18.06 Professor Johnson Quiz 1 October 3, 2007

SOLUTIONS

1 (30 pts.) A given circuit network (directed graph) which has an $m \times n$ incidence matrix A (rows = edges, columns = nodes) and a conductance matrix C [diagonal = inverse of the (positive) resistance of each edge] given by:

$$A = \begin{pmatrix} 0 & 1 & -1 \\ 0 & 1 & -1 \\ 1 & 0 & -1 \\ 1 & -1 & 0 \end{pmatrix} \qquad C = \begin{pmatrix} 1/R_1 & 0 & 0 & 0 \\ 0 & 1/R_2 & 0 & 0 \\ 0 & 0 & 1/R_3 & 0 \\ 0 & 0 & 0 & 1/R_4 \end{pmatrix}.$$

Suppose the unknowns are the vector $\mathbf{v}$ of voltages at each node, and you are given a vector $\mathbf{d}$ of applied voltage drops across each edge (e.g. if you connect a battery to each edge). In this case, Kirchhoff's laws plus Ohm's law gives the equation:

$$A^T C A \mathbf{v} = A^T C \mathbf{d}.$$

(a) Sketch the network, labelling each edge from 1 to 4 corresponding to each row of A, and each node from ① to ③ corresponding to each column of A, and put an arrow to show the direction of each edge.

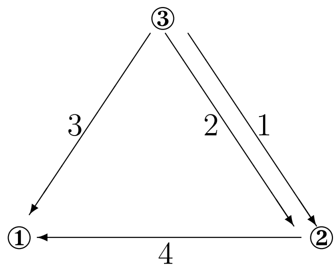
(b) Is $A^T C A \mathbf{v} = A^T C \mathbf{d}$ always solvable for all $\mathbf{d}$? Why or why not? [You can use the fact, from class, that $\text{rank}(A^T C A) = \text{rank}(A) = n - 1$. Hint: think about the subspaces; little or no calculation is necessary. This is *not* the same as whether $A^T C A \mathbf{v} = \mathbf{s}$ is solvable for all $\mathbf{s}$.]

(c) Solve for $\mathbf{v}$ when $C = I$ (all resistances = 1) and $\mathbf{d} = \begin{pmatrix} 5 & 0 & 0 & 0 \end{pmatrix}^T$. To get a unique solution, set the voltage on node 1 to $v_1 = 0$ ("ground")—this simplifies life to a 3×2 matrix problem, since you then only have 2 unknowns v_2 and v_3. [Recall that the null space of A is the span of $\begin{pmatrix} 1 & 1 & 1 \end{pmatrix}^T$, so we can add any constant to the solutions.]

(d) For the same $\mathbf{d}$ as in (c), what is the minimum value (minimum over all $\mathbf{v}$) of $\|A\mathbf{v} - \mathbf{d}\|^2$?

Solution:

(a)



(b) Yes.

Since $\text{rank}(A^TCA) = \text{rank}(A) = n - 1$, and $\text{rank}(A^TCA) \leq \text{rank}(A^TC) \leq \text{rank}(A^T) = \text{rank}(A)$, we see that $\text{rank}(A^TCA) = \text{rank}(A^TC)$. Since the column space of A^TC contains the column space of A^TCA, and both have the same dimension, we see that the two column space are the same. Thus for any $\mathbf{d}$, $A^TC\mathbf{d}$ lies in the column space of A^TCA. In other words, the equation is always solvable.

Alternatively, it is also sufficient to say that the column space $C(A^TC)$ is clearly at least *contained* in $C(A^T) = C(A^TCA)$, with the latter equality because $C(A^T)$ and $C(A^TCA)$ have the same dimension [$\text{rank}(A^TCA) = \text{rank}(A) = \text{rank}(A^T)$] and $C(A^TCA) \subseteq C(A^T)$. (Recall that $C(AB) \subseteq C(A)$ for any A and B, since $AB\mathbf{x}$ is made of the columns of A.)

Common errors: Many students wrote that, since A^TCA is singular, there isn't always a solution—this is incorrect because the right-hand side is not an arbitrary vector, it is only vectors $A^TC\mathbf{d}$ in $C(A^TC)$. Several students wrote that we must have $A\mathbf{v} = \mathbf{d}$, which is not true since A^TC is not invertible (or even square). Many students wrote that, if you ignore the C, this is just like a least-squares problem and least-squares problems are always solvable—this is on the right track (the A^T on the right-hand side is truly the key here), but the least-squares problems we've studied were only when A has full column rank, which isn't true here.

(c) The equation becomes $A^T A\mathbf{v} = A^T\mathbf{d}$. We calculate:

$$A^T A = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 3 & -2 \\ -1 & -2 & 3 \end{pmatrix}, \qquad A^T\mathbf{d} = \begin{pmatrix} 0 \\ 5 \\ -5 \end{pmatrix},$$

so we can solve by elimination:

$$\begin{pmatrix} 2 & -1 & -1 & 0 \\ -1 & 3 & -2 & 5 \\ -1 & -2 & 3 & -5 \end{pmatrix} \to \begin{pmatrix} 2 & -1 & -1 & 0 \\ 0 & 5/2 & -5/2 & 5 \\ 0 & -5/2 & 5/2 & -5 \end{pmatrix} \to \begin{pmatrix} 2 & -1 & -1 & 0 \\ 0 & 5/2 & -5/2 & 5 \\ 0 & 0 & 0 & 0 \end{pmatrix} \to \begin{pmatrix} 2 & -1 & -1 & 0 \\ 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

