The Gram-Schmidt Process

The Gram-Schmidt Process is an algorithm used to convert any basis for a subspace $W$ of $\mathbb{R}^n$ into a new orthogonal basis for $W$. This orthogonal basis can then be normalized, if desired, to get an orthonormal basis for $W$. Since every nonzero subspace $W$ has a basis to which the Gram-Schmidt Process can be applied, this means that every nonzero subspace of $\mathbb{R}^n$ has an orthonormal basis.

It is important to think about what the algorithm is doing: that makes it much easier to remember than just memorizing formulas. Therefore this presentation, while doing the very same thing as in the text, separates the idea of the Gram-Schmidt Process from the calculations.

**Lemma** Suppose $\{v_1, \ldots, v_i\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$ and $W = \text{Span} \{v_1, \ldots, v_i\}$. Pick a vector $x \notin W$ and write

$$x = \bar{x} + v$$

where $\bar{x} = \text{proj}_W x \in W$ and $v = x - \bar{x} = x - \text{proj}_W x \in W^\perp$.

The enlarged set $\{v_1, \ldots, v_i, x\}$ is orthogonal (because the $v_i$'s are in $W$ and $v$ is in $W^\perp$), and

$$\text{Span} \{v_1, \ldots, v_i, x\} = \text{Span} \{v_1, \ldots, v_i, v\}$$

**NOTE:** Here, I'm using the letter $v$ (rather than the letter $z$ as in the Orthogonal Decomposition Theorem). In this context, I think that makes the discussion read more smoothly.

Informally, what does the Lemma say?

The $v_i$'s are orthogonal, so $\{v_1, \ldots, v_i\}$ is a linearly independent set, therefore a basis for $W$. Suppose we add a new vector $x$ from “outside $W”$ to the basis, so the subspace $W$ is enlarged to $\text{Span} \{v_1, \ldots, v_i, x\}$. Since $x$ is not a linear combination of $v_1, \ldots, v_i$, the set $\{v_1, \ldots, v_i, x\}$ is linearly independent and is a basis for the enlarged subspace $\text{Span} \{v_1, \ldots, v_i, x\}$.

But the “$\bar{x}$ piece” of $x$ was already in $W$; what actually enlarges $W$ is adding $v$, the “piece” of $x$ from $W^\perp$). The lemma says that just adding $v$ to $\{v_1, \ldots, v_i\}$ creates the same enlarged subspace as adding “the whole vector” $x$ to $\{v_1, \ldots, v_i\}$: that is,

$$\text{Span} \{v_1, \ldots, v_i, x\} = \text{Span} \{v_1, \ldots, v_i, v\}$$

and, since $v \in W^\perp$, the new set $\{v_1, \ldots, v_i, v\}$ is an orthogonal basis for the enlarged subspace.

**Proof** To show that $\text{Span} \{v_1, \ldots, v_i, x\} = \text{Span} \{v_1, \ldots, v_i, v\}$, we need to show that

a) every vector $w$ in $\text{Span} \{v_1, \ldots, v_i, x\}$ is also in $\text{Span} \{v_1, \ldots, v_i, v\}$, and (vice-versa)

b) every vector $w$ in $\text{Span} \{v_1, \ldots, v_i, v\}$ is also in $\text{Span} \{v_1, \ldots, v_i, x\}$

For short, let’s use the abbreviated notation $\text{lc}(v_1, \ldots, v_i)$ to mean “a linear combination of $v_1, \ldots, v_i$” without being specific about the weights involved in the linear combination.

In this shorthand, for example, “$\text{lc}(v_1, \ldots, v_i) + \text{lc}(v_1, \ldots, v_i) = \text{lc}(v_1, \ldots, v_i)$”. This just says that the sum of two linear combinations of $v_1, \ldots, v_i$ is a new linear combination of $v_1, \ldots, v_i$
Since \( \mathcal{E} \) is in \( W \), we know that \( \mathcal{E} = \text{lc}(v_1, \ldots, v_i) \).

