11
Vector Spaces, Span, and Basis

11.1 Vector Spaces

Vector spaces are collections of vectors. The most common spaces are $\mathbb{R}^2$, $\mathbb{R}^3$, and $\mathbb{R}^n$ – the spaces that include all 2-, 3-, and n-dimensional vectors. We can construct *subspaces* by specifying only a subset of the vectors in a space. For example, the set of all 3-dimensional vectors with only integer entries is a subspace of $\mathbb{R}^3$.

Remember that $\mathbb{R}^2$ is not a subspace of $\mathbb{R}^3$; they are completely separate, non-overlapping spaces.

11.2 Span

A set of m vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ is said to *span* a space V if any vector $\mathbf{u}$ in V can be written as a linear combination of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$. This is equivalent to saying there exist scalars $a_1, a_2, \ldots, a_m$ such that

$$\mathbf{u} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_m\mathbf{v}_m$$

Writing a vector $\mathbf{u}$ as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ is called *decomposing* $\mathbf{u}$ over $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$. If a set of vectors spans a space, they can be used to decompose any other vector in the space.

We've already seen vector composition using a special set of vectors $\hat{\mathbf{e}}_j$, the unit vectors with only one nonzero entry at element j. For example, the vector

$$\begin{pmatrix} -2 \\ 4 \\ 5 \end{pmatrix} = -2\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 4\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + 5\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Thus the vectors $\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3$ spans $\mathbb{R}^3$. In general, the set of vectors $\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \ldots, \hat{\mathbf{e}}_n$ spans the space $\mathbb{R}^n$. Are these the only sets of vectors that span these spaces?

No, there are infinitely many sets of vectors that span each space. Consider the vectors $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$. We can show that these vectors span $\mathbb{R}^2$ by showing that any vector $\mathbf{u}$ in $\mathbb{R}^2$ can be written as a linear combination

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = a_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + a_2 \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

Finding the coefficients a_1 and a_2 is akin to solving the system of linear equations

$$\begin{aligned} a_1 - a_2 &= u_1 \\ a_1 + a_2 &= u_2 \end{aligned}$$

which has the unique solution

$$a_1 = \frac{u_1 + u_2}{2}, \quad a_2 = \frac{u_2 - u_1}{2}$$

To demonstrate, let $\mathbf{u} = \begin{pmatrix} -2 \\ 4 \end{pmatrix}$. Then $a_1 = 1$ and $a_2 = 3$. Then

$$a_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + a_2 \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + 3 \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1-3 \\ 1+3 \end{pmatrix} = \begin{pmatrix} -2 \\ 4 \end{pmatrix}$$

We've shown that there are least two sets of vectors that span $\mathbb{R}^2$, $\left\{\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right\}$ and $\left\{\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}\right\}$. How can we say there are infinitely many? If vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ span a space V, then the vectors $k_1\mathbf{v}_1$ and $k_2\mathbf{v}_2$ also span V for any scalars k_1 and k_2. To prove this, remember that any vector $\mathbf{u}$ can be decomposed onto $\mathbf{v}_1$ and $\mathbf{v}_2$, i.e.

$$\begin{aligned} \mathbf{u} &= a_1\mathbf{v}_1 + a_2\mathbf{v}_2 \\ &= \frac{a_1}{k_1}(k_1\mathbf{v}_1) + \frac{a_2}{k_2}(k_2\mathbf{v}_2) \end{aligned}$$

Since (a_1/k_1) and (a_2/k_2) are simply scalars, we've shown that $\mathbf{u}$ can be decomposed onto the vectors $k_1\mathbf{v}_1$ and $k_2\mathbf{v}_2$. Therefore, $k_1\mathbf{v}_1$ and $k_2\mathbf{v}_2$ must also span $\mathbb{R}^2$. For example, the vectors $\begin{pmatrix} 3 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ -1/2 \end{pmatrix}$ are scalar multiples of $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. The former two vectors must therefore span $\mathbb{R}^2$, so we can decompose the vector $\begin{pmatrix} -2 \\ 4 \end{pmatrix}$ onto them.

