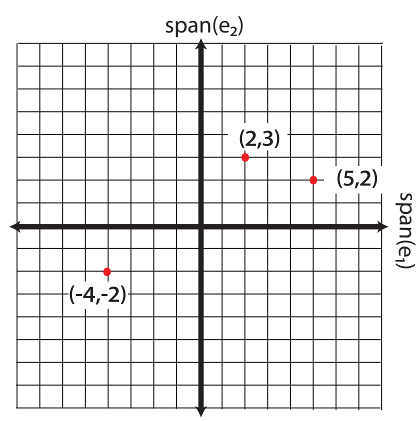
CHAPTER 9

BASIS AND CHANGE OF BASIS

When we think of coordinate pairs, or coordinate triplets, we tend to think of them as points on a grid where each axis represents one of the coordinate directions:



When we think of our data points this way, we are considering them as linear combinations of elementary **basis vectors**

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ and } \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

For example, the point $(2, 3)$ is written as

$$\begin{pmatrix} 2 \\ 3 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 3 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 2\mathbf{e}_1 + 3\mathbf{e}_2. \tag{9.1}$$

We consider the coefficients (the scalars 2 and 3) in this linear combination as **coordinates** in the basis $\mathcal{B}_1 = \{\mathbf{e}_1, \mathbf{e}_2\}$. The coordinates, in essence, tell us how much "information" from the vector/point $(2,3)$ lies along each basis direction: to create this point, we must travel 2 units along the direction of $\mathbf{e}_1$ and then 3 units along the direction of $\mathbf{e}_2$.

We can also view Equation 9.1 as a way to separate the vector $(2,3)$ into orthogonal components. Each component is an **orthogonal projection** of the vector onto the span of the corresponding basis vector. The orthogonal projection of vector **a** onto the span another vector **v** is simply the closest point to **a** contained on the span(**v**), found by "projecting" **a** toward **v** at a 90° angle. Figure 9.1 shows this explicitly for $\mathbf{a} = (2,3)$.

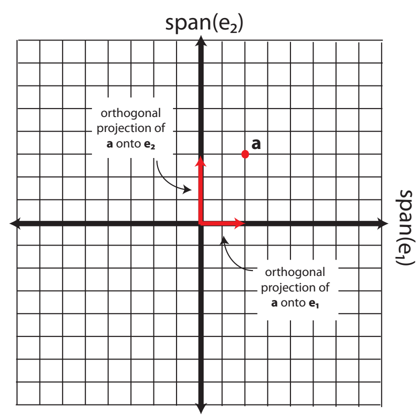


Figure 9.1: Orthogonal Projections onto basis vectors.

Definition 9.0.1: Elementary Basis

For any vector $\mathbf{a} = (a_1, a_2, \dots, a_n)$, the basis $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ (recall $\mathbf{e}_i$ is the i^{th} column of the identity matrix $\mathbf{I}_n$) is the **elementary basis** and **a** can be written in this basis using the **coordinates** $a_1, a_2, \dots, a_n$ as follows:

$$\mathbf{a} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + \dots a_n\mathbf{e}_n.$$

The elementary basis $\mathcal{B}_1$ is convenient for many reasons, one being its orthonormality:

$$\begin{aligned} \mathbf{e}_1^T\mathbf{e}_1 &= \mathbf{e}_2^T\mathbf{e}_2 = 1 \\ \mathbf{e}_1^T\mathbf{e}_2 &= \mathbf{e}_2^T\mathbf{e}_1 = 0 \end{aligned}$$

However, there are many (infinitely many, in fact) ways to represent the data points on different axes. If I wanted to view this data in a different way, I could use a different basis. Let's consider, for example, the following orthonormal basis, drawn in green over the original grid in Figure 9.2:

$$\mathcal{B}_2 = \{\mathbf{v}_1, \mathbf{v}_2\} = \left\{ \frac{\sqrt{2}}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \frac{\sqrt{2}}{2}\begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}$$