To show a) If \( w \) is in \( \text{Span}\{v_1, \ldots, v_i, x\} \), then
\[
\begin{align*}
  w &= \text{lc}(v_1, \ldots, v_i) + cx \\
  &= \text{lc}(v_1, \ldots, v_i) + c(\mathcal{E} + v) \\
  &= \text{lc}(v_1, \ldots, v_i) + c\mathcal{E} + cv \\
  &= \text{lc}(v_1, \ldots, v_i) + c\mathcal{E} + \text{lc}(v_1, \ldots, v_i) + cv \\
  &= \text{lc}(v_1, \ldots, v_i) + cv, \text{ so} \\
\end{align*}
\]
so \( w \) is in \( \text{Span}\{v_1, \ldots, v_i, v\} \).

To show b) If \( w \) is in \( \text{Span}\{v_1, \ldots, v_i, v\} \), then
\[
\begin{align*}
  w &= \text{lc}(v_1, \ldots, v_i) + cv \\
  &= \text{lc}(v_1, \ldots, v_i) + c(x - \mathcal{E}) \\
  &= \text{lc}(v_1, \ldots, v_i) + cx - \text{lc}(v_1, \ldots, v_i) \\
  &= \text{lc}(v_1, \ldots, v_i) + cx, \text{ so} \\
\end{align*}
\]
so \( w \) is in \( \text{Span}\{v_1, \ldots, v_i, x\} \).

The Gram-Schmidt Process (GSP) If you understand the preceding lemma, the idea behind the Gram-Schmidt Process is very easy. We want to convert an orthogonal basis \( \{x_1, \ldots, x_p\} \) for \( W \) into an orthogonal basis \( \{v_1, \ldots, v_p\} \). We build the orthogonal basis by replacing each vector \( x_i \) with a vector \( v_i \). We do this “from the ground up,” one \( x_i \) at a time, starting with \( x_1 \).

Step 1. To start, let \( v_1 = x_1 \), and call \( W_1 = \text{Span}\{x_1\} = \text{Span}\{v_1\} \).

Step 2: Now enlarge \( W_1 \): let \( W_2 = \text{Span}\{x_1, x_2\} = \text{Span}\{v_1, x_2\} \) (since \( v_1 = x_1 \)).
Decompose \( x_2 = \bar{x}_2 + v_2 \), where \( \bar{x}_2 = \text{proj}_{W_1}x_2 \) and \( v_2 \in W_1^\perp \).
But we don’t need “the whole vector \( x_2 \)” to get \( W_2 \). According to the Lemma, \( W_2 = \text{Span}\{x_1, x_2\} = \text{Span}\{v_1, x_2\} = \text{Span}\{v_1, v_2\} \), and, by the Lemma, \( \{v_1, v_2\} \) is an orthogonal set.

Of course, we already know a formula for \( v_2 \): since \( W_1 = \text{Span}\{v_1\} \),
\[
v_2 = x_2 - \text{proj}_{W_1}x_2 = x_2 - \frac{x_2 \cdot v_1}{v_1 \cdot v_1} v_1
\]

Step 3: Now enlarge \( W_2 \): let \( W_3 = \text{Span}\{x_1, x_2, x_3\} = \text{Span}\{v_1, v_2, x_3\} \).
Decompose \( x_3 = \bar{x}_3 + v_3 \), where \( \bar{x}_3 = \text{proj}_{W_2}x_3 \) and \( v_3 \in W_2^\perp \).
But we don’t need “the whole vector \( x_3 \)” to get \( W_3 \). According to the Lemma, \( W_3 = \text{Span}\{x_1, x_2, x_3\} = \text{Span}\{v_1, v_2, x_3\} = \text{Span}\{v_1, v_2, v_3\} \), and by the Lemma, \( \{v_1, v_2, v_3\} \) is an orthogonal set.

Of course, we already know a formula for \( v_3 \): since \( W_2 = \text{Span}\{v_1, v_2\} \),
\[
v_3 = x_3 - \text{proj}_{W_2}x_3 = x_3 - \frac{x_3 \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_3 \cdot v_2}{v_2 \cdot v_2} v_2
\]
Next: Keep repeating this process as long as you can. Suppose we have arrived at
\[ W_i = \text{Span}\{x_1, ..., x_i\} = \text{Span}\{v_1, ..., v_i\}, \]
where \( \{v_1, ..., v_i\} \) is an orthogonal set.

