The Gram-Schmidt Algorithm.

Suppose that V is an inner product space and that H is a finite dimensional non-trivial subspace. Note that if H has an orthogonal basis $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$, then we may compute the coordinates of any $\mathbf{v} \in H$ relative to H as follows: writing $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$, we take the inner product of both sides with $\mathbf{v}_j$ and discover that $\langle \mathbf{v}, \mathbf{v}_j \rangle = c_j \langle \mathbf{v}_j, \mathbf{v}_j \rangle$. Hence the jth coordinate of $[\mathbf{v}_j]_{\mathcal{B}}$ is $c_j = \langle \mathbf{v}, \mathbf{v}_j \rangle / \|\mathbf{v}_j\|^2$. Thus

$$\mathbf{v} = \mathrm{Proj}_H(\mathbf{v}) := \sum_{j=1}^{k} \frac{\langle \mathbf{v}, \mathbf{v}_j \rangle}{\|\mathbf{v}_j\|^2} \mathbf{v}_j$$

for all $\mathbf{v} \in H$. Now if $\mathbf{v} \in V$ is a vector outside H, the righthand expression still defines a vector $\mathrm{Proj}_H(\mathbf{v}) \in H$, but this cannot possibly equal $\mathbf{v}$. Instead $\mathrm{Proj}_H : V \to H$ is a linear (why? Think about it.) transformation known as the 'orthogonal projection' of V onto H. As the word 'the' implies, the transformation turns out the same no matter which orthogonal basis we use to define it.

Proposition 0.1. *Let* $\mathrm{Proj}_H : V \to H$ *be as above. Given* $\mathbf{v} \in V$*, let* $\mathbf{v}_H = \mathrm{Proj}_H(\mathbf{v})$ *and* $\mathbf{v}_\perp = \mathbf{v} - \mathbf{v}_H$*. Then* $\mathbf{v}_H$ *and* $\mathbf{v}_\perp$ *are the unique vectors satisfying* $\mathbf{v}_H \in H$*,* $\mathbf{v}_\perp \in H^\perp$ *and* $\mathbf{v} = \mathbf{v}_H + \mathbf{v}_\perp$*. Moreover,* $\mathbf{v}_H$ *is the unique vector in* H *minimizing the distance* $\|\mathbf{v} - \mathbf{v}_H\|$*.*

Proof. By linearity and the definition of Proj_H, we have

$$\langle \mathbf{v}_\perp, \mathbf{v}_j \rangle = \langle \mathbf{v}, \mathbf{v}_j \rangle - \sum_{\ell=1}^{K} \frac{\langle \mathbf{v}, \mathbf{v}_\ell \rangle \langle \mathbf{v}_\ell, \mathbf{v}_j \rangle}{\|\mathbf{v}_\ell\|^2} = \langle \mathbf{v}, \mathbf{v}_j \rangle - \frac{\langle \mathbf{v}, \mathbf{v}_j \rangle \langle \mathbf{v}_j, \mathbf{v}_j \rangle}{\|\mathbf{v}_j\|^2} = 0.$$

So $\mathbf{v}_\perp$ is orthogonal to each vector $\mathbf{v}_j \in \mathcal{B}$, and it follows that $\mathbf{v}_\perp \in H^\perp$. Clearly $\mathbf{v} = \mathbf{v}_H + \mathbf{v}_\perp$. If $\mathbf{w}_H \in H$ and $\tilde{\mathbf{w}}_\perp \in H^\perp$ also satisfy $\mathbf{v} = \mathbf{w}_H + \mathbf{v}_\perp$, then $\mathbf{w}_H + \mathbf{w}_\perp = \mathbf{v}_H + \mathbf{v}_\perp$ implies that $\mathbf{w}_H - \mathbf{v}_H = \mathbf{v}_\perp - \mathbf{w}_\perp \in H \cap H^\perp$. On the other hand, non-degeneracy of the inner product says that $\mathbf{0}$ is the only vector orthogonal to itself. I conclude that $\mathbf{w}_H - \mathbf{v}_H = \mathbf{v}_\perp - \mathbf{w}_\perp = \mathbf{0}$. That is, $\mathbf{v}_H$ and $\mathbf{v}_\perp$ give the unique decomposition of $\mathbf{v}$ into vectors in and orthogonal to H.

For the final assertion, I let $\mathbf{w} \in H$ be any vector and estimate

$$\|\mathbf{v} - \mathbf{w}\|^2 = \|\mathbf{v}_\perp - (\mathbf{w} - \mathbf{v}_H)\|^2 = \langle \mathbf{v}_\perp - (\mathbf{w} - \mathbf{v}_H), \mathbf{v}_\perp - (\mathbf{w} - \mathbf{v}_H) \rangle .$$

But $\mathbf{w} - \mathbf{v}_H \in H$ is orthogonal to $\mathbf{v}_\perp$, so when I expand the last expression, two of the four terms vanish leaving me with.

$$\|\mathbf{v} - \mathbf{w}\|^2 = \langle \mathbf{v}_\perp, \mathbf{v}_\perp \rangle + \langle \mathbf{w} - \mathbf{v}_H, \mathbf{w} - \mathbf{v}_H \rangle = \|\mathbf{v}_\perp\|^2 + \|\mathbf{w} - \mathbf{v}_H\|^2 \geq \|\mathbf{v}_\perp\|^2 ,$$

with equality only if $\mathbf{w} = \mathbf{v}_H$. □

Of course, the definition of orthogonal projection required that H have at least one orthogonal basis. Our final result guarantees us that this will always be the case when H is finite dimensional.

Theorem 0.2. *Let* H *be a non-trivial finite dimensional subspace of an inner product space* V*. Then* H *has an orthogonal basis.*

Note that the proof doesn't just show that an orthogonal basis exists; it actually gives a recursive method for constructing it. This is known as the 'Gram-Schmidt algorithm.'

Proof. I work by induction on $\dim H$. In the case $\dim H = 1$, any basis $\{\mathbf{v}_1\}$ for H consists of a single non-zero vector and is therefore also an orthogonal set.

Supposing the theorem is true when $\dim H \leq k$, I consider the case $\dim H = k + 1$. Let $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_{k+1}\}$ be a basis for H and let $W = \operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ be the smaller subspace generated by the first k basis vectors. Then by my inductive hypothesis, there is an orthogonal basis $\{\mathbf{w}_1, \dots, \mathbf{w}_k\}$ for W. By the Proposition, I have $\mathbf{v}_{k+1} = \operatorname{Proj}_W(\mathbf{v}_{k+1}) + \mathbf{w}_{k+1}$ where $\mathbf{w}_{k+1} \in W^\perp \cap H$.

I claim $\mathbf{w}_{k+1} \neq \mathbf{0}$. If not, then $\mathbf{v}_{k+1} = \operatorname{Proj}_W(\mathbf{v}_{k+1}) \in W$ must be a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_k$, which contradicts independence of $\mathcal{B}$. On the other hand, since $\mathbf{w}_1, \dots, \mathbf{w}_k \in W$ and $\mathbf{w}_{k+1} \in W^\perp$, I have $\langle \mathbf{w}_{k+1}, \mathbf{w}_j \rangle = 0$ for all $j \leq k$. Hence $\mathcal{C} := \{\mathbf{w}_1, \dots, \mathbf{w}_{k+1}\}$ is an orthogonal set of non-zero vectors and therefore independent. Since all $\mathbf{w}_j \in H$ and $\dim H = k + 1$, it follows that $\mathcal{C}$ is an orthogonal basis for H. This completes the inductive step and the proof. $\square$