Since k_1 and k_2 are arbitrary, this allows us to generate infinitely many sets of vectors that span any space from a single spanning set.

$$\begin{pmatrix} -2 \\ 4 \end{pmatrix} = -\frac{2}{3}\begin{pmatrix} 3 \\ 0 \end{pmatrix} - 8\begin{pmatrix} 0 \\ 1/2 \end{pmatrix}$$

Similarly, if $\mathbf{v}_1$ and $\mathbf{v}_2$ span a space V, the vectors $\mathbf{v}_1$ and $(\mathbf{v}_1 + \mathbf{v}_2)$ also span V:

$$\begin{aligned} \mathbf{u} &= a_1\mathbf{v}_1 + a_2\mathbf{v}_2 \\ &= a_1\mathbf{v}_1 + a_2(\mathbf{v}_1 + \mathbf{v}_2) - a_2\mathbf{v}_1 \\ &= (a_1 - a_2)\mathbf{v}_1 + a_2(\mathbf{v}_1 + \mathbf{v}_2) \end{aligned}$$

If scalars a_1 and a_2 decompose $\mathbf{u}$ over $\mathbf{v}_1$ and $\mathbf{v}_2$, then $(a_1 - a_2)$ and a_2 decompose $\mathbf{u}$ over $\mathbf{v}_1$ and $(\mathbf{v}_1 + \mathbf{v}_2)$.

11.3 Review: Linear Independence

We said before that a set of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are *linearly independent* if and only if

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n = \mathbf{0}$$

implies that all coefficients $a_1, a_2, \ldots, a_n$ are zero. No linear combination of a set of linearly independent vectors can be the zero vector except for the trivial case where all the coefficients are zero. We often say that a set of vectors are linearly dependent if one of the vectors can be written as a linear combination of the others, i.e.

$$\mathbf{v}_i = a_1\mathbf{v}_1 + \cdots + a_{i-1}\mathbf{v}_{i-1} + a_{i+1}\mathbf{v}_{i+1} + \cdots + a_n\mathbf{v}_n$$

Moving the vector $\mathbf{v}_i$ to the right hand side

$$\mathbf{0} = a_1\mathbf{v}_1 + \cdots + a_{i-1}\mathbf{v}_{i-1} - \mathbf{v}_i + a_{i+1}\mathbf{v}_{i+1} + \cdots + a_n\mathbf{v}_n$$

we see 1.) a linear combination of the vectors sums to the zero vector on the left, and 2.) at least one of the coefficients (the -1 in front of $\mathbf{v}_i$) is nonzero. This is consistent with are above definition of linear independence. We said these vectors were not linearly independent, so it is possible for a linear combination to sum to zero using at least one nonzero coefficient.

11.4 Basis

The concepts of span and linear independence are a powerful combination. Any linearly independent set of vectors that span a space V are called a *basis* for V. Any vector in a space be decomposed over a set of vectors that span the space. **However, every vector in a space has a unique decomposition over an associated basis.** Said another way, for every vector in a space, there are only one set of coefficients $a_1, a_2, \ldots, a_n$ that decompose it over a basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$.

We can prove that a decomposition over a basis is unique by contradiction. Suppose there were two sets of coefficients – $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$ – that decomposed a vector $\mathbf{u}$ over a basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$. Then

$$\mathbf{u} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_n\mathbf{v}_n$$

We can move all the right hand size over to the left and group terms to give

$$(a_1 - b_1)\mathbf{v}_1 + (a_2 - b_2)\mathbf{v}_2 + \cdots + (a_n - b_n)\mathbf{v}_n = \mathbf{0}$$

Remember that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is a basis, so the vectors must be linearly independent. By the definition of linear independence, the only way the above equation can be true is if all the coefficients are zero. This implies that $a_1 = b_1$, $a_2 = b_2$, and so on. Clearly this is a violation of our original statement that $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots b_n$ where different. Therefore, there can only be one way to decompose any vector onto a basis.

Testing if vectors form a basis

Every basis for a vector space has the same number of vectors. This number is called the *dimension* of the vector space. For standard vectors spaces like $\mathbb{R}^n$, the dimension is n. – the dimension of $\mathbb{R}^2$ is 2, and the dimension of $\mathbb{R}^3$ is 3.

Most people think of dimension as the number of elements in a vector. While the true definition of dimension is the number of vectors in the basis, counting elements in a vector works for spaces like $\mathbb{R}^n$. To see why, remember that the Cartesian unit vectors $\hat{\mathbf{e}}_i$ form a basis for $\mathbb{R}^n$, but we need n of these vectors, one per element.

Any set of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is a basis for a space V if and only if:

1. The number of vectors (n) matches the dimension of V.

2. The vectors span V.

3. The vectors are linearly independent.

Proving any two of the above statements automatically implies the third is true.