If \( i < p \) (that is, if there are still some \( x_i \)'s remaining), then use \( x_{i+1} \) to enlarge \( W_i \) again:
\[ W_{i+1} = \text{Span}\{x_1, ..., x_i, x_{i+1}\} = \text{Span}\{v_1, ..., v_i, x_{i+1}\}. \]

Once again, we don't need “the whole vector \( x_{i+1} \)”; as before, it's enough to use
\[ v_{i+1} = \text{the component of } x_{i+1} \text{ orthogonal to } W_i. \]

By the Lemma, we have \( \text{Span}\{x_1, ..., x_i, x_{i+1}\} = \text{Span}\{v_1, ..., v_i, x_{i+1}\} \)
and \( \{v_1, ..., v_i, x_{i+1}\} \) is an orthogonal set.

And, of course, we know a formula for \( v_{i+1} \): since \( W_i = \text{Span}\{v_1, v_2, ..., v_i\}, \)
\[
v_{i+1} = x_{i+1} - \text{proj}_W x_{i+1} = x_{i+1} - \frac{x_{i+1} \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_{i+1} \cdot v_2}{v_2 \cdot v_2} v_2 - ... - \frac{x_{i+1} \cdot v_i}{v_i \cdot v_i} v_i
\]

Finally, we run out of \( x_i \)'s and arrive at \( W_p = \text{Span}\{x_1, ..., x_p\} = \text{Span}\{v_1, ..., v_p\} \), and
\( \{v_1, ..., v_p\} \) is an orthogonal set that is a basis for the original subspace \( W = W_p. \)

For certain purposes it's useful to remember that at each step in the process, the set
\( \{x_1, ..., x_i\} \) has the same span as the set \( \{v_1, ..., v_i\} \)

You'll feel much better about the Gram-Schmidt Process if you understand the preceding material
about what's going on in the calculations. However, when you solve problems you might just use
a “programmed” recipe for the vectors in the orthogonal basis.

Suppose \( \{x_1, ..., x_p\} \) is a basis for the subspace \( W \) of \( \mathbb{R}^n \). Then \( \{v_1, v_2, ..., v_p\} \) is an
orthogonal basis for \( W \) where
\[
v_1 = x_1 \\
v_2 = x_2 - \text{proj}_{v_1} x_2 = x_2 - \frac{x_2 \cdot v_1}{v_1 \cdot v_1} v_1 \\
v_3 = x_3 - \text{proj}_{v_1} x_3 - \text{proj}_{v_2} x_3 = x_3 - \frac{x_3 \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_3 \cdot v_2}{v_2 \cdot v_2} v_2 \\
v_4 = x_4 - \text{proj}_{v_1} x_4 - \text{proj}_{v_2} x_4 - \text{proj}_{v_3} x_4 \\
= x_4 - \frac{x_4 \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_4 \cdot v_2}{v_2 \cdot v_2} v_2 - \frac{x_4 \cdot v_3}{v_3 \cdot v_3} v_3 \\
\vdots \\
v_p = x_p - \text{proj}_{v_1} x_p - \text{proj}_{v_2} x_p - ... - \text{proj}_{v_{p-1}} x_p \\
= x_p - \frac{x_p \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_p \cdot v_2}{v_2 \cdot v_2} v_2 - ... - \frac{x_p \cdot v_{p-1}}{v_{p-1} \cdot v_{p-1}} v_{p-1}
**Example** Use the Gram-Schmidt Process to find an orthogonal basis for

\[
W = \text{Span} \left\{ \begin{bmatrix} 0 \\ 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}
\]

and explain some of the details at each step.

\[
\begin{array}{c}
\uparrow \\
x_1 \\
\uparrow \\
x_2 \\
\uparrow \\
x_3
\end{array}
\]

You can check that \(x_1, x_2, x_3\) are linearly independent and therefore form a basis for \(W\). What would happen if we mindlessly applied the Gram-Schmidt Process to a linearly dependent set of vectors \(\{x_1, \ldots, x_p\}\)?

