Definitions and examples

Definition 1. We say vectors $\vec{v}, \vec{w} \in \mathbb{R}^n$ are called orthogonal if $\vec{v} \cdot \vec{w} = 0$. A set of vectors $S = \{\vec{v}_1, \ldots, \vec{v}_k\}$ is called orthogonal if $\vec{v}_i \cdot \vec{v}_j = 0$ for $i \neq j$, i.e. all vectors are orthogonal to each other. A set of vectors $S = \{\vec{v}_1, \ldots, \vec{v}_k\}$ is called orthonormal if $S$ is orthogonal and all $\vec{v}_i$ are unit vectors, i.e. $\|\vec{v}_i\| = 1$ for all $i$.

Example 2. The set $S = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 1 \\ -1 \end{pmatrix} \right\}$ is an orthogonal subset of $\mathbb{R}^4$. The set $S = \left\{ \begin{pmatrix} 1/2 \\ 1/2 \\ -1/2 \\ -1/2 \end{pmatrix}, \begin{pmatrix} 1/2 \\ -1/2 \\ 1/2 \\ 1/2 \end{pmatrix}, \begin{pmatrix} -1/2 \\ -1/2 \\ 1/2 \\ 1/2 \end{pmatrix} \right\}$ is an orthonormal subset of $\mathbb{R}^4$. The set $S = \left\{ \begin{pmatrix} 1/2 \\ 1/2 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1/2 \\ 0 \\ 0 \\ 1/2 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ -1/\sqrt{2} \end{pmatrix} \right\}$ is neither orthogonal or orthonormal (but all vectors are unit vectors).

The reason why we care about orthogonal/orthonormal subsets is because they make two important computations in linear algebra very easy: if $\beta$ is an orthonormal basis of $\mathbb{R}^n$ (i.e. a basis of $\mathbb{R}^n$ that’s orthonormal) computing the $\beta$-coordinates $[\vec{x}]_\beta$ of some vector $\vec{x} \in \mathbb{R}^n$ becomes very simple. Additionally, a set of orthogonal/orthonormal vectors is automatically linearly independent.

Theorem 3. Let $S = \{\vec{u}_1, \ldots, \vec{u}_k\}$ be an orthonormal subset of $\mathbb{R}^n$.

1. $S$ is linearly independent
2. If $\vec{v} \in \text{Span}(S)$, then $\vec{x} = (\vec{x} \cdot \vec{u}_1)\vec{u}_1 + \ldots + (\vec{x} \cdot \vec{u}_k)\vec{u}_k$

Proof. 1. Suppose that $c_1\vec{u}_1 + \ldots + c_k\vec{u}_k = \vec{0}$ for some $c_i \in \mathbb{R}$. Since the $\vec{u}_i$ are orthogonal, we can recover the coefficient $c_i$ by dotting both sides with $\vec{u}_i$: $\vec{x} \cdot \vec{u}_i = (c_1\vec{u}_1 + \ldots + c_k\vec{u}_k) \cdot \vec{u}_i = c_i(\vec{u}_i \cdot \vec{u}_i) = c_i = 0$, where $\vec{u}_i \cdot \vec{u}_i = 1$ because $\vec{u}_i$ are unit vectors. This then says all $c_i = 0$, so the vectors are linearly independent.

2. If $\vec{x} \in \text{Span}(S)$, we can write $\vec{x} = c_1\vec{u}_1 + \ldots + c_k\vec{u}_k$ for some coefficients $c_i$. To determine the values $c_i$, we do the same thing as above: dotting both sides with $\vec{u}_i$, $\vec{x} \cdot \vec{u}_i = (c_1\vec{u}_1 + \ldots + c_k\vec{u}_k) \cdot \vec{u}_i = c_i(\vec{u}_i \cdot \vec{u}_i) = c_i$. This proves what we want. \hfill \Box

One important rephrasing of the second statement is the following:
Theorem 4. Let $\beta = \{\vec{u}_1, \ldots, \vec{u}_n\}$ be an orthonormal basis of $\mathbb{R}^n$. For $\vec{x} \in \mathbb{R}^n$, we have $[\vec{x}]_\beta = \begin{pmatrix} \vec{x} \cdot \vec{u}_1 \\ \vdots \\ \vec{x} \cdot \vec{u}_n \end{pmatrix}$

Orthogonal Projections

We saw earlier how to construct the orthogonal projection of a vector $\vec{x} \in \mathbb{R}^2$ onto a line $L$, as well as how to project a vector $\vec{x}$ onto either a line $L$ or a plane $S$ in $\mathbb{R}^3$. We now give a general construction for how to project a vector $\vec{x}$ onto a $k$-dimensional subspace of $\mathbb{R}^n$.