We get to choose which two of the above three statements to prove when verifying that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is a basis. The first statement is usually trivial – does the number of vectors match the dimension? – so we almost always choose to prove the first statement. Proving that vectors are linearly independent is always easier that proving the vectors span the space. If we collect the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ into a matrix, the rank of this matrix should be n if the vectors are linearly independent.

Decomposing onto a basis

How do we decompose a vector onto a basis? Remember that decomposing $\mathbf{u}$ over $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is equivalent to finding a set of coefficients $a_1, a_2, \ldots, a_n$ such that

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n = \mathbf{u}$$

Let's collect the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ into a matrix $\mathbf{V}$, where each vector $\mathbf{v}_i$ is the ith column in $\mathbf{V}$. Then

$$\begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \mathbf{Va} = \mathbf{u}$$

We see that finding the coefficients $a_1, a_2, \ldots, a_n$ that decompose a vector $\mathbf{u}$ onto a basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is equivalent to solving the linear system $\mathbf{Va} = \mathbf{u}$.

By formulating vector decomposition as a linear system, we can easily see why the decomposition of a vector over a basis is unique. In $\mathbb{R}^n$, the basis contains n vectors, each with n elements. So, the matrix $\mathbf{V}$ is a square, $n \times n$ matrix. Since the vectors in the basis (and therefore the columns in $\mathbf{V}$) are linearly independent, the matrix $\mathbf{V}$ has full rank. Thus, the solution to $\mathbf{Va} = \mathbf{u}$ must be unique, implying that the decomposition of every vector onto a basis is unique.

Since $\mathbf{V}$ is square and full rank, it's inverse ($\mathbf{V}^{-1}$) must exist. The system must have a unique solution $\mathbf{a} = \mathbf{V}^{-1}\mathbf{u}$.

11.5 Orthogonal and Orthonormal Vectors

A set of vectors is an *orthogonal set* if every vector in the set is orthogonal to every other vector in the set. If every vector in an orthogonal set has been normalized, we say the vectors form an *orthonormal set*. Orthogonal and orthonormal sets are ideal candidates for basis vectors. Since there is no "overlap" among the vectors, we can easily decompose other vectors onto orthogonal basis vectors.

Imagine you have an orthogonal set of vectors you want to use as a basis. We'll assume you have the correct number of vectors (equal to the dimension of your space) for this to be possible. Based on the above requirements for basis vectors, we need only to show that these vectors are linearly independent. If so, they are a basis. For a set of n orthogonal vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, linear independence requires that

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n = \mathbf{0}$$

if and only if all the coefficients $a_1, a_2, \ldots, a_n$ are equal to zero. Let's take the dot product of both sides of the above equation with the vector $\mathbf{v}_1$

$$(a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n = \mathbf{0}) \cdot \mathbf{v}_1 = \mathbf{0} \cdot \mathbf{v}_1$$

On the right hand size, we know that $\mathbf{0} \cdot \mathbf{v}_1 = 0$ for any vector $\mathbf{v}_1$. We also distribute the dot product on the left hand side to give

$$a_1\mathbf{v}_1 \cdot \mathbf{v}_1 + a_2\mathbf{v}_2 \cdot \mathbf{v}_1 + \cdots + a_n\mathbf{v}_n \cdot \mathbf{v}_1 = 0$$

Since all the vectors are orthogonal, $\mathbf{v}_i \cdot \mathbf{v}_1$ is zero except when $i = 1$. Canceling out all the dot products equal to zero shows that

$$a_1 \mathbf{v}_1 \cdot \mathbf{v}_1 = a_1 \|\mathbf{v}_1\|^2 = 0$$

We know that $\|\mathbf{v}_1\|^2 \neq 0$, so the only way $a_1 \mathbf{v}_1 \cdot \mathbf{v}_1$ can equal zero is if a_1 is zero. If we repeat this entire process by taking the dot product with $\mathbf{v}_2$ instead of $\mathbf{v}_1$, we will find that $a_2 = 0$. This continues with $\mathbf{v}_3, \ldots, \mathbf{v}_n$ until we can say that $a_1 = a_2 = \cdots = a_n = 0$. Therefore, if the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are an orthogonal (or orthonormal) set, they are linearly independent.

Decomposing onto orthonormal vectors

We saw previously that finding the coefficients to decompose a vector onto a basis requires solving a system of linear equations. For high-dimensional spaces, solving such a system can be computationally expensive. Fortunately, decomposing a vector onto an orthonormal basis is far more efficient.

