Introduction to Diagonalization

For a square matrix A, a process called "diagonalization" can sometimes give us more insight into how the transformation $\boldsymbol{x} \mapsto A\boldsymbol{x}$ "works." The insight has a strong geometric flavor, but we will see later that it can also be very useful in applications.

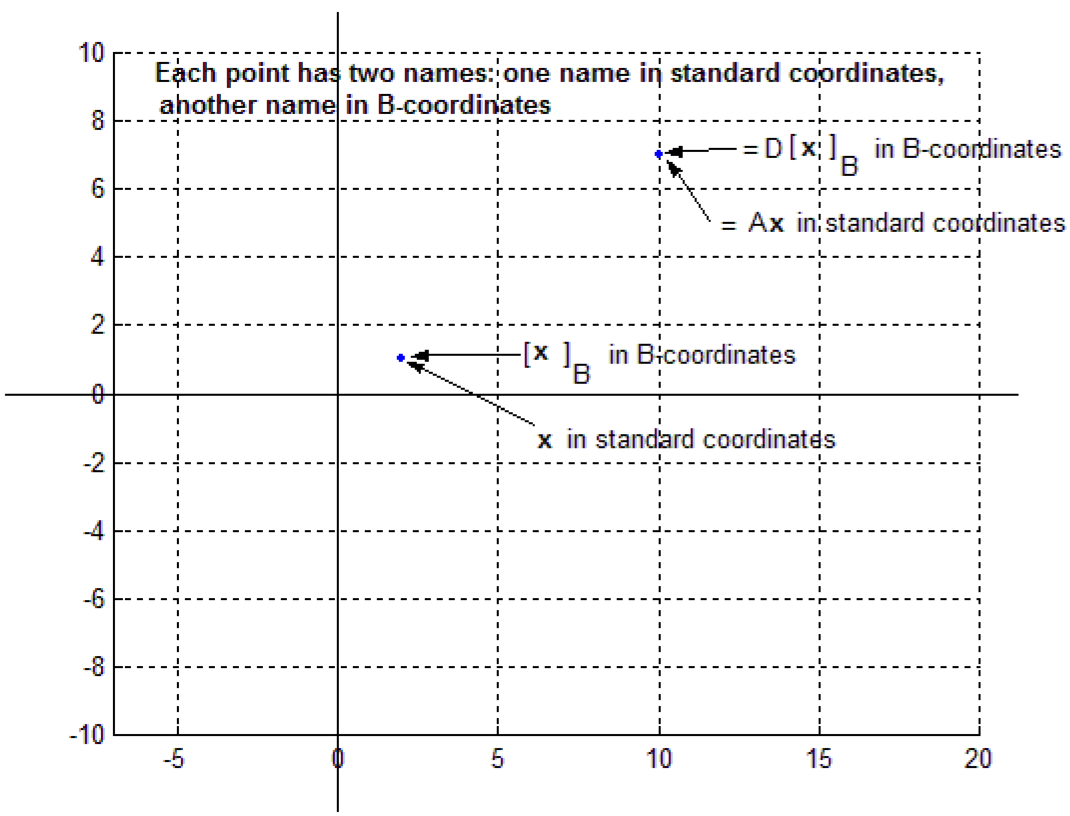
The idea is to try to find to a new basis $\mathcal{B}$ for $\mathbb{R}^n$ so that, if we compute in terms of $\mathcal{B}$-coordinates, A "works like" a diagonal matrix D. We would like to find $\mathcal{B}$-coordinates and a diagonal matrix D so that

$\boldsymbol{x} \mapsto A\boldsymbol{x}$ (a computation in standard coordinates)

<u>is the same as</u> *same geometric point but named in two different coordinate systems*

$[\boldsymbol{x}]_{\mathcal{B}} \mapsto D[\boldsymbol{x}]_{\mathcal{B}}$ (a computation is done in $\mathcal{B}$-coordinates)

The point pictured below at the lower left (named $\boldsymbol{x}$ or $[\boldsymbol{x}]_{\mathcal{B}}$ in the two different coordinate systems) is moved by the transformation to the point pictured at the upper right (named $A\boldsymbol{x}$ in one coordinate system or $D[\boldsymbol{x}_{\mathcal{B}}]$ in the other coordinate system). Geometrically, the one point is moved to the other point; but we describe that action using two different coordinate systems: as $\boldsymbol{x} \mapsto A\boldsymbol{x}$ or as $[\boldsymbol{x}]_{\mathcal{B}} \mapsto D[\boldsymbol{x}]_{\mathcal{B}}$.