Definition 5. Let $S$ be a $k$-dimensional subspace of $\mathbb{R}^n$. The orthogonal complement of $S$, $S^\perp$, is defined to be the set of vectors in $\mathbb{R}^n$ that are orthogonal to every vector in $S$.

Example 6. If $S$ is a line in $\mathbb{R}^3$, then $S^\perp$ is the plane orthogonal to that line. The set $S^\perp$ has the following properties: (of which we will only need the last for now)

Theorem 7. Let $S \subset \mathbb{R}^n$ be a $k$-dimensional subspace.

1. $S^\perp$ is a subspace of $\mathbb{R}^n$.
2. $\dim(S^\perp) = n - k$
3. $(S^\perp)^\perp = S$
4. $\mathbb{R}^n = S \oplus S^\perp$, i.e. every vector $\vec{x} \in \mathbb{R}^n$ is of the form $\vec{x} = \vec{x}_S + \vec{x}_\perp$ where $\vec{x}_S \in S$ and $\vec{x}_\perp \in S^\perp$.

Definition 8. The vector $\vec{x}_S$ (sometimes denoted $\text{Proj}_S(\vec{x})$) in the above theorem is called the orthogonal projection of $\vec{x}$ onto $S$.

We can compute $\vec{x}_S$ as follows. Suppose we have an orthonormal basis $\{\vec{u}_1, \ldots, \vec{u}_k\}$ of $S$. Then we can write $\vec{x} = \vec{x}_S + \vec{x}_\perp$, and since $\vec{x}_S \in S$, we can write $\vec{x}_S = c_1 \vec{u}_1 + \ldots + c_k \vec{u}_k$. So $\vec{x} = c_1 \vec{u}_1 + \ldots + c_k \vec{u}_k + \vec{x}_\perp$. Since $\vec{x}_\perp \in S^\perp$, it’s orthogonal to all vectors in $S$, and in particular, the vectors $\vec{u}_i$. So dotting both sides with $\vec{u}_i$ says $\vec{x} \cdot \vec{u}_i = (c_1 \vec{u}_1 + \ldots + c_k \vec{u}_k + \vec{x}_\perp) \cdot \vec{u}_i = c_i (\vec{u}_i \cdot \vec{u}_i) = c_i$. This gives us the following formula:

Theorem 9. Let $S \subset \mathbb{R}^n$ be a $k$-dimensional subspace with orthonormal basis $\{\vec{u}_1, \ldots, \vec{u}_k\}$. Then $\text{Proj}_S(\vec{x}) = (\vec{x} \cdot \vec{u}_1)\vec{u}_1 + \ldots + (\vec{x} \cdot \vec{u}_k)\vec{u}_k$.

Example 10. Let $S = \text{Span}\left\{ \begin{pmatrix} 1 \\ 7 \\ 1 \\ 7 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \end{pmatrix} \right\}$. Let’s compute $\text{Proj}_S(\vec{x})$ for arbitrary $\vec{x} \in \mathbb{R}^4$. Notice that $S$ is spanned by an orthogonal set, but not an orthonormal set. To make it orthonormal, we just need to normalize the vectors. Then $\beta = \left\{ \begin{pmatrix} 1/10 \\ 7/10 \\ 1/10 \\ 7/10 \end{pmatrix}, \begin{pmatrix} -1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ -1/\sqrt{2} \end{pmatrix} \right\}$.