Theorem. *The decomposition of a vector* $\mathbf{u}$ *onto an orthonormal basis* $\hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2, \ldots, \hat{\mathbf{v}}_n$ *given by*

$$\mathbf{u} = a_1 \hat{\mathbf{v}}_1 + a_2 \hat{\mathbf{v}}_2 + \cdots + a_n \hat{\mathbf{v}}_n$$

has coefficients

$$\begin{aligned} a_1 &= \mathbf{u} \cdot \hat{\mathbf{v}}_1 \\ a_2 &= \mathbf{u} \cdot \hat{\mathbf{v}}_2 \\ &\vdots \\ a_n &= \mathbf{u} \cdot \hat{\mathbf{v}}_n \end{aligned}$$

We use the symbol $\hat{\mathbf{v}}_i$ for vectors in an orthonormal set to remind us that each vector has been normalized.

Proof. We use a similar strategy as when we proved the linear independence of orthogonal sets. Let's start with the formula for vector decomposition

$$\mathbf{u} = a_1 \hat{\mathbf{v}}_1 + a_2 \hat{\mathbf{v}}_2 + \cdots + a_n \hat{\mathbf{v}}_n$$

Taking the dot product of both sides with the vector $\hat{\mathbf{v}}_1$ yields (after distributing the right hand side)

$$\mathbf{u} \cdot \hat{\mathbf{v}}_1 = a_1 \hat{\mathbf{v}}_1 \cdot \hat{\mathbf{v}}_1 + a_2 \hat{\mathbf{v}}_2 \cdot \hat{\mathbf{v}}_1 + \cdots + a_n \hat{\mathbf{v}}_n \cdot \hat{\mathbf{v}}_1$$

Because all the vectors $\hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2, \ldots, \hat{\mathbf{v}}_n$ are orthogonal, the only nonzero term on the right hand side is $a_1 \hat{\mathbf{v}}_1 \cdot \hat{\mathbf{v}}_1$, so

$$\mathbf{u} \cdot \hat{\mathbf{v}}_1 = a_1 \hat{\mathbf{v}}_1 \cdot \hat{\mathbf{v}}_1$$

By definition of the dot product, $\hat{\mathbf{v}}_1 \cdot \hat{\mathbf{v}}_1 = \|\hat{\mathbf{v}}_1\|^2$. Since $\hat{\mathbf{v}}_1$ is a unit vector, $\|\hat{\mathbf{v}}_1\|^2 = 1$. Thus $a_1 = \mathbf{u} \cdot \hat{\mathbf{v}}_1$. By repeating the same procedure with $\hat{\mathbf{v}}_2$, we find that $a_2 = \mathbf{u} \cdot \hat{\mathbf{v}}_2$, and so on. $\square$

Decomposing vectors over an orthonormal basis is efficient, requiring only a series of dot products to compute the coefficients. For example, we can decompose the vector $\mathbf{u} = \begin{pmatrix} 7 \\ -5 \\ 10 \end{pmatrix}$ over the orthonormal basis $\left\{ \hat{\mathbf{v}}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \hat{\mathbf{v}}_2 = \begin{pmatrix} 0 \\ 3/5 \\ 4/5 \end{pmatrix}, \hat{\mathbf{v}}_3 = \begin{pmatrix} 0 \\ 4/5 \\ -3/5 \end{pmatrix} \right\}$.

$$a_1 = \mathbf{u} \cdot \hat{\mathbf{v}}_1 = \begin{pmatrix} 7 \\ -5 \\ 10 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = 7 + 0 + 0 = 7$$

$$a_2 = \mathbf{u} \cdot \hat{\mathbf{v}}_2 = \begin{pmatrix} 7 \\ -5 \\ 10 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 3/5 \\ 4/5 \end{pmatrix} = 0 - 3 + 8 = 5$$

$$a_3 = \mathbf{u} \cdot \hat{\mathbf{v}}_3 = \begin{pmatrix} 7 \\ -5 \\ 10 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 4/5 \\ -3/5 \end{pmatrix} = 0 - 4 - 6 = -10$$

The decomposition of $\mathbf{u}$ is

$$\mathbf{u} = 7\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 5\begin{pmatrix} 0 \\ 3/5 \\ 4/5 \end{pmatrix} - 10\begin{pmatrix} 0 \\ 4/5 \\ -3/5 \end{pmatrix} = \begin{pmatrix} 7+0+0 \\ 0+3-8 \\ 0+4+6 \end{pmatrix} = \begin{pmatrix} 7 \\ -5 \\ 10 \end{pmatrix}$$

Checking an orthonormal set

Given a set of vectors $\hat{\mathbf{v}}_1, \hat{\mathbf{v}}_2, \ldots, \hat{\mathbf{v}}_n$, how can we verify that they are orthonormal? We need to test two things.