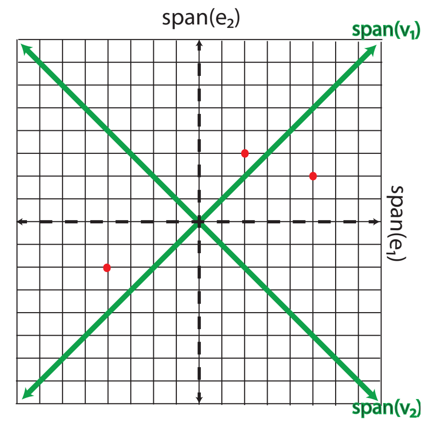


Figure 9.2: New basis vectors, $\mathbf{v}_1$ and $\mathbf{v}_2$, shown on original plane

The scalar multipliers $\frac{\sqrt{2}}{2}$ are simply normalizing factors so that the basis vectors have unit length. You can convince yourself that this is an orthonormal basis by confirming that

$$\begin{aligned} \mathbf{v}_1^T\mathbf{v}_1 &= \mathbf{v}_2^T\mathbf{v}_2 = 1 \\ \mathbf{v}_1^T\mathbf{v}_2 &= \mathbf{v}_2^T\mathbf{v}_1 = 0 \end{aligned}$$

If we want to *change the basis* from the elementary $\mathcal{B}_1$ to the new green basis vectors in $\mathcal{B}_2$, we need to determine a new set of coordinates that direct us to the point using the green basis vectors as a frame of reference. In other words we need to determine (α_1, α_2) such that travelling α_1 units along the direction $\mathbf{v}_1$ and then α_2 units along the direction $\mathbf{v}_2$ will lead us to the point in question. For the point $(2, 3)$ that means

$$\begin{pmatrix} 2 \\ 3 \end{pmatrix} = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 = \alpha_1 \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix} + \alpha_2 \begin{pmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{pmatrix}.$$

This is merely a system of equations $\mathbf{Va} = \mathbf{b}$:

$$\frac{\sqrt{2}}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

The 2×2 matrix $\mathbf{V}$ on the left-hand side has linearly independent columns and thus has an inverse. In fact, $\mathbf{V}$ is an orthonormal matrix which means its inverse is its transpose. Multiplying both sides of the equation by $\mathbf{V}^{-1} = \mathbf{V}^T$ yields the solution

$$\mathbf{a} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \mathbf{V}^T\mathbf{b} = \begin{pmatrix} \frac{5\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{pmatrix}$$

This result tells us that in order to reach the red point (formerly known as (2,3) in our previous basis), we should travel $\frac{5\sqrt{2}}{2}$ units along the direction of $\mathbf{v}_1$ and then $-\frac{\sqrt{2}}{2}$ units along the direction $\mathbf{v}_2$ (Note that $\mathbf{v}_2$ points toward the southeast corner and we want to move northwest, hence the coordinate is negative). Another way (a more mathematical way) to say this is that *the length of the orthogonal projection of* $\mathbf{a}$ *onto the span of* v_1 *is* $\frac{5\sqrt{2}}{2}$*, and the length of the orthogonal projection of* $\mathbf{a}$ *onto the span of* v_2 *is* $-\frac{\sqrt{2}}{2}$. While it may seem that these are difficult distances to plot, they work out quite well if we examine our drawing in Figure 9.2, because the diagonal of each square is $\sqrt{2}$.

In the same fashion, we can re-write all 3 of the red points on our graph in the new basis by solving the same system simultaneously for all the points. Let $\mathbf{B}$ be a matrix containing the original coordinates of the points and let $\mathbf{A}$ be a matrix containing the new coordinates:

$$\mathbf{B} = \begin{pmatrix} -4 & 2 & 5 \\ -2 & 3 & 2 \end{pmatrix} \quad \mathbf{A} = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \end{pmatrix}$$

Then the new data coordinates on the rotated plane can be found by solving:

$$\mathbf{VA} = \mathbf{B}$$

And thus

$$\mathbf{A} = \mathbf{V}^T\mathbf{B} = \frac{\sqrt{2}}{2} \begin{pmatrix} -6 & 5 & 7 \\ -2 & -1 & 3 \end{pmatrix}$$

Using our new basis vectors, our alternative view of the data is that in Figure 9.3.

