Recall that any linearly independent set of $\mathbb{R}^m$ has at most $m$ elements. For any subspace $X$ of $\mathbb{R}^m$ (including $\mathbb{R}^m$ itself), we can find a basis for that subspace; call it $B = \{b_1, \ldots, b_n\}$. We can find an orthonormal basis $Q$ for $X$ via the Gram-Schmidt algorithm applied to $B$; since the columns of $B$ span $X$, the columns of $Q$ span $X$. On the other hand, the columns of a matrix with orthonormal columns are linearly independent: suppose $Qa = 0$, then $a = Ia = Q^TQa = 0$.

1 Orthogonal complement

Now let $W \subset \mathbb{R}^m$. Define $W^\perp$ by

$$W^\perp = \{x \in \mathbb{R}^n, \langle x, w \rangle = 0 \forall w \in W\}$$

Note that even if $W$ is not a subspace of $\mathbb{R}^n$, $W^\perp$ is: if $x, y \in W^\perp$, $a \in \mathbb{R}$, and $w \in W$,

$$\langle x + y, w \rangle = \langle x, w \rangle + \langle y, w \rangle = 0,$$

and

$$\langle ax, w \rangle = a \langle x, w \rangle = 0,$$

So $W^\perp$ is closed under addition and scalar multiplication.

If $W$ itself is a subspace, say of dimension $n$, we can divide up $\mathbb{R}^n$ into $W$ and $W^\perp$ in a very natural way: every $v \in \mathbb{R}^n$ can be written uniquely as

$$v = v_W + v_{W^\perp},$$

where $v_W \in W$, and $v_{W^\perp} \in W^\perp$. To see this, let $Q$ be an orthonormal basis for $W$, and let $v \in \mathbb{R}^n$. Set

$$v_W = QQ^Tv,$$

and

$$v_{W^\perp} = v - v_W.$$

It is clear that $v_W + v_{W^\perp}$; it is also clear that $v_W \in W$, because each column of $Q$ is in $W$, and $v_W$ is a linear combination of columns of $Q$, and $W$ is closed under linear combinations. On the other hand, if $w \in W$, since $Q$ is a basis for $W$, there exists $a$ such that $w = Qa$. Then since $Q^TQ = I_n$,

$$w^Tv_{W^\perp} = (Qa)^T(v - QQ^Tv) = a^TQ^Tv - a^TQ^Tv = 0,$$

so $v_{W^\perp} \in W^\perp$, and we have found a decomposition of $v$ into an element from $W$ and from $W^\perp$. These elements are unique: if $v = a + b$, where $a \in W$ and $b \in W^\perp$,

$$v_W + v_{W^\perp} = v = a + b$$

so

$$v_W - a = b - v_{W^\perp},$$

and since the left hand side is in $W$, and the right hand side in $W^\perp$, both $v_W - a$ and $b - v_{W^\perp}$ are in $W \cap W^\perp$. But any vector $u \in W \cap W^\perp$ satisfies $\langle u, u \rangle = 0$, by the definition of $W^\perp$, and so $u = 0$. Thus $a = v_W$ and $b = v_{W^\perp}$.

We can now show that if $W$ is a subspace, $(W^\perp)^\perp = W$. If $w \in W$, and $x \in W^\perp$, by definition, $\langle w, x \rangle = 0$, so $w \in W^\perp$. On the other hand, if $v \notin W$, $v_{W^\perp} \neq 0$, and so

$$\langle v, v_{W^\perp} \rangle = \langle v_W + v_{W^\perp}, v_{W^\perp} \rangle = \langle v_{W^\perp}, v_{W^\perp} \rangle = ||v_{W^\perp}||^2 > 0,$$

and so $v \notin W^\perp$. 

1
2 Orthogonal projection

We can thus define a mapping $P_W : \mathbb{R}^m \mapsto \mathbb{R}^m$ via $v \mapsto v_W$ (the above arguments show that this map is defined and unambiguous for every $v \in \mathbb{R}^m$, and is linear). This is called the orthogonal projection onto $W$.