1. All vectors are normalized ($\|\hat{\mathbf{v}}_i\| = 1$ for all $\hat{\mathbf{v}}_i$).
2. All vectors are mutually orthogonal ($\hat{\mathbf{v}}_i \cdot \hat{\mathbf{v}}_j = 0$ for all $i \neq j$).

The first test is straightforward. The second can be a little cumbersome, as we need to test all $n^2 - n/2$ pairs of vectors for orthogonality. A simpler, albeit more sometimes more computationally intensive approach, is to collect the vectors into a matrix $\mathbf{V} = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{pmatrix}$. Then the set of vectors is orthonormal if an only if $\mathbf{V}^{-1} = \mathbf{V}^\mathsf{T}$. While inverting a matrix is “expensive", for small to medium size vector sets this method avoids the need to iterate over all pairs of vectors to test for orthonormality.

Interestingly, the proof of this method (not shown here) reveals that if the columns of $\mathbf{V}$ are an orthonormal set, the rows of $\mathbf{V}$ are also an orthonormal set!

Projections

Our next goal will be to create orthonormal sets of vectors from sets that are not orthogonal. Before introducing such an algorithm, we need to develop a geometric tool – the vector *projection*. The projection of vector $\mathbf{v}$ onto vector $\mathbf{u}$ is a vector that points along $\mathbf{u}$ with length equal to the "shadow" of $\mathbf{v}$ onto $\mathbf{u}$. Previously we used the dot product to calculate the magnitude of the projection of $\mathbf{v}$ onto $\mathbf{u}$, which was a scalar equal to $\|\mathbf{v}\| \cos\theta$, where θ is the angle between $\mathbf{v}$ and $\mathbf{u}$. To calculate the actual projection, we multiply the magnitude of the projection ($\|\mathbf{v}\|\cos\theta$) by a unit vector that points along $\mathbf{u}$. Thus the projection of $\mathbf{v}$ onto $\mathbf{u}$ is defined as

$$\mathrm{proj}_{\mathbf{u}}(\mathbf{v}) = (\|\mathbf{v}\| \cos\theta)\hat{\mathbf{u}}$$

By definition, $\hat{\mathbf{u}} = \mathbf{u}/\|\mathbf{u}\|$. Also, we note that $\mathbf{v}\cdot\mathbf{u} = \|\mathbf{v}\|\,\|\mathbf{u}\|\cos\theta$, so the expression $\|\mathbf{u}\|\cos\theta$ can be written in terms of the dot product $(\mathbf{v}\cdot\mathbf{u})/\|\mathbf{u}\|$. We can rewrite our formula for the projection using only dot products:

$$\mathrm{proj}_{\mathbf{u}}(\mathbf{v}) = (\|\mathbf{v}\|\cos\theta)\hat{\mathbf{u}} = \frac{\mathbf{v}\cdot\mathbf{u}}{\|\mathbf{u}\|}\frac{\mathbf{u}}{\|\mathbf{u}\|} = \frac{\mathbf{v}\cdot\mathbf{u}}{\|\mathbf{u}\|^2}\mathbf{u} = \frac{\mathbf{v}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}$$

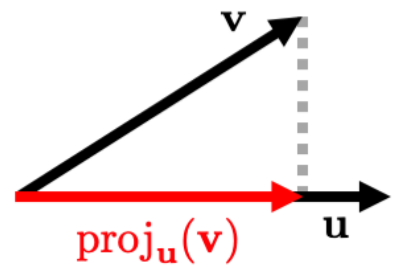

Figure 11.1: The projection is a vector "shadow" of one vector onto another.

We can use the projection to make any two vectors orthogonal, as demonstrated by the following theorem.

