Math 314

Topics since the third exam

The final exam is on Wednesday, May 5, from 10:00am to noon. It will cover the material from the entire course, with a slight emphasis on the material from this sheet.

Chapter 5: Norms and inner products (again)

§3: Gram-Schmidt orthogonalization

We've seen how a basis consitsing of vectors orthogonal to one another can prove useful; this section is about how to *build* such a basis.

The starting point is our old formula for the projection of one vector onto another;

$$v - \frac{< w, v >}{< w, w >} w \text{ is perpendicular to } w.$$

Gram-Scmidt orthogonalization consists of repeatedly using this formula to replace a collection of vectors with ones that are orthogonal to one, **without changing their span**. Starting with a collection $\{v_1, \ldots, v_n\}$ of vectors in V,

let $w_1 = v_1$, then let $w_2 = v_2 - \frac{<w_1,v_2>}{<w_1,w_1>} w_1$.

Then w_1 and w_2 are orthogonal, and since w_2 is a linear combination of $w_1 = v_1$ and v_2, while the above equation can also be rewritten to give v_2 as a linaear combination of w_1 and w_2, the span is unchanged. Continuing,

let $w_3 = v_3 - \frac{< w_1, v_3 >}{< w_1, w_1 >} w_1 - \frac{< w_2, v_3 >}{< w_2, w_2 >} w_2$; then since w_1 and w_2 are orthogonal, it is not hard to check that w_3 is orthogonal to **both** of them, and using the same argument, the span is unchanged (in this case, span$\{w_1, w_2, w_3\}$ =span$\{w_1, w_2, v_3\}$=span$\{v_1, v_2, v_3\}$).

Continuing this, we let $w_k = v_k - \frac{< w_1, v_k >}{< w_1, w_1 >} w_1 - \ldots - \frac{< w_{k-1}, v_k >}{< w_{k-1}, w_{k-1} >} w_{k-1}$

Doing this all the way to n will replace $v_1, \ldots, v_n$ with orthogonal vectors $w_1, \ldots, w_n$, without changing the span.

One thing worth noting is that the if two vectors are orthogonal, then any scalar multiples of them are, too. This means that if the coordinates of one of our w_k are not to our satisfaction (having an ugly denomenator, perhaps), we can scale it to change the coordinates to something more pleasant. It is interesting to note that in so doing, the the later vectors w_k are unchanged, since our scalar, can be pulled out of both the top inner product and the bottom one in later calculations, and cancelled.

We've seen that if $w_1, \ldots, w_n$ is an **orthogonal basis** for a subspace W of V, and $w \in W$, then $w = \frac{< w_1, w >}{< w_1, w_1 >} w_1 + \ldots + \frac{< w_{k-1}, w >}{< w_{k-1}, w_{k-1} >} w_{k-1}$

On the other hand, if $v \in V$, we can define the orthogonal projection

$$\text{proj}_W(v) = \frac{< w_1, v >}{< w_1, w_1 >} w_1 + \ldots + \frac{< w_{k-1}, v >}{< w_{k-1}, w_{k-1} >} w_{k-1}$$

of v into W. This vector is in W, and by the Gram-Schmidt argument, $v - \text{proj}_W(v)$ is orthogonal to all of the w_i, so it is orthogonal to every linear combination, i.e., it is orthonal to every vector in W. As a result:

$||v - \text{proj}_W(v)|| \leq ||v - w||$ for **every** vector w in W. (**)

In the case that the w_i are not just orthogonal but also *orthnormal*, we can simplify this somewhat:

$\text{proj}_W(v) = < w_1, v > w_1 + \cdots + < w_n, v > w_n = (w_1 w_1^T + \cdots + w_n w_n^T)v = Pv$,

where $P = (w_1 w_1^T + \cdots + w_n w_n^T)$ is the **projection matrix** giving us orthogonal projection.

This projection matrix has three useful properties: (1) since it has the property (**), the matrix you get will be the same no matter what orthonormal basis you will use to build it; (2) it is symmetric ($P^T = P$), and (3) it is idempotent, meaning $P^2 = P$ (this is because the orthogonal projection of a vector in W (e.g., Pv) is the same vector).