Changing coordinates: remember that multiplication by the change of basis matrix $P_{\mathcal{B}}$ converts $\mathcal{B}$-coordinates into standard coordinates, and the inverse matrix converts standard coordinates into $\mathcal{B}$-coordinates:

$$\begin{aligned} P_{\mathcal{B}}[\boldsymbol{x}]_{\mathcal{B}} &= \boldsymbol{x} \\ [\boldsymbol{x}]_{\mathcal{B}} &= P_{\mathcal{B}}^{-1}\boldsymbol{x} \end{aligned}$$

So what we want is to find (if possible) a basis $\mathcal{B}$ and a diagonal matrix D so that

$$P_{\mathcal{B}} \quad \cdot \quad D \quad \cdot \quad \underbrace{P_{\mathcal{B}}^{-1}\boldsymbol{x}} \quad = \quad A\boldsymbol{x}$$

$\leftarrow$ converts $\boldsymbol{x}$ into $[\boldsymbol{x}]_{\mathcal{B}}$ ($\mathcal{B}$ coordinates)

$\uparrow$ compute $D[\boldsymbol{x}]_{\mathcal{B}}$

$\uparrow$ finally, convert the result $D[\boldsymbol{x}]_{\mathcal{B}}$ back into standard coordinates

In other words, we want a basis $\mathcal{B}$ and diagonal matrix D so that A factors as

$$P_{\mathcal{B}}DP_{\mathcal{B}}^{-1} = A \quad (*)$$

The same ideas apply just as well in $\mathbb{R}^n$ except that it's harder to draw a picture in $\mathbb{R}^3$ (and impossible to draw in $\mathbb{R}^n$ for $n > 3$).

Definition If there is a basis $\mathcal{B}$ for $\mathbb{R}^n$ and a diagonal matrix D such that the factorization (*) is true, we say that the matrix A is diagonalizable.

If we can do this, it's a good thing – because (if we compute in $\mathcal{B}$ coordinates):

i) diagonal matrices are so easy to work with, and

ii) it's easy to visualize geometrically what a mapping $\boldsymbol{z} \mapsto D\boldsymbol{z}$ does to $\boldsymbol{z}$.

Both i) and ii) can be important in applications.

Unfortunately, it is not always possible to "diagonalize" a square matrix A. We'll learn something in these notes about when it is possible, but more details will need to wait until Chapter 5. For now, introducing the basic ideas

i) will let us practice with changing bases, and
ii) will be good preparation for the lster material in Chapter 5.

The first example gives an illustration of why diagonalization is useful.

Example This very elementary example is in $\mathbb{R}^2$. Exactly the same ideas apply for $n \times n$ matrices A, but working in $\mathbb{R}^2$ with a 2×2 matrix A makes the visualization much easier.

If D is a 2×2 diagonal matrix, what does the mapping $\boldsymbol{x} \mapsto D\boldsymbol{x}$ do to $\boldsymbol{x}$ geometrically? That is, how does D operate on $\mathbb{R}^2$? To be specific, suppose $D = \begin{bmatrix} 3 & 0 \\ 0 & 6 \end{bmatrix}$.

When multiply the basis vectors $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ by D, we just get a rescaling. These vectors are mapped onto scalar multiples of themselves: the scalars involved are the diagonal elements of D.

$$D\boldsymbol{e_1} = D\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix} = 3\boldsymbol{e_1}, \text{ and}$$