Let \(v_1 = x_1 = \begin{bmatrix} 0 \\ 2 \\ 0 \\ 1 \end{bmatrix}\), and \(W_1 = \text{Span} \{x_1\} = \text{Span} \{v_1\}\).

Let \(W_2 = \text{Span} \{x_1, x_2\} = \text{Span} \{v_1, x_2\}\).

Using the Lemma, we can replace \(x_2\) with \(v_2 = \text{the component of } x_2 \text{ orthogonal to } W_1 = x_2 - \text{proj}_{W_1} x_2\)

\[
\text{proj}_{W_1} x_2 = \frac{x_2 \cdot v_1}{v_1 \cdot v_1} v_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \\ 0 \\ 1 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 0 \\ 2 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{2}{5} \\ 0 \\ \frac{1}{5} \end{bmatrix}, \text{ so}
\]

\[
\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ \frac{2}{5} \\ 0 \\ \frac{1}{5} \end{bmatrix} = \begin{bmatrix} 1 \\ -\frac{2}{5} \\ 0 \\ \frac{4}{5} \end{bmatrix}
\]

Optional: rescale to avoid working with the fractions in the remaining calculations: multiply by 5 and instead use

\[
v_2 = 5 \begin{bmatrix} \frac{1}{5} \\ -\frac{2}{5} \\ 0 \\ \frac{4}{5} \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \\ 0 \\ 4 \end{bmatrix} \text{ (explain why rescaling doesn’t “mess anything up.”)}
\]

Then \(\text{Span} \{x_1, x_2\} = \text{Span} \{v_1, x_2\} = \text{Span} \{v_1, v_2\}\).

(Continued.)
Let $W_3 = \text{Span} \{x_1, x_2, x_3\} = \text{Span} \{v_1, v_2, x_3\}$

Using the Lemma, we can replace $x_3$ with $v_3$ = the component of $x_3$ orthogonal to $W_2$

$= x_3 - \text{proj}_{W_2} x_3$

$= x_3 - \text{proj}_{v_1} x_3 - \text{proj}_{v_2} x_3$

\[
\text{proj}_{W_2} x_3 = \frac{x_3 \cdot v_1}{v_1 \cdot v_1} v_1 + \frac{x_3 \cdot v_2}{v_2 \cdot v_2} v_2 = \begin{bmatrix}
1 & 0 \\
1 & 1 \\
2 & 2 \\
0 & 0 \\
1 & 1
\end{bmatrix} \begin{bmatrix}
0 \\
2 \\
0 \\
1
\end{bmatrix} + \begin{bmatrix}
1 & 0 & -2 \\
0 & 1 & 0 \\
1 & 1 & 4
\end{bmatrix} \begin{bmatrix}
5 \\
-2 \\
-2
\end{bmatrix} = \frac{1}{5} \begin{bmatrix} 0 \\ 2 \\ 0 \\ 1 \end{bmatrix} + \frac{9}{15} \begin{bmatrix} -2 \\ 0 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},
\]

so $v_3 = x_3 - \text{proj}_{W_2} x_3 = \begin{bmatrix}
1 \\
0 \\
1
\end{bmatrix} - \begin{bmatrix}
1 \\
0 \\
1
\end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

Then $W_3 = \text{Span} \{x_1, x_2, x_3\} = \text{Span} \{v_1, v_2, v_3\}$, and we have replaced all the $x_i$'s. $W_3$ is the subspace $W$ we were originally given, and an orthogonal basis for $W$ is

\[
\{v_1, v_2, v_3\} = \left\{ \begin{bmatrix}
0 \\
2 \\
1
\end{bmatrix}, \begin{bmatrix}
-5 \\
-2 \\
4
\end{bmatrix}, \begin{bmatrix}
0 \\
0 \\
1
\end{bmatrix} \right\}.
\]