If we think of the vectors w_i as the columns of a matrix A, then $W = \mathcal{C}(A)$, and so the result (**) is talking about the least squares solution to the equation $Ax = v$! The closest vector Ax to v is then Pv, which, looking at what we did before, means that $P = A(A^TA)^{-1}A^T$. This, however, makes sense even if the columns of A are **not** orthogonal; if we picked orthonormal ones, and computed P, we would **still** get the least squares solution, which this formula **also** gives!

§4: Orthogonal matrices

We've seen that having a basis consisting of orthonormal vectors can simplify some of our previous calculations. Now we'll see where some of them come from.

An $n \times n$ matrix Q is called **orthogonal** if it's columns form an orthonormal basis for R^n. This means <(ith column of Q),(jth column of Q¿ = 1 if $i = j$, 0 otherwise . This in turn means that $Q^TQ = I$, which in turn means $Q^T = Q^{-1}$! So an orthogonal matrix is one whose inverse is equal to its own transpose.

A basic fact about an orthogonal matrix Q : for any $v, w \in R^n$, $< Qv, Qw > = < v, w >$.

A basic fact about a symmetric matrix A : if v_1 and v_2 are eigenvectors for A with different eigenvalues λ_1, λ_2, then v_1 and v_2 are orthogonal.

This is a main ingredient needed to show: If A is a symmetric $n \times n$ matrix, then A is always diagonalizable; in fact there is an orthonormal basis for R^n consisting of eigenvectors of A. This means that the matrix P, with $AP = PD$, whose columns are a basis of eigenvectors for A, can (when A is symmetric) be chosen to be an **orthogonal** matrix.

Wow, short section.

§5: Orthogonal complements

This notion of orthogonal vectors can even be used to reinterpret some of our dearly-held results about systems of linear equations, where all of this stuff began.

Starting with $Ax = 0$, this can be interpreted as saying that <(every row of A),x >=0, i.e., x is orthogonal to every row of A. This in turn implies that x is orthogonal to every linear combination of rows of A, i.e., x is orthogonal to every vector in the row space of A.

This leads us to introduce a new concept: the **orthogonal complement** of a subspace W in a vector space V, denoted $W^\perp$, is the collection of vectors v with $v \perp w$ for **every** vector $w \in W$. It is not hard to see that these vectors form a subspace of V; the sum of two vectors orthogonal to w, for example, is orthogonal to w, so the sum of two vectors in $W^\perp$ is also in $W^\perp$. The same is true for scalar multiples.

Some basic facts:

For every subspace W, $W \cap W^\perp = \{0\}$ (since anything in both is orthogonal to *itself*, and only the 0-vector has that property).

Any vector $v \in V$ can be written, uniquely, as $v = w + w^\perp$, for $w \in W$ and $w^\perp \in W^\perp$; w in fact is $\text{proj}_W(v)$. $v - \text{proj}_W(v)$ will be in $W^\perp$, more or less by definition of $\text{proj}_W(v)$. The uniqueness comes from the result above about intersections.

Even further, a basis for W and a basis for $W^\perp$ together form a basis for V; this implies that $\dim(W)+\dim(W^\perp) = \dim(V)$.

Finally, $(W^\perp)^\perp = W$; this is because W is contained in $(W^\perp)^\perp$ (a vector in W is orthogonal to every vector that is orthogonal to things in W), and the dimensions of the two spaces are the same.

The importance that this has to systems of equations stems from the following facts:

$\mathcal{N}(A) = \mathcal{R}(A)^\perp$ (this is what we noted, actually, at the beginning of this section!)

$\mathcal{R}(A) = \mathcal{N}(A)^\perp$

$\mathcal{C}(A) = \mathcal{N}(A^T)^\perp$

So, for example, to compute a basis for $W^\perp$, start with a basis for W, writing them as the columns of a matrix A, so $W = \mathcal{C}(A)$, then $W^\perp = \mathcal{C}(A)^\perp = \mathcal{R}(A^T)^\perp = \mathcal{N}(A^T)$, which we know how to compute a basis for!