Theorem. *Given any vectors* $\mathbf{v}$ *and* $\mathbf{u}$*, the vector*

$$\mathbf{v} - \mathrm{proj}_{\mathbf{u}}(\mathbf{v})$$

is orthogonal to $\mathbf{u}$*.*

Proof. If the vector $\mathbf{v} - \mathrm{proj}_{\mathbf{u}}(\mathbf{v})$ is orthogonal to $\mathbf{u}$, the dot product between these vectors must be zero.

$$\begin{aligned}
\left(\mathbf{v} - \mathrm{proj}_{\mathbf{u}}(\mathbf{v})\right)\cdot\mathbf{u} &= \left(\mathbf{v} - \frac{\mathbf{v}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}\right)\cdot\mathbf{u} \\
&= \mathbf{v}\cdot\mathbf{u} - \frac{\mathbf{v}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}\cdot\mathbf{u} \\
&= \mathbf{v}\cdot\mathbf{u} - \mathbf{v}\cdot\mathbf{u} \\
&= 0
\end{aligned}$$

□

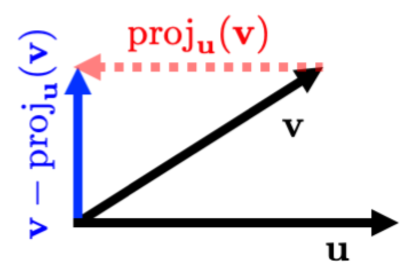

Figure 11.2: Subtracting the projection of $\mathbf{v}$ onto $\mathbf{u}$ from $\mathbf{v}$ makes the vectors orthogonal.

Subtracting the projection of $\mathbf{v}$ onto $\mathbf{u}$ from the vector $\mathbf{v}$ "corrects" $\mathbf{v}$ by removing its overlap with $\mathbf{u}$. The resulting vector is a vector closest to $\mathbf{v}$ that is still orthogonal to $\mathbf{u}$.

Creating orthonormal basis vectors

We can make any two vectors orthogonal by adjusting one based on its projection onto the other. We can apply these corrections iteratively to make any set of linearly independent vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ into an orthonormal basis set $\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, \ldots, \hat{\mathbf{u}}_n$. First, we set

We must begin with linearly independent vectors. Otherwise, orthogonalization will turn one of the vectors into the zero vector, which is not allowed in a basis.

$$\mathbf{u}_1 = \mathbf{v}_1$$

We leave this first vector unchanged. All other vectors will be made orthogonal to it (and each other). Next, we create $\mathbf{u}_2$ by making $\mathbf{v}_2$ orthogonal to $\mathbf{u}_1$:

$$\mathbf{u}_2 = \mathbf{v}_2 - \text{proj}_{\mathbf{u}_1}(\mathbf{v}_2)$$

Now we have two orthogonal vectors, $\mathbf{u}_1$ and $\mathbf{u}_2$. We continue by creating $\mathbf{u}_3$ from $\mathbf{v}_3$, but this time we must make $\mathbf{v}_3$ orthogonal to both $\mathbf{u}_1$ and $\mathbf{u}_2$:

$$\mathbf{u}_3 = \mathbf{v}_3 - \text{proj}_{\mathbf{u}_1}(\mathbf{v}_3) - \text{proj}_{\mathbf{u}_2}(\mathbf{v}_3)$$

We continue this process for all n vectors, making each vector $\mathbf{v}_i$ orthogonal to all the newly created orthogonal vectors $\mathbf{u}_1, \ldots, \mathbf{u}_{i-1}$.

This approach is called the Gram-Schmidt algorithm. Given a set of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, we create a set of orthogonal vectors

Said more succinctly,

$$\mathbf{u}_i = \mathbf{v}_i - \sum_{k=1}^{i-1} \text{proj}_{\mathbf{u}_k}(\mathbf{v}_i)$$

$$\begin{aligned}
\mathbf{u}_1 &= \mathbf{v}_1 \\
\mathbf{u}_2 &= \mathbf{v}_2 - \text{proj}_{\mathbf{u}_1}(\mathbf{v}_2) \\
\mathbf{u}_3 &= \mathbf{v}_3 - \text{proj}_{\mathbf{u}_1}(\mathbf{v}_3) - \text{proj}_{\mathbf{u}_2}(\mathbf{v}_3) \\
&\vdots \\
\mathbf{u}_i &= \mathbf{v}_i - \text{proj}_{\mathbf{u}_1}(\mathbf{v}_i) - \cdots - \text{proj}_{\mathbf{u}_{i-1}}(\mathbf{v}_i) \\
&\vdots \\
\mathbf{u}_n &= \mathbf{v}_n - \text{proj}_{\mathbf{u}_1}(\mathbf{v}_n) - \cdots - \text{proj}_{\mathbf{u}_{n-1}}(\mathbf{v}_n)
\end{aligned}$$

The Gram-Schmidt algorithm products an orthogonal set of vectors. To make the set orthonormal, we must subsequently normalize each vector.