarrow{T} S \) is still true; what changes are the coordinate names for the points \( Q \) and \( S \). Since the matrix \( A \) depended on \( e_1, \ldots, e_n \), it shouldn't be surprising that a change of
basis changes the matrix that describes the same mapping \( T \) (which, geometrically, hasn’t changed at all). The original message, at (**) is that sometimes we can find \( B \) so that the new matrix is a diagonal matrix, \( D \).

The change of coordinates matrix is \( P_B = \begin{bmatrix} b_1 & b_2 & \ldots & b_n \end{bmatrix} \) and \[
\begin{align*}
P_B |x|_B &= x \\
P_B^{-1} x &= |x|_B
\end{align*}
\]

Then \( T \) can be described by two different matrix multiplications: multiplying by \( A \) when working in standard coordinates, and \( D \) (a diagonal matrix) when written in \( B \)

\[
\begin{bmatrix} x \\
P_B^{-1} \\
[x]_B \\
Q \rightarrow S \end{bmatrix} \rightarrow \begin{bmatrix} Ax \\
\uparrow P_B \\
D[x]_B \end{bmatrix}
\]

The multiplications by \( P_B \) and \( P_B^{-1} \) indicated in the diagram are just to go back and forth between the coordinate systems.

Assuming that we can find such a basis \( B \) so that, in the new coordinate system the matrix for the transformation is diagonal matrix \( D \). Look at the different transformations in the diagram above.

1) first convert \( x \) to \( B \) coordinates (compute \( P_B^{-1} x \) to get \( [x]_B \))
2) then perform the transformation = rescaling the \( B \)-coordinates (multiply \( DP_B^{-1} x \) to get \( D[x]_B \))
3) then convert back to standard coordinates to arrive at the original result \( Ax \) (compute \( P_SD[PB^{-1}]x = Ax \))

then we get the same result as \( Ax \). In other words, \( Ax = P_BDP_B^{-1}x \). (***)

This motivates an official definition:

Suppose \( A \) is an \( n \times n \) matrix. We say that \( A \) is diagonalizable if there is an invertible \( n \times n \) matrix \( P \) and a diagonal matrix \( D \) such that

\[
A = PDP^{-1}
\]

When this happens, the columns of \( P \) can be used as a new basis \( B \) for \( \mathbb{R}^n \). Then \( P = P_B \) is the “change of coordinates matrix” so that this formula is exactly as presented in (***).
When this can be done, we could think of it as saying that the matrix $A$ was not diagonal originally because we started out working in the “wrong” coordinate system: the “ugly” transformation is beautiful in a different coordinate system.

It is not always possible to diagonalize an $n \times n$ matrix $A$, but the examples and discussion below will give us some insight about when we can do so. There will be more details throughout Chapter 5.

**Example** Suppose $T(x) = Ax$, where $A = \begin{bmatrix} 4 & 2 \\ 1 & 5 \end{bmatrix}$. It’s a fact that

$$A = \begin{bmatrix} 4 & 2 \\ 1 & 5 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 6 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix} \quad (****)$$

$P$ is invertible. Therefore, by the Invertible Matrix Theorem, its columns $b_1$, $b_2$ are linearly independent and span $\mathbb{R}^2$, so we can use $B = \{b_1, b_2\} = \left\{ \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$ as a new basis for $\mathbb{R}^2$. Then $P = P_B = \text{the change of coordinates matrix } P_B$ :

$$P_B[x]_B = x$$

For $x = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, we can compute $A \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 10 \\ 7 \end{bmatrix}$.

