Gram-Schmidt Ortho Notes PhL 12.12.04

Theorem 2A (GS orthogonal): Suppose the column vectors of a matrix U are linearly independent, *but not necessarily orthogonal or normalized*. Then we can find an upper triangular matrix R with unity diagonal elements (hence detR=1), which converts the matrix U to the matrix V whose column vectors *are* orthogonal (but not necessarily normalized). The conversion is $V = UR^{-1}$. Basically, this conversion is the Gram-Schmidt orthogonalization procedure written in matrix form. [It could be that the matrix U consists of the column vectors u_i of an eigenvalue problem $A\mathbf{u_i} = \lambda_i\mathbf{u_i}$, but I don't think this fact is relevant to GS.]

Proof: If we look at M&M page 313, we get this notion of doing GS

v1 = u1	
v2 = u2 - αv1	such that v2•v1 = 0
v3 = u3 - α'v1 - β' v2	such that v3•v1= 0 and v3•v2 = 0
v4 = u4 - α" v1 - β"v2 - γ"v3	etc

I think this is completely reasonable and bulletproof. Now let's write again as

v1 = u1
v2 = u2 - a21 v1
v3 = u3 - a31 v1 - a32 v2
v4 = u4 - a41 v1 - a42 v2 - a43 v3

Now solve for the u's to get

u1 = v1
u2 = v2 + a21 v1
u3 = v3 + a31 v1 + a32 v2
u4 = v4 + a41 v1 + a42 v2 + a43 v3

Now remember that each **u** and each **v** is a column vector. So rewrite the above in matrix notation this way

$$\begin{bmatrix} & & & \\ u1 & u2 & u3 & u4 \\ & & & \end{bmatrix} = \begin{bmatrix} & & & \\ v1 & v2 & v3 & v4 \\ & & & \end{bmatrix} \begin{bmatrix} 1 & a21 & a31 & a41 \\ & 1 & a32 & a42 \\ & & 1 & a43 \\ & & & 1 \end{bmatrix}$$

which we can rewrite as

$$U = VR$$

where R is the upper triangular matrix shown. Note that det(R) = 1 according to Theorem 2H above. Note also that the matrix R is uniquely defined by the above procedure. There is only one matrix R that does the job we have assigned to it above.

Here is a second approach to the same matrix forms. Start again with the above equations.

v1 = u1
v2 = u2 - a21 v1
v3 = u3 - a31 v1 - a32 v2
v4 = u4 - a41 v1 - a42 v2 - a43 v3

We want to rewrite these so we have all u's on the right side, so "just do it"

v1 = u1
v2 = u2 - a21 u1
v3 = u3 - a31 u1 - a32 (u2 - a21 u1) = [u3 - a32 u2 - (a31-a32*a21) u1]
v4 = u4 - a41 u1 - a42 (u2 - a21 u1) - a43 [u3 - a32 u2 - (a31-a32*a21) u1]

which we now rewrite as

v1 = u1
v2 = u2 - b21 u1
v3 = u3 - b31 u1 - b32 u2
v4 = u4 - b41 u1 - b42 v2 - b43 u3

and this is the same as our solution for u in terms of the v, but we just have different coefficients. We then get this matrix equation

V = UP

where P is the same as R with the a replaced by b. Notice that detP = detR = 1, so both matrices are invertible. We conclude that PR = 1, so our two matrices are just inverses of each other.

Comments: The transformation U = VR does not describe a *rotation* of the column vectors v_i into u_i. Instead, this transformation causes a *mixing* of the column vectors. Look at the first two equations,

u1 = v1 $= e_1$
u2 = v2 + a21 v1 $= e_2 + a_{21}\, e_1$

If we start with V = unit vectors as shown, then $v_1 \bullet v_2 = 0$. However, $u_1 \bullet u_2 = a_{21} \neq 0$. Since a rotation must preserve dot products, this cannot be a rotation, despite the fact that detR = 1.

Suppose we had the equation U = RV instead. This at least has the right *form* to be a rotation, because it says that $\mathbf{u_i} = R\, \mathbf{v_i}$, saying we are shuffling components of each column vector to make the new vector, not mixing vectors. If R were real orthogonal, this would in fact describe a rotation of the column vectors, whereas U = VR still would *not* describe a rotation of same.

