Week 3 Solutions

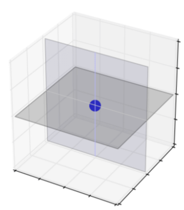
Gov January Linear Algebra Review

2021-01-19

1. (Strang 4.1.13) Put bases for the subspaces $\boldsymbol{V}$ and $\boldsymbol{W}$ into the columns of the matrices V and W. Explain why the test for orthogonal subspaces can be written $V^T W =$ zero matrix. This matches $\boldsymbol{v}^T \boldsymbol{w} = 0$ for orthogonal vectors.

Remember from page 194, that **subspaces** – some subset of vectors belonging to some greater **space**, e.g. $\mathbb{R}^2$ that can be **spanned** by some set of independent vectors (a **basis**) – $\boldsymbol{V}$ and $\boldsymbol{W}$ are **orthogonal** when $v^T w = 0$ for every v in $\boldsymbol{V}$ and every w in $\boldsymbol{W}$. Two vectors are orthogonal if they're perpendicular or *completely* unaligned directionally. If two subspaces are orthogonal, all pairs of vectors across the two subspaces are orthogonal. Since every vector in a subspace is a linear combination of basis vectors, this means that each subspace's basis vector must be orthogonal to the other subspace's basis vectors.

An example are the horizontal and vertical planes in $\mathbb{R}^3$ that go through $(0, 0, 0)$:



$V^T W =$ zero matrix makes each column of V orthogonal to each column of W. This means: each basis vector for $\boldsymbol{V}$ is orthogonal to each basis vector for $\boldsymbol{W}$. Then every v in $\boldsymbol{V}$ (combinations of the basis vectors) is orthogonal to every w in $\boldsymbol{W}$.

2. (Strang 4.1.25) Find $A^T A$ if the columns of A are unit vectors, all mutually perpendicular.

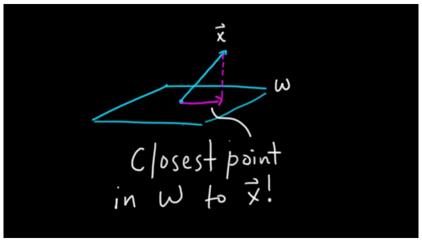
If the columns of A are unit vectors, all mutually perpendicular, then $A^T A = I$. Simple but important! We usually call such a matrix Q.

3. (Strang 4.2.13) Suppose A is the 4 by 4 identity matrix with its last column removed. A is 4 by 3. Project $\boldsymbol{b} = (1, 2, 3, 4)$ onto the column space of A. What shape is the projection matrix P and what is P?

Recall in our first review session the intuitive geometric idea of projecting vector a onto b: finding a point on vector b that is the "closest" to vector b.

The **projection** of a vector b onto a subspace $\boldsymbol{A}$ is the closest vector p in $\boldsymbol{A}$. Incidentally, $b -p $ ends up being orthogonal to $\boldsymbol{A}$.

An example of some vector $x \in \mathbb{R}^3$ being projected down to the subspace in $\mathbb{R}^3$ made up by a horizontal plane is shown here:



The formula for projection of a vector $\boldsymbol{b}$ onto a column space of A (given on page 206 and some motivation on the following pages as well as this lecture video) is given by $p = A(A^TA)^{-1}A^Tb$. The first part of the right-hand side, $A(A^TA)^{-1}A^T$, is called the **projection matrix**.

Let's get each of the components of the projection matrix:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$A^TA = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I$$

which you can get either by doing it out by hand, or in R:

```
A <- rbind(c(1, 0, 0),
           c(0, 1, 0),
           c(0, 0, 1),
           c(0, 0, 0))
AtA <- t(A) %*% A
print(AtA)
```

```
##      [,1] [,2] [,3]
## [1,]    1    0    0
## [2,]    0    1    0
## [3,]    0    0    1
```

Note that the inverse of an identity matrix is itself, so $(A^TA)^{-1} = I$.

Finally, we can get P which involves computing $A \cdot A^T$ (which is different from $A^T \cdot A$):

$$P = A \cdot (A^T A)^{-1} \cdot A^T = A \cdot A^T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

which we can do by hand or in R:

```
AAt <- A %*% t(A)
print(AAt)
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    1    0    0    0
## [2,]    0    1    0    0
## [3,]    0    0    1    0
## [4,]    0    0    0    0
```

Now to get the projection p we must do $P \cdot b$:

$$P \cdot b = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 0 \end{bmatrix}$$

4. (Strang 4.2.21) Multiply the matrix $P = A(A^TA)^{-1}A^T$ by itself. Cancel to prove that $P^2 = P$. Explain by $P(P\boldsymbol{b})$ always equals $P\boldsymbol{b}$. The vector $P\boldsymbol{b}$ is in the column space of A so its projection onto that column space is __________.