But using the factorization in (****), we can calculate $Ax$ another way – longer, but giving us more insight:

$$A \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 6 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 10 \\ 7 \end{bmatrix}$$

Look carefully at what happens to $x$, step-by-step as each matrix in (****) does its job:

$$x = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \xrightarrow{\text{multiplication by } P_B^{-1}} \begin{bmatrix} \frac{1}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ \frac{4}{3} \end{bmatrix}.$$  Then  $$\begin{bmatrix} 3 & 0 \\ 0 & 6 \end{bmatrix} \begin{bmatrix} \frac{1}{3} \\ \frac{4}{3} \end{bmatrix} = \begin{bmatrix} 1 \\ 8 \end{bmatrix}$$

$P_B^{-1}$ converts standard coordinates into $B$-coordinates by factors of 3 and 6 (in the coordinate directions corresponding to $b_1$ and $b_2$)  (cont. → )
Finally,
\[
\begin{bmatrix}
2 & 1 \\
-1 & 1
\end{bmatrix}
\begin{bmatrix}
1 \\
8
\end{bmatrix}
= 
\begin{bmatrix}
10 \\
7
\end{bmatrix}
\]

\[ \text{multiplication by } P_B \]
converts the (stretched) \(B\)-coordinates back into
standard coordinates

Locate points and follow the comments below in the picture:

1) \( \mathbf{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \): the same geometric point, named in \(B\)-coordinates, is \([\mathbf{x}]_B = \begin{bmatrix} \frac{1}{3} \\ \frac{4}{3} \end{bmatrix}\)

2) Rescaling \([\mathbf{x}]_B\) by the factors 3 and 6 in the \(b_1\) and \(b_2\) directions
gives a new point whose name in \(B\)-coordinates is \(\begin{bmatrix} 1 \\ 8 \end{bmatrix}\).

3) This new point has a different name in standard coordinates. If we convert the
\(B\)-coordinates \(\begin{bmatrix} 1 \\ 8 \end{bmatrix}\) back to standard coordinates, we get \(\begin{bmatrix} 10 \\ 7 \end{bmatrix} = A \begin{bmatrix} 2 \\ 1 \end{bmatrix}\).

Thus
\[
\begin{align*}
\begin{bmatrix} 2 \\ 1 \end{bmatrix} & \rightarrow A \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 10 \\ 7 \end{bmatrix} & \text{expressed in standard coordinate} \\
\begin{bmatrix} \frac{1}{3} \\ \frac{4}{3} \end{bmatrix} & \rightarrow D \begin{bmatrix} \frac{1}{3} \\ \frac{4}{3} \end{bmatrix} = \begin{bmatrix} 1 \\ 8 \end{bmatrix} & \text{expressed in } B\text{-coordinates}
\end{align*}
\]
**Example (continued)** In the example, notice what multiplication by $A$ does to the new basis vectors $b_1$ and $b_2$: this observation is important because it reveals (at least in theory) how a basis $B$ was chosen to diagonalize $A$. What are $Ab_1$ and $Ab_2$? We could just multiply out $Ab_1$ but computation using the factorization in (****) gives us more insight:

$$Ab_1 = \begin{bmatrix} 4 & 2 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 6 \end{bmatrix} \begin{bmatrix} 1/3 \\ -1/3 \end{bmatrix} \begin{bmatrix} -2 \\ 2 \end{bmatrix}$$

$\uparrow$ (visualize the effect of each multiplication in the preceding figure)

$$= \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ -3 \end{bmatrix} = 3 \begin{bmatrix} 2 \\ -1 \end{bmatrix} = 3b_1$$

these are the $B$-coordinates $\uparrow$ rescaled $\uparrow$ returned to $B$-coordinates$\uparrow$ standard to coordinates

Verify similarly that $Ab_2 = 6b_2$

So: multiplication by $A$ sends each vector $b_1$ and $b_2$ to a scalar multiple of itself.

Nonzero vectors with this property are called **eigenvectors** of $A$ and the scalars are called **eigenvalues**. The scalar 3 is the eigenvalue associated with the eigenvector $b_1$, and 6 is the eigenvalue associated with the eigenvector $b_2$.

The reason we were able to do such a nice analysis of how $x \mapsto Ax$ works is that we had a new very special basis $B$ for $\mathbb{R}^2$ — namely, a basis $B$ whose members are eigenvectors of matrix $A$.