In the above example, we changed our basis from the original elementary basis to a new orthogonal basis which provides a different view of the data. All of this amounts to a rotation of the data around the origin. No real information has been lost - the points maintain their distances from each other in nearly every distance metric. **Our new variables, $\mathbf{v}_1$ and $\mathbf{v}_2$ are linear combinations of our original variables $\mathbf{e}_1$ and $\mathbf{e}_2$,** thus we can transform the data *back* to its original coordinate system by again solving a linear system (in this example, we'd simply multiply the new coordinates again by $\mathbf{V}$).

In general, we can change bases using the procedure outlined in Theorem 9.0.1.

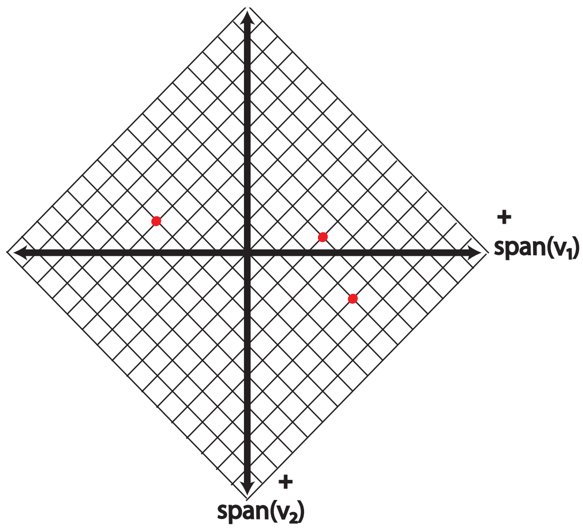


Figure 9.3: Points plotted in the new basis, $\mathcal{B}$

Theorem 9.0.1: Changing Bases

Given a matrix of coordinates (in columns), $\mathbf{A}$, in some basis, $\mathcal{B}_1 = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, we can change the basis to $\mathcal{B}_2 = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ with the new set of coordinates in a matrix $\mathbf{B}$ by solving the system

$$\mathbf{XA} = \mathbf{VB}$$

where $\mathbf{X}$ and $\mathbf{V}$ are matrices containing (as columns) the basis vectors from $\mathcal{B}_1$ and $\mathcal{B}_2$ respectively.
Note that when our original basis is the elementary basis, $\mathbf{X} = \mathbf{I}$, our system reduces to

$$\mathbf{A} = \mathbf{VB}.$$

When our new basis vectors are orthonormal, the solution to this system is simply

$$\mathbf{B} = \mathbf{V}^T\mathbf{A}.$$

Definition 9.0.2: Basis Terminology

A **basis** for the vector space $\mathbb{R}^n$ can be any collection of n linearly independent vectors in $\mathbb{R}^n$; n is said to be the **dimension** of the vector space $\mathbb{R}^n$. When the basis vectors are orthonormal (as they were in our example), the collection is called an **orthonormal basis**.

The preceding discussion dealt entirely with bases for $\mathbb{R}^n$ (our example was for points in $\mathbb{R}^2$). However, we will need to consider bases for *subspaces* of $\mathbb{R}^n$. Recall that the span of two linearly independent vectors in $\mathbb{R}^3$ is a plane. This plane is a 2-dimensional subspace of $\mathbb{R}^3$. Its dimension is 2 because 2 basis vectors are required to represent this space. However, not all points from $\mathbb{R}^3$ can be written in this basis - only those points which exist on the plane. In the next chapter, we will discuss how to proceed in a situation where the point we'd like to represent does not actually belong to the subspace we are interested in. This is the foundation for Least Squares.

Exercises

1. Show that the vectors $\mathbf{v}_1 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$ and $\mathbf{v}_2 = \begin{pmatrix} -2 \\ 6 \end{pmatrix}$ are orthogonal. Create an orthonormal basis for $\mathbb{R}^2$ using these two direction vectors.
2. Consider $\mathbf{a}_1 = (1, 1)$ and $\mathbf{a}_2 = (0, 1)$ as coordinates for points in the elementary basis. Write the coordinates of $\mathbf{a}_1$ and $\mathbf{a}_2$ in the orthonormal basis found in exercise 1. Draw a picture which reflects the old and new basis vectors.
3. Write the orthonormal basis vectors from exercise 1 as linear combinations of the original elementary basis vectors.
4. What is the length of the orthogonal projection of $\mathbf{a}_1$ onto $\mathbf{v}_1$?