$$\begin{aligned}
P^2 = \left(A(A^TA)^{-1}A^T\right)^2 &= A(A^TA)^{-1}A^T \cdot A(A^TA)^{-1}A^T && (1) \\
&= A(A^TA)^{-1}(A^TA)(A^TA)^{-1}A^T && \text{(group together middle term)} \quad (2) \\
&= A\cancel{(A^TA)^{-1}(A^TA)}(A^TA)^{-1}A^T && \text{(cancel out inverses)} \quad (3) \\
&= A(A^TA)^{-1}A^T && \text{(cancel out inverses)} \quad (4) \\
&= P && (5)
\end{aligned}$$

The implication is that we can't *re-project* a vector once it's *already projected*. Pb is in the column space (where P projects), and the projection of the projection $P(Pb)$ is also Pb.

5. (Strang 4.2.22) Prove that $P = A(A^TA)^{-1}A^T$ is symmetric by computing P^T. Remember that the inverse of a symmetric matrix is symmetric.

A square matrix is **symmetric** if the upper triangular matrix entries are equal to the lower triangular matrix entries, e.g.

$$\begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 2 & 3 & 0 \end{bmatrix}$$

Or in other words, $A^T = A$. Let's show this for P:

$$P^T = \left(\underline{A} \cdot \underline{(A^TA)^{-1}A^T}\right)^T \tag{6}$$

$$= \left(\underline{(A^TA)^{-1}}\underline{A^T}\right)^T \cdot \left(A\right)^T \quad \text{(rule that } (XY)^T = Y^TX^T\text{)} \tag{7}$$

$$= \left(A\right) \cdot \left((A^TA)^{-1}\right)^T \cdot \left(A\right)^T \quad \text{(re-apply this rule, but on the left-hand side)} \tag{8}$$

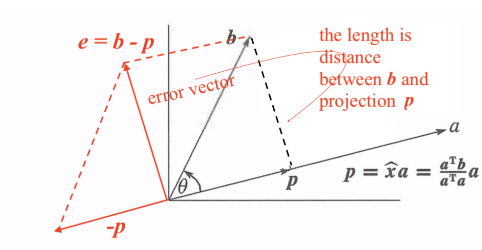
$$\tag{9}$$

We would get exactly what we need if we could just show that $\left((A^TA)^{-1}\right)^T = (A^TA)^{-1}$. Is this true?

Note that A^TA is always going to produce a symmetric matrix (you can draw out an example in low dimensions). Therefore, by the given prompt, $(A^TA)^{-1}$ is *also* symmetric. By definition of symmetry, then, $(A^TA)^{-1} = A^TA$, so indeed this is true and we're done. Huzzah!

6. (Strang 4.2.23) Is the error vector $\boldsymbol{e}$ orthogonal to $\boldsymbol{b}$ or $\boldsymbol{p}$ or $\boldsymbol{e}$ or $\widehat{\boldsymbol{x}}$? Show that $||\boldsymbol{e}||^2$ equals $\boldsymbol{e}^T\boldsymbol{b}$ which equals $\boldsymbol{b}^T\boldsymbol{b} - \boldsymbol{p}^T\boldsymbol{b}$.

Let's back up and get a visual of how these quantities relate to each other (a slightly less technically correct version is shown on p. 208 of Strang):



Here, $\widehat{\boldsymbol{x}}$ is the single-dimensional equivalent of the projection matrix – just one number that tells us where b should project down to along the line a.

We can see clearly here that the error vector $\boldsymbol{e}$ – obtained by vector addition of p and b – is orthogonal to $\boldsymbol{p}$, but not $\boldsymbol{b}$ or $\boldsymbol{e}$ (itself). $\widehat{\boldsymbol{x}}$ is simply a number and a vector cannot be orthogonal to a number.

We can walk through the simplifications needed to show the first equality using the definition of length, the distributive property, the definition of e, and the above orthogonality we've discovered:

$$||\boldsymbol{e}||^2 = \boldsymbol{e}^T\boldsymbol{e} = \boldsymbol{e}^T \cdot (\boldsymbol{b} - \boldsymbol{p}) = \boldsymbol{e}^T \cdot \boldsymbol{b} - \cancel{\boldsymbol{e}^T \boldsymbol{p}} = \boldsymbol{e}^T\boldsymbol{b}$$

We can again use the definition of e and the fact that transposition can be distributed to get the rest of the way:

$$\boldsymbol{e}^T\boldsymbol{b} = (\boldsymbol{b} - \boldsymbol{p})^T\boldsymbol{b} = \boldsymbol{b}^T\boldsymbol{b} - \boldsymbol{p}^T\boldsymbol{b}$$

Huzzah, again!