18.06.20: Projections and Gram–Schmidt

Lecturer: Barwick

Wednesday, 30 March 2016

Once again …the Gram–Schmidt orthogonalization/orthonormalization process:

(1) We start with the vector $\vec{v}_1$. The only problem there is that it's not a unit vector. So we take $\vec{u}_1 \coloneqq \vec{v}_1$, and we normalize it: $\hat{u}_1 \coloneqq \frac{1}{\|\vec{u}_1\|}\vec{u}_1$.

(2) Next, we take the vector $\vec{v}_2$, and we remove the best approximation to $\vec{v}_2$ that lies in W_1:

$$\vec{u}_2 \coloneqq \vec{v}_2 - \pi_{W_1}(\vec{v}_2) = \vec{v}_2 - \pi_{\vec{u}_1}(\vec{v}_2),$$

and we normalize it: $\hat{u}_2 = \frac{1}{\|\vec{u}_2\|}\vec{u}_2$.

(3) Now, we take $\vec{v}_3$. Here we have

$$\vec{u}_3 \coloneqq \vec{v}_3 - \pi_{W_2}(\vec{v}_3) = \vec{v}_3 - \pi_{\vec{u}_1}(\vec{v}_3) - \pi_{\vec{u}_2}(\vec{v}_3),$$

and we normalize: $\hat{u}_3 = \frac{1}{\|\vec{u}_3\|}\vec{u}_3$.

(4) We can keep doing this. We write

$$\vec{u}_i = \vec{v}_i - \pi_{W_{i-1}}(\vec{v}_i) = \vec{v}_i - \pi_{\vec{u}_1}(\vec{v}_i) - \cdots - \pi_{\vec{u}_{i-1}}(\vec{v}_i),$$

and we normalize: $\hat{u}_i = \frac{1}{\|\vec{u}_i\|}\vec{u}_i$.

Animation in $\mathbf{R}^3$ … stolen from Wikipedia!

Let's look at this collection $\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4\}$ of 4 linearly independent vectors in $\mathbf{R}^5$:

$$\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} \right\},$$

and let's begin by just orthogonalizing it, without worrying about normalizing.

There are 4 steps:

(1) We won't even touch the first vector:

$$\vec{u}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

(2) Next, let's remove the projection of $\vec{v}_2$ onto $\vec{u}_1$ from $\vec{v}_2$:

$$\vec{u}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -1/2 \\ 1/2 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

(3) Next, we remove the projections of $\vec{v}_3$ onto $\vec{u}_1$ and $\vec{u}_2$ from $\vec{v}_3$:

$$\vec{u}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} - 0 - \frac{1}{3/2}\begin{pmatrix} -1/2 \\ 1/2 \\ 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1/3 \\ -1/3 \\ 1/3 \\ 1 \\ 0 \end{pmatrix}.$$

(4) Finally, we remove the projections of $\vec{v}_4$ onto $\vec{u}_1$, $\vec{u}_2$ and $\vec{u}_3$ from $\vec{v}_4$:

$$\vec{u}_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} - 0 - 0 - \frac{1}{4/3} \begin{pmatrix} 1/3 \\ -1/3 \\ 1/3 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -1/4 \\ 1/4 \\ -1/4 \\ 1/4 \\ 1 \end{pmatrix}.$$

This gives us our desired orthogonal collection of vectors:

$$\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -1/2 \\ 1/2 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1/3 \\ -1/3 \\ 1/3 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1/4 \\ 1/4 \\ -1/4 \\ 1/4 \\ 1 \end{pmatrix} \right\},$$

and we note with pride that each $\vec{u}_i$ here can be written as a linear combination of $\vec{v}_1, \ldots, \vec{v}_i$. Cool.

Suppose I want to project the vector $\vec{b} := \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ 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. To compute projections, we're supposed to work with an orthogonal basis, but $\{\vec{v}_1, \vec{v}_2\}$ probably won't be orthogonal, so we'll have to orthogonalize to get a new basis $\{\vec{u}_1, \vec{u}_2\}$
3. Finally, we can compute $\pi_W(\vec{b}) = \pi_{\vec{u}_1}(\vec{b}) + \pi_{\vec{u}_2}(\vec{b})$.

Computationally, this approach may not make you very happy.

We can be more efficient by abstracting our process some. (This is a general lesson in math! Well-adapted abstractions yield efficiency!)

If we're projecting a vector $\vec{b} \in \mathbf{R}^n$ onto a k-dimensional subspace $W \subset \mathbf{R}^n$ spanned by some linearly independent (but not necessarily orthogonal!!) vectors $\vec{a}_1, \ldots, \vec{a}_k$, then we know that the difference $\vec{b} - \pi_W(\vec{b})$ will be perpindicular to W. That means it will be perpindicular to each element of our basis $\vec{a}_1, \ldots, \vec{a}_k$.

So we have:

$$(\vec{a}_i)^\intercal(\vec{b} - \pi_W(\vec{b})) = \vec{a}_i \cdot (\vec{b} - \pi_W(\vec{b})) = 0$$

for each i. Putting all k of those equations gives us

$$A^\intercal(\vec{b} - \pi_W(\vec{b})) = 0,$$

where $A = \begin{pmatrix} \vec{a}_1 & \cdots & \vec{a}_k \end{pmatrix}$. (Note that A is an $n \times k$ matrix, so $A^\intercal$ is a $k \times n$ matrix.) Thus

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

There are probably lots of vectors $\vec{c}$ out there such that $A^\intercal\vec{b} = A^\intercal\vec{c}$, but one thing singles out our friend $\pi_W(\vec{b})$: it lies in W! That is, it is in the *image* of A.

So … there's some vector $\vec{w} \in \mathbf{R}^k$ such that $\pi_W(\vec{b}) = A\vec{w}$, and for this vector we have

$$A^\intercal \vec{b} = A^\intercal A\vec{w}.$$

Now here's the (actually kind of surprising) fact: the fact that the vectors $\vec{a}_1, \ldots, \vec{a}_k$ are linearly independent actually implies that $A^\intercal A$ (which is a $k \times k$ matrix) is *invertible*. That means that the equation above actually *uniquely* specifies $\vec{w}$ in terms of $\vec{b}$.

We can thus write a formula for $\vec{w}$:

$$\vec{w} = (A^\intercal A)^{-1} A^\intercal \vec{b},$$

and we get a formula for $\pi_W(\vec{b})$ as well:

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

Let's appreciate how good this is: let's write a *formula* for the projection of 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 $A(A^\intercal A)^{-1} A^\intercal$.

Bam! One-stop shopping for projections.

We can use this to modify Gram–Schmidt slightly. Let's try it with

$$\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} \right\},$$