{\vec{u}_1, \vec{u}_2, \vec{u}_3} is an orthonormal basis for $S$. To compute $\text{Proj}_S(\vec{x})$, we first find the matrix of the orthogonal projection. We see that $\text{Proj}_S(\vec{e}_1) = (\vec{e}_1 \cdot \vec{u}_1)\vec{u}_1 + (\vec{e}_1 \cdot \vec{u}_2)\vec{u}_2 + (\vec{e}_1 \cdot \vec{u}_3)\vec{u}_3 = \begin{pmatrix} 51/100 \\ 7/100 \\ -49/100 \\ 7/100 \end{pmatrix}$. Similarly, we compute $\text{Proj}_S(\vec{e}_2) = \begin{pmatrix} 7/10 \\ -99/100 \\ 51/100 \\ 7/100 \end{pmatrix}$, $\text{Proj}_S(\vec{e}_3) = \begin{pmatrix} 7/10 \\ -49/100 \\ 7/100 \\ -1/100 \end{pmatrix}$, and $\text{Proj}_S(\vec{e}_4) = \begin{pmatrix} 7/10 \\ 7/100 \\ 51/100 \\ 7/100 \end{pmatrix}$.
What is the minimal distance from a point \( \vec{x} \) along the line through \( \vec{x} \)? Hopefully, it's pretty clear geometrically that the shortest path to the plane is obtained by traveling parallel to the plane. We can extend the orthonormal basis \( \{ \vec{u}_1, \vec{u}_2, \vec{u}_3 \} \) of \( S \) by choosing \( \vec{u}_4 = (\vec{x} - \text{Proj}_S(\vec{x}))/\|\vec{x} - \text{Proj}_S(\vec{x})\| \) for any vector \( \vec{x} \in \mathbb{R}^4 \) (because \( \vec{x} - \text{Proj}_S(\vec{x}) \in S^\perp \), so after normalizing we have a set of 4 orthonormal vectors which forces it to be a basis!), although the vector \( \vec{u}_4 \) will end up being irrelevant. The matrix of \( P \) in \( \beta \)-coordinates, \( P_\beta \), is given by

\[
P_\beta = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0
\end{pmatrix}.
\]

If \( E \) is the standard basis of \( \mathbb{R}^4 \), the change of basis matrix \( S_\beta^E \) is simply the matrix \( \begin{pmatrix}
\vec{u}_1 & \vec{u}_2 & \vec{u}_3 & \vec{u}_4 \\
\end{pmatrix} \), and we have \( S_\beta^E (S_\beta^E)^{-1} = (S_\beta^E)^t = \begin{pmatrix}
- \vec{u}_1 & - \\
- \vec{u}_2 & - \\
- \vec{u}_3 & - \\
- \vec{u}_4 & -
\end{pmatrix} \).

The change of basis formula then says that \( P = S_\beta^E P_\beta S_\beta^E \), \( P = \vec{u}_1 \vec{u}_1^t + \vec{u}_2 \vec{u}_2^t + \vec{u}_3 \vec{u}_3^t \) (verify this is how the matrix multiplication works out if this is not clear!). In general, we get the following:

**Theorem 11.** Let \( S \) be a \( k \)-dimensional subspace of \( \mathbb{R}^n \) with orthonormal basis \( \{ \vec{u}_1, \ldots, \vec{u}_k \} \). The matrix of the orthogonal projection \( \text{Proj}_S \) is given by \( P = \vec{u}_1 \vec{u}_1^t + \ldots + \vec{u}_k \vec{u}_k^t \).

As an application of projections, we give the following geometric example:

**Example 12.** What is the minimal distance from a point \( \vec{x} \in \mathbb{R}^3 \) to the plane \( S : x + 2y + z = 0 \)? It is hopefully pretty clear geometrically that the shortest path to the plane is obtained by traveling along the line through \( \vec{x} \) that’s orthogonal to the plane, i.e., the minimal distance is given by \( \|\vec{x} - \text{Proj}_S(\vec{x})\| = \|\vec{x} - \vec{a}\| \). It turns out an orthonormal basis of \( S \) is given by \( \begin{pmatrix}
-2/\sqrt{5} \\
1/\sqrt{5} \\
0
\end{pmatrix} \), \( \begin{pmatrix}
1/\sqrt{30} \\
2/\sqrt{30} \\
-5/\sqrt{30}
\end{pmatrix} \).

The above theorem says that the matrix \( P \) of the projection is given by \( \begin{pmatrix}
5/6 & -1/3 & -1/3 \\
-1/3 & 1/3 & -1/3 \\
-1/6 & -1/3 & 5/6
\end{pmatrix} \).

If we write \( \vec{x} = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \), it turns out that \( \|\vec{x} - \text{Proj}_S(\vec{x})\| = \|(I_3 - P)\vec{x}\| = \sqrt{\frac{a^2 + 2b + c}{6}} \) (if you do out the computation).