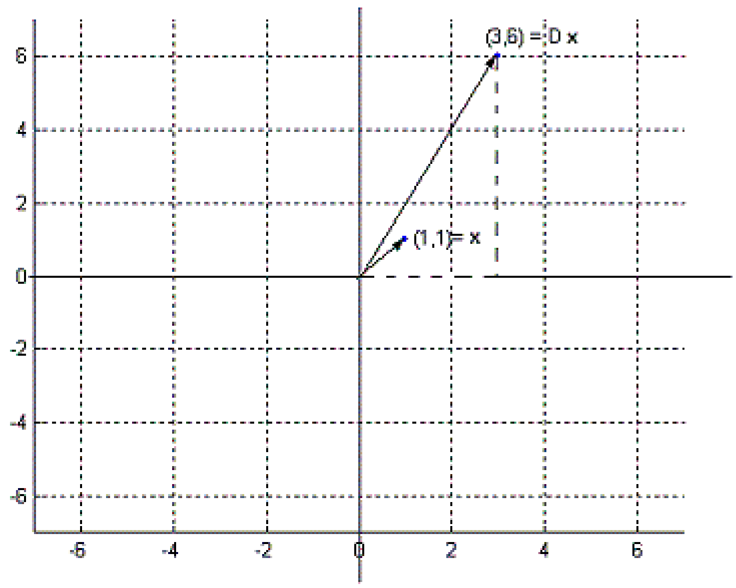
$$D\boldsymbol{e_2} = D\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \end{bmatrix} = 6\boldsymbol{e_2}$$

For any $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, something similar happens:

$$D\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 6 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3x_1 \\ 6x_2 \end{bmatrix}$$

D simply rescales each coordinate of $\boldsymbol{x}$ – by a factor of 3 in the first coordinate and and by a factor of 6 in the second coordinate. In this particular example, the diagonal entries of D are both bigger than 1, so multiplying $\boldsymbol{x}$ by D "stretches" or "dilates" each coordinate. This is pictured below (for standard coordinates) where

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



Example For a more general 2×2 like $A = \begin{bmatrix} 4 & 2 \\ 1 & 5 \end{bmatrix}$, it's not so clear geometrically what A does to $\boldsymbol{x}$. However we can get some insight by doing our calculations in terms of a new basis for $\mathbb{R}^2$.

First of all, it's a fact (*just check by multiplying, if you like*) that

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

$$= \quad \underset{P}{\uparrow} \qquad \underset{D}{\uparrow} \qquad \underset{P^{-1}}{\uparrow}$$

For now, don't worry about where the matrix P came from. Such questions come up in Chapter 5.

P is invertible so, by the Invertible Matrix Theorem, its columns are linearly independent and span $\mathbb{R}^2$. Therefore we can use the <u>columns of P</u> as a new basis for $\mathbb{R}^2$:

$$\mathcal{B} = \{\boldsymbol{b_1}, \boldsymbol{b_2}\} = \{ \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \}.$$

(*We looked at this basis as an example in the preceding lecture.*) Then P is the matrix that the textbook calls $P_{\mathcal{B}} =$ "the change of coordinates matrix from $\mathcal{B}$-coordinates to standard coordinates".

$$P_{\mathcal{B}}[\boldsymbol{x}]_{\mathcal{B}} = \boldsymbol{x}$$

The new basis lets us see more clearly how A operates. (*See also the figure, below*)

Suppose, for example, that $\boldsymbol{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ (standard coordinates)

$$A\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 1 & 5 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 10 \\ 7 \end{bmatrix}.$$

We can also calculate $A\boldsymbol{x}$ in a more roundabout way – but one which gives us new insight:

$$A\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \underset{\substack{\uparrow \\ A}}{\begin{bmatrix} 4 & 2 \\ 1 & 5 \end{bmatrix}}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \underset{\substack{\uparrow \\ P_{\mathcal{B}}}}{\begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix}}\underset{\substack{\uparrow \\ D}}{\begin{bmatrix} 3 & 0 \\ 0 & 6 \end{bmatrix}}\underset{\substack{\uparrow \\ P_{\mathcal{B}}^{-1}}}{\begin{bmatrix} \frac{1}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}}\underset{\substack{\uparrow \\ \boldsymbol{x}}}{\begin{bmatrix} 2 \\ 1 \end{bmatrix}} = \begin{bmatrix} 10 \\ 7 \end{bmatrix}$$

In the second calculation, look at what happens to $\boldsymbol{x}$, step-by-step, as each of the three matrices, in turn, does its work:

$$\boldsymbol{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} \frac{1}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \underset{\substack{\uparrow \\ [\boldsymbol{x}]_{\mathcal{B}}}}{\begin{bmatrix} \frac{1}{3} \\ \frac{4}{3} \end{bmatrix}}.$$

multiplication by $P_{\mathcal{B}}^{-1}$ converts standard coordinates into $\mathcal{B}$-coordinates

