18.06.19: Gram–Schmidt Crackers

Lecturer: Barwick

Monday, 28 March 2016

A collection of vectors $\vec{v}_1, \ldots, \vec{v}_k \in \mathbf{R}^n$ are said to be *orthogonal* if any two of them are perpindicular, i.e., if

$$\vec{v}_i \cdot \vec{v}_j = 0 \text{ if } i \neq j.$$

More particularly, $\vec{v}_1, \ldots, \vec{v}_k \in \mathbf{R}^n$ are said to be *orthnormal* if any one of them is a unit vector, and any two of them are perpindicular, i.e., if

$$\vec{v}_i \cdot \vec{v}_j = \delta_{ij} = \begin{cases} 0 & \text{if } i \neq j; \\ 1 & \text{if } i = j. \end{cases}$$

Recall from the first problem set that an orthonormal collection of vectors is linearly independent. Of course, there are lots of linearly independent collections of vectors that aren't orthonormal. That's an issue when it comes to understanding their geometry.

Why? Well, suppose we have a vector subspace $W \subseteq \mathbf{R}^n$, and suppose we have an orthonormal basis $\{\hat{u}_1, \dots, \hat{u}_k\}$ of W. So take some vectors $\vec{a}, \vec{b} \in W$ and write them as

$$\vec{a} = \sum_{i=1}^{k} a_i \hat{u}_i \quad \text{and} \quad \vec{b} = \sum_{i=1}^{k} b_i \hat{u}_i.$$

We then find

$$\vec{a} \cdot \vec{b} = \sum_{i=1}^{k} a_i b_i.$$

In other words, the geometry of $\vec{a}$ and $\vec{b}$ can be extracted with a minimum of thought from these coefficients.

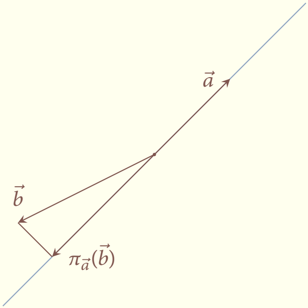
But now we run into a problem: when we're thinking about the kinds of bases for subspaces we can get our hands on, nothing makes them orthonormal. Our ways of computing kernels of bases, for example, don't ensure any orthonormality.

So there's the question: we'd like a way of taking some linearly independent vectors $\vec{v}_1, \dots, \vec{v}_k \in \mathbf{R}^n$ and generating a new, orthonormal, collection of vectors $\hat{u}_1, \dots, \hat{u}_k \in \mathbf{R}^n$ such that for any i, we have

$$\operatorname{im}\begin{pmatrix} \hat{u}_1 & \dots & \hat{u}_i \end{pmatrix} = \operatorname{im}\begin{pmatrix} \vec{v}_1 & \dots & \vec{v}_i \end{pmatrix}.$$

Recall: the *projection* of a vector $\vec{b}$ onto a vector $\vec{a}$ is the vector

$$\pi_{\vec{a}}(\vec{b}) \coloneqq (\hat{a} \cdot \vec{b})\hat{a} = \frac{\vec{a} \cdot \vec{b}}{\vec{a} \cdot \vec{a}}\vec{a}.$$



We spoke a little about this, but we let's examine more. First of all, what happens if you scale the vector $\vec{a}$ onto which you're projecting?

$$\pi_{r\vec{a}}(\vec{b}) = \frac{(r\vec{a}) \cdot \vec{b}}{(r\vec{a}) \cdot (r\vec{a})}(r\vec{a}) = \frac{r(\vec{a} \cdot \vec{b})}{r^2(\vec{a} \cdot \vec{a})}(r\vec{a}) = \frac{\vec{a} \cdot \vec{b}}{\vec{a} \cdot \vec{a}}\vec{a} = \pi_{\vec{a}}(\vec{b}).$$

So we can think of this not as the projection of $\vec{b}$ onto $\vec{a}$, but as the projection of $\vec{b}$ onto the *line* L spanned by $\vec{a}$. We write $\pi_L(\vec{b})$.

Next, the projection is sometimes called the *orthogonal projection*, because you're taking $\vec{b}$, and $\pi_L(\vec{b})$ is the approximation to $\vec{b}$ in L that *differs* from $\vec{b}$ by a perpindicular vector. That is,

$$(\vec{b} - \pi_L(\vec{b})) \cdot \vec{a} = 0.$$

(Why?)

Let's try to generalize this. Suppose I have a vector subspace $W \subseteq \mathbf{R}^n$. We might like to define the *projection* of $\vec{b}$ onto W.

How might we do this? Well, suppose we have a orthogonal basis $\{\vec{u}_1, \ldots, \vec{u}_k\}$ of W. Then we can write

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

The difference $\vec{b} - \pi_W(\vec{b})$ is perpindicular to W; that is, if $\vec{w} \in W$, then

$$(\vec{b} - \pi_W(\vec{b})) \cdot \vec{w} = 0.$$

To explain this, let's remember a vector is a perpindicular to a subspace if and only if it's perpindicular to a basis for that subspace. So all we need to show is that for any j,

$$(\vec{b} - \pi_W(\vec{b})) \cdot \vec{u}_j = 0.$$

But that's true:

$$\begin{aligned}(\vec{b} - \pi_W(\vec{b})) \cdot \vec{u}_i &= (\vec{b} - \sum_{i=1}^{k} \pi_{\vec{u}_i}(\vec{b})) \cdot \vec{u}_j \\ &= (\vec{b} - \pi_{\vec{u}_j}(\vec{b})) \cdot \vec{u}_j - \sum_{i \neq j} \pi_{\vec{u}_i}(\vec{b}) \cdot \vec{u}_j = \vec{0} - \sum_{i \neq j} \vec{0} = \vec{0},\end{aligned}$$

because each $\pi_{\vec{u}_i}(\vec{b})$ is a scalar multiple of $\vec{u}_i$.

Here's an key fact: if you have a different orthogonal basis $\{\vec{v}_1, \dots, \vec{v}_k\}$ of W, then

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

So if we have a linearly independent set $\vec{v}_1, \dots, \vec{v}_k \in \mathbf{R}^n$ of vectors, let's think about how to use these projections to extract what we want. Again, the goal is to get a new, orthonormal, collection of vectors $\hat{u}_1, \dots, \hat{u}_k \in \mathbf{R}^n$ such that for any i, we have

$$\operatorname{im}\begin{pmatrix} \hat{u}_1 & \dots & \hat{u}_i \end{pmatrix} = W_i,$$

where

$$W_i \coloneqq \operatorname{im}\begin{pmatrix} \vec{v}_1 & \dots & \vec{v}_i \end{pmatrix},$$

so that we have this sequence of subspaces

$$0 = W_0 \subset W_1 \subset \cdots \subset W_k = W.$$

So now we can think about the 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 something important happens: we want to write

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

but to compute this, we need an orthogonal basis of W_2. But good news! We created it in the last step! So

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

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