18.06.21: Last bit on projections

Lecturer: Barwick

Friday, April Fool's Day 2016

Let's remember our story. We have a vector $\vec{b} \in \mathbf{R}^n$. We have a k-dimensional subspace $W \subseteq \mathbf{R}^n$ with basis $\{\vec{a}_1, \ldots, \vec{a}_k\}$. We are trying to compute the projection $\pi_W(\vec{b})$ of $\vec{b}$ onto W.

The first method we discussed was to apply Gram–Schmidt to $\vec{a}_1, \ldots, \vec{a}_k$ to get an orthogonal basis $\vec{u}_1, \ldots, \vec{u}_k$, and then write

$$\pi_W(\vec{b}) = \sum_{i=1}^{k} \pi_{\vec{u}_i}(\vec{b}).$$

The second method was a simple formula: one forms the $n \times k$ matrix

$$A = \begin{pmatrix} \vec{a}_1 & \cdots & \vec{a}_k \end{pmatrix},$$

and we write

$$\pi_W(\vec{b}) = A(A^\intercal A)^{-1} A^\intercal \vec{b}.$$

To deduce this formula, we only used two facts about the projection $\pi_W(\vec{b})$:

(1) $\vec{b} - \pi_W(\vec{b})$ is perpindicular to W, and

(2) $\pi_W(\vec{b}) \in W$.

As always, we should test out the extreme cases. When $k = n$, we're "projecting" $\vec{b}$ onto the whole damn $\mathbf{R}^n$. In other words, we're not doing anything. And the formula above matches that, because $(A^\intercal A)^{-1} = A^{-1}(A^\intercal)^{-1}$.

Onthe other hand, when $k = 1$, our matrix A is just the column vector $\vec{a}_1$ itself. Then we get

$$\pi_W(\vec{b}) = \vec{a}_1(\vec{a}_1^\intercal \vec{a}_1)^{-1}\vec{a}_1^\intercal \vec{b} = \vec{a}_1 \frac{\vec{a}_1 \cdot \vec{b}}{\vec{a}_1 \cdot \vec{a}_1} = \pi_{\vec{a}_1}(\vec{b}).$$

Good.

So let's appreciate how good this is: let's project any vector $\vec{b} \in \mathbf{R}^3$ onto the plane W given by the equation $x - y + z = 0$. Here's what I have to do:

(1) Find a basis $\{\vec{v}_1, \vec{v}_2\}$ for that plane. That's the kernel of the 1×3 matrix $\begin{pmatrix} 1 & -1 & 1 \end{pmatrix}$.

(2) Now we put that basis into a matrix A, and we compute $A(A^\intercal A)^{-1} A^\intercal$.

Suppose A is an $n \times k$ matrix with $\ker(A) = 0$. Now set

$$\Pi_A \coloneqq A(A^\intercal A)^{-1} A^\intercal.$$

This is our *orthogonal projection matrix onto the image of A.*

Question. What is Π_A^2? Explain why first computationally, and then geometrically.

There's one thing we need to double-check with this formula before we stop. I claimed that $A^\intercal A$ is invertible when $\vec{a}_1, \ldots, \vec{a}_k$ are linearly independent. To prove that, it suffices to show that $\ker(A^\intercal A) = \ker(A)$. (Why??)

Clearly if $A\vec{x} = \vec{0}$, then $A^\intercal A\vec{x} = \vec{0}$, so $\ker(A) \subseteq \ker(A^\intercal A)$.

Let's prove the other inclusion. If $A^\intercal A\vec{x} = \vec{0}$, then

$$\|A\vec{x}\|^2 = (A\vec{x}) \cdot (A\vec{x}) = \vec{x}^\intercal A^\intercal A\vec{x} = 0.$$

So $A\vec{x} = 0$. This proves that $\ker(A) \supseteq \ker(A^\intercal A)$. $\square$

(Note that I didn't really use that $\vec{a}_1, \ldots, \vec{a}_k$ are linearly independent in that proof. It's *always* the case that $\ker(A^\intercal A) = \ker(A)$.)

EXAM III COVERS EVERYTHING UP TO HERE

(except for the fun stuff on special relativity)

That's Lectures 1–21, excluding 17 and 18. It's also the first 4 chapters of Strang, excluding 4.3.

OK … sigh … *determinants*.

There are two serious pedagogical problems with introducing determinants:

(1) They're extremely useful, but generally extremely annoying to compute.

(2) They have beautiful formal properties, but to show you why, I'd need to introduce a whole pack of auxiliary, abstract, notions that you won't see again until a much later math course.

No matter what, it's hard to make people happy when talking about determinants.