<u>Then</u>

$$\begin{bmatrix} 3 & 0 \\ 0 & 6 \end{bmatrix}\begin{bmatrix} \frac{1}{3} \\ \frac{4}{3} \end{bmatrix} = \begin{bmatrix} 1 \\ 8 \end{bmatrix}$$

the <u>diagonal matrix</u> D stretches the new $\mathcal{B}$-coordinates by factors of 3 and 6 (in directions corresponding to $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$)

(stretched $\mathcal{B}$-coordinates)

<u>Finally</u>,

$$\begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 8 \end{bmatrix} = \begin{bmatrix} 10 \\ 7 \end{bmatrix}$$

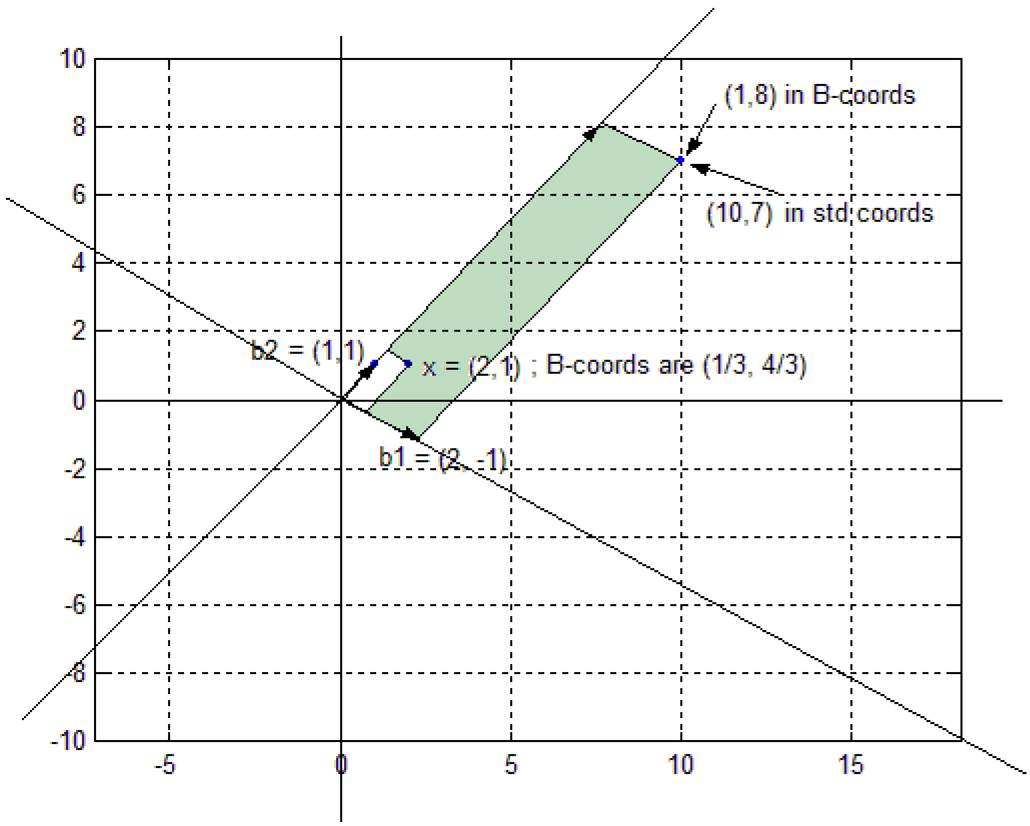
multiplication by $P_{\mathcal{B}}$ converts the (stretched) $\mathcal{B}$-coordinates back into standard coordinates

Use the figure below to check each of the preceding steps graphically, as accurately as you can.

1) $\boldsymbol{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ has $\mathcal{B}$-coordinates $[\boldsymbol{x}]_{\mathcal{B}} = \begin{bmatrix} \frac{1}{3} \\ \frac{4}{3} \end{bmatrix}$

2) Stretching $[\boldsymbol{x}]_{\mathcal{B}}$ by factors of 3 and 6 in the $\boldsymbol{b_1}$ and $\boldsymbol{b_2}$ directions gives the point with $\mathcal{B}$-coordinates $\begin{bmatrix} 1 \\ 8 \end{bmatrix}$.