*Remember, in the beginning, I simply gave you the matrix $P$, and its columns $b_1$ and $b_2$ turned out to be eigenvectors of $A$. We haven’t discussed yet how you might start with $A$ and try to find such a basis — or even know whether $A$ has any eigenvectors! More on this throughout Chapter 5.*

Here is the general definition.

**Definition** A nonzero vector $x$ is called an **eigenvector** of the $n \times n$ matrix $A$ if $Ax = \lambda x$ for some scalar $\lambda$. The scalar $\lambda$ is called an **eigenvalue** of $A$ (associated with the eigenvector $x$). Eigenvalues and eigenvectors of $A$ are also called eigenvectors or eigenvalues of the transformation $T$ where $T(x) = Ax$.

*Eigenvalue and eigenvector* are words with German roots. Some books call these “characteristic vectors” and “characteristic values” (loose translation). It’s traditional to use the Greek letter $\lambda$ (“lambda”) to denote an eigenvalue. Note that an eigenvalue $\lambda = 0$ might occur; but the vector $0$ is not considered as an eigenvector. For example, if $x \neq 0$ is in $\text{Nul } A$, then $x$ is an eigenvector with eigenvalue $0$, since $Ax = 0 = 0 \cdot x$. 
There is nothing special about the eigenvalues 3 and 6 in the example. The diagonal entries in $D$ might have been some other numbers $\lambda_1$ and $\lambda_2$ (perhaps 0; perhaps with $\lambda_1 = \lambda_2$). To summarize for the $2 \times 2$ case:

Suppose $A$ can be factored as $A = P \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} P^{-1}$, where $P$ has columns $b_1, b_2$.

We can use $B = \{b_1, b_2\}$ as a new basis (why?) for coordinates in $\mathbb{R}^2$ and the change of coordinates matrix is $P_B = P$.

A simple calculation shows that $b_1$ and $b_2$ must be eigenvectors of $A$ with eigenvalues $\lambda_1$ and $\lambda_2$. For example, for $b_1$:

$$Ab_1 = P_B \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} P_B^{-1} b_1 = P_B \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

the $B$-coordinates of $b_1$

$$= P_B \begin{bmatrix} \lambda_1 \\ 0 \end{bmatrix} = [b_1 \ b_2] \begin{bmatrix} \lambda_1 \\ 0 \end{bmatrix} = \lambda_1 b_1 + 0 b_2 = \lambda_1 b_1$$

in $B$-coordinates

Similarly $Ab_2 = \lambda_2 b_2$.

With $\lambda_1$ and $\lambda_2$ replacing 3 and 6, the geometric interpretation of how $A$ operates on a point $\mathbf{x}$ is just as it was earlier: $\mathbf{x} \mapsto A\mathbf{x}$ is the same as a rescaling of $B$ coordinates, with rescaling factors $\lambda_1$ and $\lambda_2$. Of course, if (say) $0 < \lambda_1 < 1$ then multiplication by $D$ is a “contraction” in the first $B$-coordinate of $\mathbf{x}$ (rather than a“stretch”); and if (say) $\lambda_2 < 0$, multiplication by $D$ reverses the sign of the second $B$-coordinate of $\mathbf{x}$ (as well as either stretching or contracting).

The preceding discussion proves the following theorem for $2 \times 2$ matrices; moreover, the discussion is perfectly parallel for $n \times n$ matrices. (For $n \times n$, there are more matrix entries and vectors to write down, and this makes the discussion look more complicated – but it isn’t!)

**Theorem 1** Let $A$ be an $n \times n$ matrix and suppose $A$ is diagonalizable, that is, suppose $A$ can be factored as

$$A = P \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} P^{-1},$$

Conclusion: then $\mathbb{R}^n$ has a basis $\mathcal{B}$ that consists of eigenvectors of $A$.

