The formula for the orthogonal projection

Let V be a subspace of $\mathbb{R}^n$. To find the matrix of the orthogonal projection onto V, the way we first discussed, takes three steps:

(1) Find a basis $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m$ for V.
(2) Turn the basis $\vec{v}_i$ into an orthonormal basis $\vec{u}_i$, using the Gram-Schmidt algorithm.
(3) Your answer is $P = \sum \vec{u}_i \vec{u}_i^T$. Note that this is an $n \times n$ matrix, we are multiplying a column vector by a row vector instead of the other way around.

It is often better to combine steps (2) and (3). (Note that you still need to find a basis!) Here is the result: Let A be the matrix with columns $\vec{v}_i$. Then

$$\boxed{P = A(A^TA)^{-1}A^T}$$

Your textbook states this formula without proof in Section 5.4, so I thought I'd write up the proof. **I urge you to also understand the other ways of dealing with orthogonal projection that our book discusses, and not simply memorize the formula.**

Example

Let V be the span of the vectors $(1\ 2\ 3\ 4)^T$ and $(5\ 6\ 7\ 8)^T$. These two vectors are linearly independent (since they are not proportional), so

$$A = \begin{pmatrix} 1 & 5 \\ 2 & 6 \\ 3 & 7 \\ 4 & 8 \end{pmatrix}.$$

Then

$$A^TA = \begin{pmatrix} 30 & 70 \\ 70 & 174 \end{pmatrix} \quad (A^TA)^{-1} = \begin{pmatrix} \frac{87}{160} & \frac{-7}{32} \\ \frac{-7}{32} & \frac{3}{32} \end{pmatrix}.$$

Note that A^T and A are not square, but the product A^TA is, so $(A^TA)^{-1}$ makes sense.

Then we have

$$A(A^TA)^{-1}A^T = \begin{pmatrix} 1 & 5 \\ 2 & 6 \\ 3 & 7 \\ 4 & 8 \end{pmatrix} \begin{pmatrix} \frac{87}{160} & \frac{-7}{32} \\ \frac{-7}{32} & \frac{3}{32} \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{pmatrix} = \begin{pmatrix} \frac{7}{10} & \frac{2}{5} & \frac{1}{10} & -\frac{1}{5} \\ \frac{2}{5} & \frac{3}{10} & \frac{1}{5} & \frac{1}{10} \\ \frac{1}{10} & \frac{1}{5} & \frac{3}{10} & \frac{2}{5} \\ -\frac{1}{5} & \frac{1}{10} & \frac{2}{5} & \frac{7}{10} \end{pmatrix}$$

You can have fun checking that this is, indeed, the matrix of orthogonal projection onto V.

Why is A^TA invertible?

This formula only makes sense if A^TA is invertible. So let's prove that it is. Suppose, for the sake of contradiction, that $\vec{c} = (c_1, c_2, \cdots, c_m)$ is a nonzero vector in the kernel of A^TA. Then $A^TA\vec{c} = 0$ and so

$$0 = \vec{c}^T A^T A \vec{c} = (A\vec{c})^T (A\vec{c}) = |A\vec{c}|^2.$$

Since the only vector with length 0 is $\vec{0}$, this shows that $A\vec{c} = 0$.

But $A\vec{c} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_m\vec{v}_m$ and we assumed that $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m$ was a basis for V, so there are no linear relations between the $\vec{v}_i$. So we can't have $c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_m\vec{v}_m = \vec{0}$. This is our contradiction and we deduce that A^TA didn't have a kernel after all.

Direct proof

Let's first check that $P\vec{w}$ is the right thing in two particular cases.

Case 1: $\vec{w}$ is in V. In this case, $\vec{w} = A\vec{c}$ for some $\vec{c}$. Then

$$P\vec{w} = A(A^TA)^{-1}A^T(A\vec{c}) = A(A^TA)^{-1}(A^TA)\vec{c} = A\vec{c} = \vec{w}$$

which is what we want.

Case 2: $\vec{w}$ is in $V^\perp$. Since $V = \text{Image}(A)$, we have $V^\perp = \text{Image}(A)^\perp = \text{Ker}(A^T)$ and we see that $A^T\vec{w} = 0$. Then

$$P\vec{w} = A(A^TA)^{-1}A^T\vec{w} = A(A^TA)^{-1}\vec{0} = \vec{0}$$

which is, again, what we want.

For a general $\vec{w}$, write $\vec{w} = \vec{w}_1 + \vec{w}_2$ where $\vec{w}_1$ is in V and $\vec{w}_2$ is in $V^\perp$. Then

$$P\vec{w} = P\vec{w}_1 + P\vec{w}_2 = \vec{w}_1 + \vec{0} = \vec{w}_1$$

where we used Case 1 to compute $P\vec{w}_1$ and Case 2 to compute $P\vec{w}_2$. Taking $\vec{w}$ to $\vec{w}_1$ is, indeed, exactly what orthogonal projection is suppose to do.

Proof using orthogonal bases

Our vectors $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_m$ are a basis for V, but not an orthonormal basis. However, V does have an orthonormal basis. Let $\vec{u}_1, \vec{u}_2, \dots, \vec{u}_m$ be this basis. Let's write Q for the matrix whose columns are the $\vec{u}_i$. The condition that $\vec{u}_i$ are orthonormal is the same as $Q^TQ = \text{Id}_m$. Let $\vec{v}_j = \sum_i R_{ij}\vec{u}_i$. We can rewrite this using matrices, $A = QR$.

I claim that R is invertible. Notice that R is square (it's $m \times m$). If $R\vec{x} = 0$, then $A\vec{x} = QR\vec{x} = 0$. But the columns of A are linearly independent, so A is injective, a contradiction. Since $\text{Rank}(R) = \text{Rank}(R^T)$, this also shows that R^T is invertible.

Now we plug in:

$$\begin{aligned}
P &= A(A^TA)^{-1}A^T \\
&= (QR)((QR)^TQR)^{-1}(QR)^T \\
&= QR(R^TQ^TQR)^{-1}R^TQ^T \\
&= QR(R^TR)^{-1}R^TQ^T && \text{because } Q^TQ = \text{Id} \\
&= QRR^{-1}(R^T)^{-1}R^TQ^T && \text{we already argued that } R \text{ and } R^T \text{ are invertible} \\
&= QQ^T
\end{aligned}$$

As your textbook explains (Theorem 5.3.10), when the columns of Q are an orthonormal basis of V, then QQ^T is the matrix of orthogonal projection onto V.

Note that we needed to argue that R and R^T were invertible before using the formula $(R^TR)^{-1} = R^{-1}(R^T)^{-1}$. By contrast, A and A^T are not invertible (they're not even square) so it doesn't make sense to write $(A^TA)^{-1} = A^{-1}(A^T)^{-1}$.