3) Converting these $\mathcal{B}$-coordinates $\begin{bmatrix} 1 \\ 8 \end{bmatrix}$ back to standard coordinates gives $\begin{bmatrix} 10 \\ 7 \end{bmatrix} = A \begin{bmatrix} 2 \\ 1 \end{bmatrix}$.



So the effect of A on a point $\boldsymbol{x}$ (when described in standard coordinates) is the same as the effect of the diagonal matrix D on the same point $\boldsymbol{x}$ (when described in $\mathcal{B}$-coordinates). D stretches the $\mathcal{B}$-coordinates by a factor of 3 in the direction of the "positive $\boldsymbol{b_1}$-axis" and by a factor of 6 in the direction of the "positive $\boldsymbol{b_2}$-axis."

There is another important observation in this example: look especially at the point $\boldsymbol{x} = \boldsymbol{b_1}$ or $\boldsymbol{x} = \boldsymbol{b_2}$. Then

$$A\boldsymbol{b_1} = \begin{bmatrix} 4 & 2 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 6 \\ -3 \end{bmatrix} = 3\begin{bmatrix} 2 \\ -1 \end{bmatrix} = 3\boldsymbol{b_1}$$

Again, it's helpful to look at what's going on in more detail, using the same 3-step process (*follow each step using the preceding figure*) :

$$A\boldsymbol{b_1} = \begin{bmatrix} 4 & 2 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 6 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 6 \end{bmatrix} \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{\substack{\mathcal{B}\text{-coordinates} \\ \text{of } \boldsymbol{b_1} = \\ \text{“unit vector” in the } \boldsymbol{b_1} \text{ direction}}} = \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix} \underbrace{\begin{bmatrix} 3 \\ 0 \end{bmatrix}}_{\substack{\text{stretched} \\ \mathcal{B}\text{-coordinates}}} = \underbrace{\begin{bmatrix} 6 \\ -3 \end{bmatrix}}_{\substack{\text{converted back into standard} \\ \text{coordinates}}} = 3\begin{bmatrix} 2 \\ -1 \end{bmatrix} = 3\boldsymbol{b_1}$$

You should analyze the equation $A\boldsymbol{b_2} = \begin{bmatrix} 4 & 2 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 6 \end{bmatrix} = 6\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 6\boldsymbol{b_2}$ in the same way.

Thus <u>the matrix A maps each vector</u> $\boldsymbol{b_1}$ <u>and</u> $\boldsymbol{b_2}$ <u>to a scalar multiple of itself</u>. Such “vectors are called <u>eigenvectors</u> of A; the scalars (“3” for the vector $\boldsymbol{b_1}$ and “6” for the vector $\boldsymbol{b_2}$) are called the <u>eigenvalues</u> associated with the eigenvectors $\boldsymbol{b_1}$ and $\boldsymbol{b_2}$.

The reason we were able to do such a nice analysis of how this matrix A works is that we were able to write down (*using the matrix P that I <u>gave</u> you in the beginning*) a very special new basis $\mathcal{B}$ for $\mathbb{R}^2$ – <u>a basis whose members are eigenvectors</u> of the matrix A.

The general definition is:.

Definition A <u>nonzero</u> vector $\boldsymbol{x}$ is called an <u>eigenvector</u> of the $n \times n$ matrix A if $A\boldsymbol{x} = \lambda\boldsymbol{x}$ for some scalar λ. The scalar λ is called an <u>eigenvalue</u> of A (associated with the eigenvector $\boldsymbol{x}$).

(It's traditional, in almost all books, to use the Greek letter λ (“lambda”) to denote an eigenvalue. Some books call “eigenvectors” and “eigenvalues” by the loose translations “characteristic vectors” and “characteristic values.”)

