COMMENTS ON THE GRAM-SCHMIDT PROCESS AND PROJECTIONS

MATH 223 SPRING 2001

The Gram-Schmidt process is an algorithm that takes a collection of linearly independent vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ in an inner product space V and converts them into an orthogonal collection of unit vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$. There are some things to note about the algorithm that lend some insight into what it is doing.

If the formulas in the book are written in terms of the intermediate collection of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ by replacing $\mathbf{e}_i$ with $\mathbf{v}_i / \|\mathbf{v}_i\|$, we have

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{u}_1, \\
\mathbf{v}_2 &= \mathbf{u}_2 - \frac{\langle \mathbf{u}_2, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1, \\
\mathbf{v}_3 &= \mathbf{u}_3 - \left(\frac{\langle \mathbf{u}_3, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 + \frac{\langle \mathbf{u}_3, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\mathbf{v}_2 \right), \\
&\cdots \\
\mathbf{v}_k &= \mathbf{u}_k - \left(\frac{\langle \mathbf{u}_k, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 + \cdots + \frac{\langle \mathbf{u}_k, \mathbf{v}_{k-1} \rangle}{\langle \mathbf{v}_{k-1}, \mathbf{v}_{k-1} \rangle}\mathbf{v}_{k-1} \right) \qquad \text{for } k \leq n.
\end{aligned}$$

These formulas are not particularly good from a computational standpoint, since the inner products $\langle \mathbf{v}_1, \mathbf{v}_1 \rangle$, $\langle \mathbf{v}_2, \mathbf{v}_2 \rangle$, etc., would have to be computed repeatedly, however there are some things that are perhaps more evident from them. We see that $\mathbf{v}_2$ is a linear combination of $\mathbf{u}_1$ ($= \mathbf{v}_1$) and $\mathbf{u}_2$. It follows that $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\} = \operatorname{span}\{\mathbf{u}_1, \mathbf{u}_2\}$. Similarly, $\mathbf{v}_3$ is a linear combination of $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{u}_3$, and so of $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{u}_3$, since $\mathbf{v}_1$ and $\mathbf{v}_2$ are linear combinations of $\mathbf{u}_1$ and $\mathbf{u}_2$. It follows that $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} = \operatorname{span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$. Continuing (there is an induction argument), we get that $\operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \operatorname{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ for each $k \leq n$. Since $\mathbf{e}_i = \mathbf{v}_i / \|\mathbf{v}_i\|$, we have $\operatorname{span}\{\mathbf{e}_1, \ldots, \mathbf{e}_k\} = \operatorname{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$. If we let

$$\begin{aligned}
V_1 &= \operatorname{span}\{\mathbf{e}_1\} = \operatorname{span}\{\mathbf{u}_1\}, \\
V_2 &= \operatorname{span}\{\mathbf{e}_1, \mathbf{e}_2\} = \operatorname{span}\{\mathbf{u}_1, \mathbf{u}_2\}, \\
&\cdots \\
V_k &= \operatorname{span}\{\mathbf{e}_1, \ldots, \mathbf{e}_k\} = \operatorname{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\} \quad \text{for } k \leq n,
\end{aligned}$$

we see that the process preserves the nested sequence of subspaces

$$V_1 \subset V_2 \subset \cdots \subset V_k \subset \cdots \subset V_n.$$

Another observation from the formulas above (and which is made in the book) is that the process involves projections. In particular, for any $\mathbf{w} \in V$ we have $\operatorname{Proj}_{\mathbf{v}_i}\mathbf{w} = \frac{\langle \mathbf{w}, \mathbf{v}_i \rangle}{\langle \mathbf{v}_i, \mathbf{v}_i \rangle}\mathbf{v}_i = \langle \mathbf{w}, \mathbf{e}_i \rangle \mathbf{e}_i = \operatorname{Proj}_{\mathbf{e}_i}\mathbf{w}$. Thus the Gram-Schmidt formulas can be written as