In fact, this basis is $\mathcal{B} = \{b_1, b_2, \ldots, b_n\}$, where the $b_i$’s are the columns of $P$. The $b_i$’s are eigenvectors of $A$, and their eigenvalues are the scalars on the diagonal of $D$ (in the same order: $\lambda_1$ for $b_1$, etc.).
It turns out that the converse of Theorem 1 is also true:

if we can somehow find a basis for $\mathbb{R}^n$ that consists of eigenvectors of $A$, then we can conclude $A$ is diagonalizable — that is, $A$ can be factored as $A = PD P^{-1}$.

The next example illustrates why this is true using a $2 \times 2$ matrix.

**Example** Suppose $A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}$. Then $b_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and $b_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$ are eigenvectors for $A$, as you can verify by calculating

$$A \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -5 \\ -5 \end{bmatrix} = 5 \begin{bmatrix} 1 \\ -1 \end{bmatrix} \text{ and}$$

$$A \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 3 \\ -6 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

The corresponding eigenvalues are 5 and 3, respectively.

*Don’t worry, for now, about how the eigenvectors were found. More detail throughout Chapter 5.*

$B = \{b_1, b_2\}$ is a basis for $\mathbb{R}^2$. Define

i) $P = [b_1 \ b_2] = \begin{bmatrix} 1 \\ -1 \\ -1 \\ -2 \end{bmatrix}$ (invertible because columns are linearly independent) and

ii) $D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$ (diagonal entries in $D$ are the eigenvalues of $b_1, b_2$ in same order as the columns of $P$: not $\begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix}!!$)

Then $A = PDP^{-1}$: to check that this factorization is true, the easiest thing to do (rather than find $P^{-1}$) is to multiply both sides on the right by $P$: $A = PDP^{-1}$ is equivalent to the equation $AP = PD$ is true, and we can easily check that:

$$AP = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} = \begin{bmatrix} 5 & 3 \\ -5 & -6 \end{bmatrix} \text{ and}$$

$$PD = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 5 & 3 \\ -5 & -6 \end{bmatrix}$$

so the factorization is correct.
The general statement of the converse for Theorem 1 is

**Theorem 2** Let \( A \) be an \( n \times n \) matrix. Suppose \( \mathbb{R}^n \) has a basis \( \{ b_1, b_2, \ldots, b_n \} \) that consists of eigenvectors of \( A \).

Conclusion: then \( A \) is diagonalizable, that is \( A = PDP^{-1} \) for some diagonal matrix \( D \).

In fact, the matrix \( P \) has the eigenvectors \( b_1, b_2, \ldots, b_n \) as its columns, and the diagonal entries of \( D \) are the corresponding eigenvalues \( \lambda_1, \lambda_2, \ldots, \lambda_n \) (written down the diagonal in the same order: \( \lambda_1 \) for \( b_1 \), etc.)

**Why is Theorem 2 true?** The argument is written here is for a \( 2 \times 2 \) matrix \( A \). (Follow the argument using the preceding example as a model). The argument works in exactly the same way when \( A \) is \( n \times n \) — there is a just bit more notation to keep track of.

Suppose \( A \) is \( 2 \times 2 \) and that \( \mathcal{B} = \{ b_1, b_2 \} \) is a basis \( \mathbb{R}^2 \) consisting of eigenvectors of \( A \). Let the corresponding eigenvalues be \( \lambda_1 \) and \( \lambda_2 \).

Create matrices \( P = [ b_1 \ b_2 ] \) and \( D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \). Since \( b_1 \) and \( b_2 \) are linearly independent, \( P \) is invertible.