In the preceding 2×2 example there is nothing special about the specific eigenvalues 3 and 6. They might be any other numbers λ_1 and λ_2 (where perhaps even $\lambda_1 = \lambda_2$).
In general,

if A can be factored as $A = P \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} P^{-1}$, where P has columns $\boldsymbol{b_1}, \boldsymbol{b_2}$,
then $\boldsymbol{b_1}$ and $\boldsymbol{b_2}$ are <u>eigenvectors</u> of A with <u>eigenvalues</u> λ_1 and λ_2 <u>because</u>:

We can take $\mathcal{B} = \{\boldsymbol{b_1}, \boldsymbol{b_2}\}$ as a new <u>basis</u> (*why?*) for $\mathbb{R}^2$, and then calculate

$$A\boldsymbol{b_1} = P_{\mathcal{B}} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} P_{\mathcal{B}}^{-1} \boldsymbol{b_1} = P_{\mathcal{B}} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{\substack{\mathcal{B}\text{-coordinates} \\ \text{of } \boldsymbol{b_1} \text{ since } \boldsymbol{b_1} = 1\boldsymbol{b_1} + 0\boldsymbol{b_2}}}$$

$$= P_{\mathcal{B}} \underbrace{\begin{bmatrix} \lambda_1 \\ 0 \end{bmatrix}}_{\text{in } \mathcal{B}\text{-coordinates}} = [\,\boldsymbol{b_1}\ \boldsymbol{b_2}\,] \begin{bmatrix} \lambda_1 \\ 0 \end{bmatrix} = \underbrace{\lambda_1 \boldsymbol{b_1} + 0\boldsymbol{b_2} = \lambda_1 \boldsymbol{b_1}}_{\text{back to standard coordinates}}$$

Similarly, $A\boldsymbol{b_2} = \lambda_2 \boldsymbol{b_2}$.

In this case, the <u>geometric</u> interpretation of how A operates on a point is just as it was earlier: convert the standard coordinates $\boldsymbol{x}$ of the point to $\mathcal{B}$-coordinates. Then multiplication by D rescales the first $\mathcal{B}$-coordinate by a factor of λ_1 and the second by a factor of λ_2. (*Of course if, say,* $0 < \lambda_1 < 1$ *then the rescaling of the first coordinate is a "contraction" of the first coordinate of* $\boldsymbol{x}$ *rather than a "stretch"; and if (say)* $\lambda_2 < 0$, *the rescaling of the second coordinate also reverses the sign of the second coordinate of* $\boldsymbol{x}$ *(as well as either stretching or contracting*)).

Again, "A acts like a diagonal matrix D" if you calculate in $\mathcal{B}$-coordinates.

<u>Exactly</u> the same algebraic manipulations and geometric interpretation applies in $\mathbb{R}^n$ with an $n \times n$ matrix A $-$ <u>when it can be factored as</u> $A = PDP^{-1}$, <u>where</u> D <u>is diagonal matrix</u>. This motivates the following definition.

To repeat the definition given earlier:

Definition Suppose A is an $n \times n$ matrix. If A can be factored as $A = PDP^{-1}$, where D is diagonal, then we say that A is a <u>diagonalizable matrix.</u>

(*It is <u>not</u> always possible to factor* A *is this way. Based on the <u>geometric</u> notions and ideas used above, can you think of a* 2×2 *matrix* A *that <u>isn't</u> diagonalizable? Can you think of a* 2×2 *matrix for which there are <u>no</u> eigenvectors ?*)

The preceding discussion proves the following theorem (where $n = 2$) :

Theorem 1 Suppose A is a 2×2 diagonalizable matrix, say $A = P \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} P^{-1}$ where $P = [\,\boldsymbol{b_1}\ \boldsymbol{b_2}\,]$. Then

i) the columns of P are eigenvectors of the matrix A, and
ii) they form a basis for $\mathbb{R}^2$

The eigenvalues λ_1 and λ_2 for these eigenvectors are the scalars found on the diagonal of the corresponding column of D.

Moreover, a completely similar argument works for an $n \times n$ matrix A if $A = PDP^{-1}$ where D is diagonal. Therefore we can say

Theorem 1 *Suppose A is an $n \times n$ matrix diagonalizable matrix, say*

$$A = P \begin{bmatrix} \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & \cdots & \lambda_3 & \cdots & 0 \\ 0 & \vdots & \vdots & \vdots & 0 \\ 0 & 0 & 0 & 0 & \lambda_n \end{bmatrix} P^{-1} \ ,$$

where $P = [\,\boldsymbol{b_1}\ \boldsymbol{b_2}\ \ldots\ \boldsymbol{b_n}\,]$. Then

i) the columns of P are eigenvectors of the matrix A, and
ii) they form a basis for $\mathbb{R}^n$.