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{u}_1, \\
\mathbf{v}_2 &= \mathbf{u}_2 - \operatorname{Proj}_{\mathbf{v}_1}\mathbf{u}_2, \\
\mathbf{v}_3 &= \mathbf{u}_3 - \left(\operatorname{Proj}_{\mathbf{v}_1}\mathbf{u}_3 + \operatorname{Proj}_{\mathbf{v}_2}\mathbf{u}_3\right), \\
&\ \ \cdots \\
\mathbf{v}_k &= \mathbf{u}_k - \left(\operatorname{Proj}_{\mathbf{v}_1}\mathbf{u}_k + \cdots + \operatorname{Proj}_{\mathbf{v}_{k-1}}\mathbf{u}_k\right) \qquad \text{for } k \le n.
\end{aligned}$$

In Section 4.5 the author (Messer) defines projection into a subspace. With that concept in mind we see that $\operatorname{Proj}_{\mathbf{v}_1}\mathbf{w} + \cdots + \operatorname{Proj}_{\mathbf{v}_{k-1}}\mathbf{w}$ is projection of $\mathbf{w}$ into the subspace V_k, written as $\operatorname{Proj}_{V_{k-1}}\mathbf{w}$, and so the formulas can be written as

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{u}_1, \\
\mathbf{v}_2 &= \mathbf{u}_2 - \operatorname{Proj}_{V_1}\mathbf{u}_2, \\
\mathbf{v}_3 &= \mathbf{u}_3 - \operatorname{Proj}_{V_2}\mathbf{u}_3, \\
&\ \ \cdots \\
\mathbf{v}_k &= \mathbf{u}_k - \operatorname{Proj}_{V_{k-1}}\mathbf{u}_k \qquad \text{for } k \le n.
\end{aligned}$$

I'll make one final observation. If S is a subspace of V and $\mathbf{w} \in V$, then, as is pointed out in Section 4.5, the vector $\mathbf{w} - \operatorname{Proj}_S\mathbf{w}$ is perpendicular to S (more precisely, it is perpendicular to every vector in S). If we let $S^\perp$ be the collection of all vectors perpendicular to S, then $\mathbf{w} - \operatorname{Proj}_S\mathbf{w}$ is in $S^\perp$. It should come as no surprise that $S^\perp$ is a subspace of V (you should be able to prove this!), called the *orthogonal complement* of S. Now the process of projecting $\mathbf{w}$ into S was one of writing $\mathbf{w}$ as the sum of two vectors, one in S and one perpendicular to S, that is, in $S^\perp$. The one that's in S is $\operatorname{Proj}_S\mathbf{w}$, and so the one that's in $S^\perp$ is $\operatorname{Proj}_{S^\perp}\mathbf{w}$, that is, $\mathbf{w} = \operatorname{Proj}_S\mathbf{w} + \operatorname{Proj}_{S^\perp}\mathbf{w}$, or $\mathbf{w} - \operatorname{Proj}_S\mathbf{w} = \operatorname{Proj}_{S^\perp}\mathbf{w}$. Thus the Gram-Schmidt formulas can be written as

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{u}_1, \\
\mathbf{v}_2 &= \operatorname{Proj}_{V_1^\perp}\mathbf{u}_2, \\
\mathbf{v}_3 &= \operatorname{Proj}_{V_2^\perp}\mathbf{u}_3, \\
&\ \ \cdots \\
\mathbf{v}_k &= \operatorname{Proj}_{V_{k-1}^\perp}\mathbf{u}_k \qquad \text{for } k \le n.
\end{aligned}$$

Thus at each step of the algorithm we consider the subspace $V_{k-1} = \operatorname{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_{k-1}\}$. The vector u_k is not in V_{k-1}. It might not be perpendicular to V_{k-1}, and so we replace it with its component that is perpendicular to V_{k-1}, that is, its component in $V_{k-1}^\perp$, which is $\operatorname{Proj}_{V_{k-1}^\perp}\mathbf{u}_k$. Note that the subspaces $V_1^\perp, \ldots, V_n^\perp$ are also nested, but in reverse order from $V_1, \ldots, V_n$:

$$V_1^\perp \supset V_2^\perp \supset \cdots \supset V_n^\perp.$$