- If $w \in W$, $P_W w = w$, and if $x \in W^\perp$, $P_W x = 0$.
- for any orthonormal basis $Q$ of $W$, by the uniqueness of the decomposition (1.1),
  \[ P_W v = QQ^T v. \]
- and
  \[ P_W v = (I_m - QQ^T) v. \]
- If $Q$ is any matrix with orthonormal columns, $QQ^T$ is the orthonormal projection onto the column space of $Q$.

In class, many times we have mentioned that if $m > n$, and $Q$ is an $m \times n$ orthonormal matrix, $QQ^T$ is not the identity; now we can see it is the identity on the column space of $Q$, and it is the zero mapping on the orthogonal complement of the column space of $Q$...

The fact that the choice of basis in the definition of the projection is irrelevant suggest that there is a basis-free way of writing the orthogonal projection onto $W$, and indeed,

\[ P_W v = \text{arg} \min_{w \in W} ||v - w||^2. \]

To see this, let $w \in W$. Then

\[ ||v - w||^2 = ||v_W + v_W^\perp - w||^2 = ||v_W - w||^2 + ||v_W^\perp||^2. \]

But

\[ ||v - v_W||^2 = ||v_W + v_W^\perp - v_W||^2 = ||v_W^\perp||^2, \]

so

\[ ||v - w||^2 = ||v_W - w||^2 + ||v - v_W||^2 \geq ||v - v_W||^2, \]

with equality if and only if $w = v_W$.

3 Some extra notes and perspectives

This section will not tested on the midterm, but may clarify some things but also may unclarify things, depending on how you like to think: feel free to ignore it.

I wrote $W \subset \mathbb{R}^m$ instead of $W \subset V$, where $V$ is a finite dimensional vector space, to emphasize that we wrote everything in the standard basis. This is unnecessary, and you may consider it inelegant or even confusing- but if this actually bothers you, you can replace each expression of the form $QQ^T v$ by an expression of the form

\[ \sum_{j=1}^n \langle Q_j, v \rangle Q_j, \]

where $Q_j$ is the $j$th element in the basis $Q$ of $W$. The point is that here, $Q$ is not a matrix, but rather a list of vectors, and is not necessarily expressed in the standard basis of $\mathbb{R}^m$.

Moreover, if you want to avoid bases altogether, you can, by making the definition of the projection of $v$ onto $W$ to be the point closest to $v$ in $W$, that is, define

\[ v_W = \text{arg} \min_{w \in W} ||v - w||^2. \]
Using a bit of calculus that is beyond the scope of this course (but is not too hard!), you can check that a minimizer exists. Then if $w \in W$, consider the degree 2 polynomial in $t$

$$||(v - v_W) + tw||^2 = ||v - v_W||^2 - 2t \langle v - v_W, w \rangle + t^2||w||^2.$$ 

Since this polynomial takes a minimum at $t = 0$ (because of the definition of $v_W$), the term $\langle v - v_W, w \rangle = 0$ (draw a picture or take a derivative in $t$ to see this). So $v - v_W$ is orthogonal to every element in $W$. Moreover, if $w \in W$,

$$||v - w||^2 = ||v - v_W + v_W - w||^2 = ||v - v_W||^2 + ||v_W - w||^2 \geq ||v - v_W||^2,$$

with equality only if $w = v_W$; that is, the projection is unique. So as soon as we know that for each $v \in V$ there is at least one minimizer for $\arg \min_{w \in W} ||v - w||^2$, we know that there is exactly one, and so the projection is well defined this way. You can further check that the mapping is linear, that it leads to the definition of the orthogonal complement, etc.

Again, I will not test on this subsection, but wanted to include this in case you were bothered by the fact that the basis appeared in the construction of the projection even though the projection doesn’t care which basis you use.