To complete the proof that \( A \) is diagonalizable, we claim that these matrices work: that \( A = PDP^{-1} \). As in the preceding example, we just need to verify that \( AP = PD \), and for that, we just need to remember the definition of matrix multiplication:

\[
PD = P \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = P \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} 0 \\ \lambda_2 \end{bmatrix} \\
= [ [ b_1 \ b_2 ] \cdot \begin{bmatrix} \lambda_1 \\ 0 \end{bmatrix} ] [ b_1 \ b_2 ] \cdot \begin{bmatrix} 0 \\ \lambda_2 \end{bmatrix} = [ \lambda_1 b_1 \ \
\lambda_2 b_2 ], \text{ and}
\]

\[
AP = A[b_1 \ b_2] = [ Ab_1 \ Ab_2 ] = [ \lambda_1 b_1 \ \
\lambda_2 b_2 ]
\]

because \( b_1 \) is an eigenvector of \( A \) with eigenvalue \( \lambda_1 \)
and \( b_2 \) is an eigenvector of \( A \) with eigenvalue \( \lambda_2 \).

\( AP \) and \( PD \) have the same columns, so \( AP = PD \)

**Question** Based on this discussion, can you give an example of a transformation \( T : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \) where \( T(x) = Ax \) and \( A \) is **not** diagonalizable?
The Big Picture, so far

In Chapter 4, we have been discussing vector spaces $V$ (where $V$ might not be $\mathbb{R}^n$). After discussing linear independence and spanning in this more general setting, we were led to the idea of a basis $B = \{b_1, \ldots, b_n\}$ for $V$. From that, the Unique Representation Theorem (p. 246) led to the idea of using $B$ to create coordinates for each $x$ in $V$. The coordinate vector $[x]_B$ is a vector in $\mathbb{R}^n$.

We saw that the mapping $x \mapsto [x]_B$ is an isomorphism (a one-to-one, onto linear mapping) from the vector space $V$ to $\mathbb{R}^n$. This association preserves all vector space operations and all linear dependency relations. For example, a linearly independent set of vectors in $V$ has a linearly independent set of coordinate vectors in $\mathbb{R}^n$, and vice-versa.

In the special case when $V = \mathbb{R}^n$, we might choose a basis $B = \{b_1, \ldots, b_n\}$ different from the standard basis. Then a vector $x$ in $\mathbb{R}^n$ gets new “nonstandard” coordinates $[x]_B$ relative to the basis $B$. The matrix $P_B = [b_1 \ldots b_n]$ is the “operator” that changes $B$ coordinates into standard coordinates according to the formula $P_B [x]_B = x$.

Since the columns of $P_B$ are linearly independent, the change of coordinates matrix $P_B$ is always invertible and $P_B^{-1}$ is the “operator” that converts standard coordinates into $B$-coordinates: $[x]_B = P_B^{-1} x$.

When an $n \times n$ matrix $A$ can be factored into the form $PDP^{-1}$, where $D$ is a diagonal matrix, then $A$ is called diagonalizable — because $A$ “acts like a diagonal matrix” when computations are done in $B$ coordinates instead of standard coordinates. The columns of $P$ are the vectors in the basis $B$.

An example led into the idea of eigenvectors and eigenvalues of the matrix $A$.

We proved that a $2 \times 2$ matrix $A$ is diagonalizable if and only if $\mathbb{R}^2$ has a basis consisting of eigenvectors of $A$ — and indicated that a completely similar proof works for an $n \times n$ matrix $A$ operating on $\mathbb{R}^n$.

The conceptual idea of diagonalization and its relation to a basis of eigenvectors is nicely motivated geometrically and not very hard. You may have noticed, however, that in the preceding examples, either

i) a factorization of a given matrix $A$ into $PDP^{-1}$ was given, or
ii) the eigenvalues and eigenvectors of $A$ were given so that $P$ and $D$ could be created.

But if you are only given an $n \times n$ matrix $A$, then trying to find its eigenvectors and eigenvalues (if it has any at all!), and determining whether $\mathbb{R}^n$ does or does not have a basis consisting of eigenvectors of $A$ are harder questions. Be aware of those questions, but try to keep worries about them suppressed until we get into Chapter 5. For now just focus on the concept of diagonalization, what it means, and how diagonalization is connected to eigenvectors and eigenvalues and changing coordinates!