**Finding orthonormal bases: the Gram-Schmidt algorithm**

Hopefully the above section illustrates why we would want to work with an orthonormal basis of some subspace \( S \) of \( \mathbb{R}^n \). How can we find one? The answer lies in the following algorithm:

**Theorem 13** (Gram-Schmidt Process). Let \( S \subset \mathbb{R}^n \) be a \( k \)-dimensional subspace with basis \( \{ \vec{v}_1, \ldots, \vec{v}_k \} \). Then the set \( \{ \vec{u}_1, \ldots, \vec{u}_k \} \) is an orthonormal basis of \( S \), where \( \vec{u}_i \) is defined by the
following:
\[ \vec{u}_1 = \vec{v}_1 / \|\vec{v}_1\| \]
\[ \vec{u}_2 = \vec{v}_2 / \|\vec{v}_2\|, \text{ where } \vec{v}_2^\perp = \vec{v}_2 - (\vec{v}_2 \cdot \vec{u}_1)\vec{u}_1 \]
\[ \vdots \]
\[ \vec{u}_k = \vec{v}_k / \|\vec{v}_k\|, \text{ where } \vec{v}_k^\perp = \vec{v}_k - (\vec{v}_k \cdot \vec{u}_1)\vec{u}_1 - \cdots - (\vec{v}_k \cdot \vec{u}_{k-1})\vec{u}_{k-1} \]

and in general, \( \vec{u}_i = \vec{v}_i / \|\vec{v}_i\| \) where \( \vec{v}_i^\perp = \vec{v}_i - (\vec{v}_i \cdot \vec{u}_1)\vec{u}_1 - \cdots - (\vec{v}_i \cdot \vec{u}_{i-1})\vec{u}_{i-1} \)

**Example 14.** Let \( L = \text{Span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\} = \text{Span}\{\vec{v}\} \) be a line in \( \mathbb{R}^3 \). Then \( L^\perp \) is a 3-dimensional subspace of \( \mathbb{R}^3 \). What’s an orthonormal basis of \( L^\perp \)? To compute this, we first find a basis of \( L^\perp \) and then run Gram-Schmidt. Any vector \( \vec{x} = \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} \in L^\perp \) must satisfy the equation \( \vec{x} \cdot \vec{v} = 0 \), i.e. \( a + b + c + d = 0 \). Solving the system says \( a = -b - c - d \) with \( b, c, d \) free, so we find that a basis of \( L^\perp \) is given by \( \beta = \left\{ \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right\} = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\} \). We now run Gram-Schmidt. We have \( \vec{u}_1 = \vec{v}_1 / \|\vec{v}_1\| = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \\ 0 \end{pmatrix} \). Then \( \vec{u}_2 = \vec{v}_2 - (\vec{v}_2 \cdot \vec{u}_1)\vec{u}_1 = \begin{pmatrix} 1/\sqrt{6} \\ -1/\sqrt{6} \\ 2/\sqrt{6} \\ 0 \end{pmatrix} \), so that \( \vec{u}_2 = \vec{v}_2 / \|\vec{v}_2\| = \begin{pmatrix} 2 \sqrt{6}^{-1/2} \\ -1 \sqrt{6}^{-1/2} \\ 2 \sqrt{6}^{-1/2} \\ 0 \end{pmatrix} \). Finally, we have \( \vec{v}_3 = \vec{v}_3 - (\vec{v}_3 \cdot \vec{u}_1)\vec{u}_1 - (\vec{v}_3 \cdot \vec{u}_2)\vec{u}_2 - (\vec{v}_3 \cdot \vec{u}_3)\vec{u}_3 \)

\[ \vec{u}_3 = \vec{v}_3 - \frac{1}{\sqrt{2}} \vec{u}_1 - \frac{1}{\sqrt{6}} \vec{u}_3 = \begin{pmatrix} -1/3 \\ -1/3 \\ -1/3 \\ 1 \end{pmatrix} \), so that \( \vec{u}_3 = \vec{v}_3 / \|\vec{v}_3\| = \begin{pmatrix} -\sqrt{3}/6 \\ -\sqrt{3}/6 \\ -\sqrt{3}/6 \\ \sqrt{3}/2 \end{pmatrix} \). This says that

\[ \beta = \left\{ \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -1/\sqrt{6} \\ -1/\sqrt{6} \\ 2/\sqrt{6} \\ 0 \end{pmatrix}, \begin{pmatrix} -\sqrt{3}/6 \\ -\sqrt{3}/6 \\ -\sqrt{3}/6 \\ \sqrt{3}/2 \end{pmatrix} \right\} \] is an orthonormal basis of \( L^\perp \).