To summarize:

(1) The transformation U = VR results in a mixing of column vectors and does not describe a rotation of these vectors even if R is real orthogonal. However, were R real orthogonal, we could say that R describes a rotation of the *row* vectors! But in our formalism, we work with column vectors. We could work in a transpose formalism, $x^T A^T = \lambda x^T$, but we don't choose to do so.

(2) The transformation U = RV does describe a rotation of the column vectors if R is real orthogonal. But this is not the equation we get by doing GS. And this transformation mixes row vectors.

****

Theorem 2B (GS orthonormal): Suppose the column vectors of a matrix U are linearly independent, *but not necessarily orthogonal or normalized*. Then we can find an upper triangular matrix C with all-positive diagonal elements (hence detC > 0), which converts the matrix U to the matrix V whose column vectors are *orthonormal*. The conversion is $V = UC^{-1}$. Basically, this conversion is the Gram-Schmidt orthonormalization procedure written in matrix form. [It could be that the matrix U consists of the column vectors u_i of an eigenvalue problem $A\mathbf{u_i} = \lambda_i \mathbf{u_i}$, but I don't think this fact is relevant to GS.]

Here we just modify the presentation given in the proof of Theorem 2A above:

v1' = u1
v1 = v1'/ ||v1'||

v2' = u2 - αv1 such that v2•v1 = 0
v2 = v2'/ ||v2'||

v3' = u3 - α'v1 - β' v2 such that v3•v1= 0 and v3•v2 = 0
v3 = v3'/ ||v3'||

v4' = u4 - α" v1 - β"v2 - γ"v3 etc
v4 = v4'/ ||v4'||

Now rewrite the results in this form: (noticing that $c_{mm} = 1/ \|v_m'\| > 0$)

v1 = c11 u1
v2 = c22 u2 - c21 v1
v3 = c33 u3 - c31 v1 - c32 v2
v4 = c44 u4 - c41 v1 - c42 v2 - c43 v3

Solve for the u as we did before, then write the result as a matrix to get

$$\begin{bmatrix} & & & \\ u1 & u2 & u3 & u4 \\ & & & \end{bmatrix} = \begin{bmatrix} & & & \\ v1 & v2 & v3 & v4 \\ & & & \end{bmatrix} \begin{bmatrix} c11 & c21 & c31 & c41 \\ & c22 & c32 & c42 \\ & & c33 & c43 \\ & & & c44 \end{bmatrix}$$

Which says that U = VC. The only difference compared to the previous method is that the diagonal elements of C are no longer unity, and of course the other elements are different as well. Also, notice that the diagonal elements are positive definite, so detC > 0 and C is therefore invertible, so $V = UC^{-1}$.

Comments on Theorem 2A and 2B. When I first wrote these up, I was thinking that the matrix like R or C was going to play a role in transforming an eigenvalue problem AU = UΛ into diagonal form with those desirable unit vector eigenvectors, or at least to some kind of "block diagonal form" like that shown in M&M Chapter 10. I was thinking maybe we were going to get something like $(C^{-1}AC)V = V\Lambda$ where the V are the nice unit eigenvectors V = I, so really we have $(C^{-1}AC) = \Lambda$. It turns out that there IS a matrix that, under some conditions, can diagonalize A in this manner, but it is not the matrix C that comes from the GS process. It is of course the matrix U itself, and as long as U is invertible, we can write
$AU = U\Lambda \Rightarrow (U^{-1}AU) = \Lambda$. So it turns out that the matrix C *plays no role whatsoever* in this transformation of the matrix A. My conclusion is that GS does not really contribute to the eigenvalue problem. It is more useful in things like finding orthogonal polynomials. Just generally, it tells you how to create an orthonormal basis from a set of linearly independent vectors. In general, you can even see how GS might be very *un-useful* in the eigenvalue problem, because it wants to mix eigenvectors of different eigenvalues. You might find some role for GS *within an eigenmanifold* of some λ.

****

Comments on the QR factorization:

We know that it is possible to write A = QR where Q is real orthogonal and R is upper triangular. Looking at what we have done above in the first section, we see U = VR where V was an orthonormal basis vector matrix (ie, real orthogonal) and R was upper triangular. This is an example of A = QR factorization for an A that has $\det A \neq 0$. Thus, you can interpret A = QR factorization as in fact being the G-S orthogonalization of the non-singular matrix A into the orthogonal matrix Q.