The eigenvalues for these eigenvectors are the scalars found on the diagonal of the corresponding column of D (for example, λ_1 for $\boldsymbol{b_1}$ etc.)

It turns out that the converse of *Theorem 1* is also true: that is,

Theorem 2 *If A is an $n \times n$ matrix and if there is a basis $\mathcal{B} = \{\boldsymbol{b_1}, \boldsymbol{b_2}, \ldots, \boldsymbol{b_n}\}$ for $\mathbb{R}^n$ consisting of eigenvectors of A (with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$), then A is diagonalizable.*

Specifically, we can factor $A = PDP^{-1}$, where

$$P = [\,\boldsymbol{b_1}\ \boldsymbol{b_2}\ \ldots\ \boldsymbol{b_n}] \text{ and } D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \lambda_n \end{bmatrix}$$

First we illustrate the meaning of Theorem 2 with another example, using $n = 2$ again to keep the notation as simple as possible. Then we will give the proof of the theorem in the case $n = 2$. The general proof for an $n \times n$ matrix is done in exactly same way. (*Not surprisingly, the proof closely parallels the ideas in the example*).

Example Suppose $A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}$. $\mathcal{B} = \{\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ -2 \end{bmatrix}\} = \{\boldsymbol{b_1}, \boldsymbol{b_2}\}$ is a basis for $\mathbb{R}^2$ (*why?*).

Since $A\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 5 \\ -5 \end{bmatrix} = 5\begin{bmatrix} 1 \\ -1 \end{bmatrix}$, we get that

$\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ is an eigenvector with eigenvalue 5.

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

$\begin{bmatrix} 1 \\ -2 \end{bmatrix}$ is an eigenvector with eigenvalue 3.

Thus, $\mathcal{B} = \{\boldsymbol{b_1}, \boldsymbol{b_2}\}$ is a basis for $\mathbb{R}^2$ consisting of eigenvectors of A. This tells use that A is diagonalizable, because:

Let $P = [\boldsymbol{b_1}\ \boldsymbol{b_2}] = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}$. Because the columns of P are linearly independent, P is invertible. Let $D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$, where the diagonal entries are the eigenvalues corresponding to the columns of P (*in the same order!*)

Theorem 2 asserts that having done this, it now turns out that $A = PDP^{-1}$. To illustrate *Theorem 2*, we now verify that this is true.

If we multiply $A = PDP^{-1}$ on both sides by P (on the right), we see that the equation $A = PDP^{-1}$ is equivalent to $AP = PD$. We check that this is true rather than writing down P^{-1} :

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

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

Proof of ***Theorem 2*** (*for the case* $n = 2$)

Suppose A is 2×2 and that $\mathcal{B} = \{\boldsymbol{b_1}, \boldsymbol{b_2}\}$ is a basis $\mathbb{R}^2$ where $\boldsymbol{b_1}, \boldsymbol{b_2}$ are eigenvectors of A, with corresponding eigenvalues λ_1 and λ_2.

Let $P = [\,\boldsymbol{b_1}\ \boldsymbol{b_2}]$ and $D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$. Since $\boldsymbol{b_1}$ and $\boldsymbol{b_2}$ are linearly independent, P is invertible. To complete the proof that A is diagonalizable, we show now that $A = PDP^{-1}$.

As in the preceding example, we just need to verify that $AP = PD$, and, to do this, we simply need to remember the definition of matrix multiplication.

$$PD = P\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = [\,P\begin{bmatrix} \lambda_1 \\ 0 \end{bmatrix} \quad P\begin{bmatrix} 0 \\ \lambda_2 \end{bmatrix}\,]$$

$$= [\ [\boldsymbol{b_1}\ \boldsymbol{b_2}\,] \cdot \begin{bmatrix} \lambda_1 \\ 0 \end{bmatrix} \quad [\,\boldsymbol{b_1}\ \boldsymbol{b_2}\,] \cdot \begin{bmatrix} 0 \\ \lambda_2 \end{bmatrix}\ ]$$

$$= [\,\lambda_1\boldsymbol{b_1}\ \ \lambda_2\boldsymbol{b_2}\,], \text{ and}$$

$$AP = A[\,\boldsymbol{b_1}\ \boldsymbol{b_2}] = [\,A\boldsymbol{b_1}\ \ A\boldsymbol{b_2}\,] \underset{\uparrow}{=} [\,\lambda_1\boldsymbol{b_1}\ \ \lambda_2\boldsymbol{b_2}\,]$$

because $\boldsymbol{b_1}$ of A with eigenvalue λ_1,
and $\boldsymbol{b_2}$ is an eigenvector of A with eigenvalue λ_2.

Therefore AP and PD have the same columns, so $AP = PD$.

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 sets in this more general setting, we were led to the idea of a basis $\mathcal{B} = \{\boldsymbol{b_1}, ..., \boldsymbol{b_n}\}$ for V. From that, the Unique Representation Theorem (p. 216) led to the idea of using $\mathcal{B}$ to define coordinates for each $\boldsymbol{x}$ in V. The

coordinate vector $[\boldsymbol{x}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$ if $\boldsymbol{x} = c_1\boldsymbol{b_1} + c_2\boldsymbol{b_2} + ... + c_n\boldsymbol{b_n}$. The coordinate vector $[\boldsymbol{x}]_{\mathcal{B}}$ is a vector in $\mathbb{R}^n$.

We saw that the mapping $\boldsymbol{x} \mapsto [\boldsymbol{x}]_{\mathcal{B}}$ from V to $\mathbb{R}^n$ is an isomorphism (a one-to-one, onto linear transformation). This mapping preserves all vector space operations and all linear dependency and independency relations. For example, a set of vectors in V is linearly independent (or dependent) if and only if the corresponding set of coordinate vectors in $\mathbb{R}^n$ is linearly independent (dependent). Vector spaces computations in V are "perfectly mirrored" by the corresponding coordinate vectors computations in $\mathbb{R}^n$..

In the special case when $V = \mathbb{R}^n$, we could choose a basis $\mathcal{B} = \{\boldsymbol{b_1}, ..., \boldsymbol{b_n}\}$ different from the standard basis $\{\boldsymbol{e_1}, ..., \boldsymbol{e_n}\}$. Each vector $\boldsymbol{x}$ in $\mathbb{R}^n$ then gets a "new name": its coordinate vector $[\boldsymbol{x}]_{\mathcal{B}}$ relative to the basis $\mathcal{B}$. The matrix $P_{\mathcal{B}} = [\,\boldsymbol{b_1} \dots \boldsymbol{b_n}]$ is the "operator" that changes $\mathcal{B}$-coordinates into standard coordinates according to the formula $P_{\mathcal{B}}\,[\boldsymbol{x}]_{\mathcal{B}} = \boldsymbol{x}$.

Since the columns of $P_{\mathcal{B}}$ are linearly independent, the change of coordinates matrix $P_{\mathcal{B}}$ is always invertible and $P_{\mathcal{B}}^{-1}$ is the "operator" that converts standard coordinates into $\mathcal{B}$-coordinates: $[\boldsymbol{x}]_{\mathcal{B}} = P_{\mathcal{B}}^{-1}\boldsymbol{x}$.

For any invertible $n \times n$ matrix A, the column vectors can be used as a basis $\mathcal{B}$ for $\mathbb{R}^n$ and for $\mathcal{B}$ coordinates. In that case, $P_{\mathcal{B}}$ is just the matrix A.

.

If a 2×2 matrix A can be factored into the form PDP^{-1}, where D is a diagonal matrix, then A is called diagonalizable – because in that case A "acts like a diagonal matrix" when computations are done relative to the basis $\mathcal{B}$ that consists of the columns. This led us to the idea of eigenvectors and eigenvalues of the matrix A.

We proved that a 2×2 matrix 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, a factorization of a given matrix A into PDP^{-1} was given, or a basis of eigenvectors was given so that the matrices P and D could be created. But if you are simply given an $n \times n$ matrix A, then trying

i) to find its eigenvalues and eigenvectors (*if it has any at all!*), and
ii) to determine 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 ignore your worries about them 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.