so the general solution is $\mathbf{v} = \begin{pmatrix} a+1 & a+2 & a \end{pmatrix}^T$, where a is an arbitrary number (the multiple of the nullspace vector). To get $v_1 = 0$, let $a = -1$, and the unique solution is:

$$\mathbf{v} = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}.$$

A faster way: We can set $v_1 = 0$ immediately after constructing $A^T A$, equivalent to deleting the first column (which is multiplied by zero), leaving us with the 3×2 problem:

$$\begin{pmatrix} -1 & -1 & 0 \\ 3 & -2 & 5 \\ -2 & 3 & -5 \end{pmatrix} \to \begin{pmatrix} -1 & -1 & 0 \\ 0 & -5 & 5 \\ 0 & 5 & -5 \end{pmatrix} \to \begin{pmatrix} -1 & -1 & 0 \\ 0 & -5 & 5 \\ 0 & 0 & 0 \end{pmatrix}$$

which has the solution $v_3 = -1$, $v_2 = 1$ as above.

Another fast way: if we set $v_1 = 0$ at the very beginning, i.e. delete the first column of A to obtain a 4×2 matrix, then $A^T A = \begin{pmatrix} 3 & -2 \\ -2 & 3 \end{pmatrix}$ and $A^T \mathbf{d} = \begin{pmatrix} 5 \\ -5 \end{pmatrix}$, and the solution is the same as above.

(d) In order to minimize $\|A\mathbf{v} - \mathbf{d}\|^2$, we would solve $A^T A\mathbf{v} = A^T\mathbf{d}$, but this is precisely what we already did in part (c)! So, the minimum value is at $\mathbf{v} = \begin{pmatrix} 0 & 1 & -1 \end{pmatrix}^T$ from above, and is given by:

$$\left\| A \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} - \mathbf{d} \right\|^2 = \left\| \begin{pmatrix} 2 \\ 2 \\ 1 \\ -1 \end{pmatrix} - \begin{pmatrix} 5 \\ 0 \\ 0 \\ 0 \end{pmatrix} \right\|^2 = (-3)^2 + 2^2 + 1^2 + (-1)^2 = 15.$$

2 (30 pts.) Fill in the blanks below: (You don't need to justify your answer.)

(a) The nullspace of AB contains the nullspace of ___$\boxed{\text{B}}$___.

If $B\mathbf{x} = \mathbf{0}$ *then* $AB\mathbf{x} = \mathbf{0}$.

(b) Let P be the projection matrix to the row space of a matrix A, then $I - P$ is the projection matrix to ___$\boxed{\text{N(A)}}$___.

Reason: $I - P$ *is the projection onto the orthogonal complement, and the orthogonal complement of the row space is the nullspace.*

(c) Suppose A is an $m \times n$ matrix, and the row space of A is n dimensional, then its nullspace is ___$\boxed{0}$___ dimensional.

The rank r *of* A *is the dimension of the row space, so* $r = n$, *and the nullspace has dimension* $n - r = n - n = 0$.

(d) Let $\hat{\mathbf{x}}$ be the least-squares solution to $A\mathbf{x} = \mathbf{b}$. Then $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to the ___$\boxed{\text{column space}}$___ of A.

The least-squares solution solves $A\hat{\mathbf{x}} = P\mathbf{b}$, *where* P *is the projection onto* $C(A)$, *so* $\mathbf{b} - A\hat{\mathbf{x}} = \mathbf{b} - P\mathbf{b} = (I - P)\mathbf{b}$, *and* $I - P$ *projects onto the complement of* $C(A)$. *Equivalently, the least-squares solution finds the* closest *point* $A\hat{\mathbf{x}}$ *to* $\mathbf{b}$ *in* $C(A)$, *so the difference must be perpendicular to* $C(A)$.

(e) If $A^T = -A$ (A is antisymmetric), and A is $n \times n$ where n is odd, then $\det A =$ ___$\boxed{0}$___. .

Reason: $\det A = \det A^T = \det(-A) = (-1)^n \det A = -\det A$ *since* n *is odd, and the only way to have* $\det A = -\det A$ *is if* $\det A = 0$.

(f) If A is symmetric and P is the projection matrix onto the nullspace $N(A)$, then $PA =$___$\boxed{0}$___.

Since A *is symmetric,* $N(A) = N(A^T) = C(A)^\perp$, *so* P *projects onto the orthogonal complement of* $C(A)$. *Thus,* $PA = 0$ *since* P *projects every column of* A *to zero.*

3 (12 pts.) Construct an example of a least-square curve-fitting problem where the solution (the least-square fit parameters) is *not unique*. (You need not solve it, just write down the $A\mathbf{x} = \mathbf{b}$ equations that you *would* solve by least-squares to minimize $\|A\mathbf{x} - \mathbf{b}\|^2$.)

$\boxed{\text{Solution:}}$ The solution is not unique if the matrix A is not of full column rank. (Problem 1(c) is an example of this type.) *There are many possible examples.*

For example, you could have more unknowns than data points. e.g., you could be fitting to a line $C + Dt$, but only have a single point (t_1, b_1)—obviously, a single point is not enough to determine a line uniquely. In terms of matrices, $A = \begin{pmatrix} 1 & t_1 \end{pmatrix}$, $\mathbf{x} = \begin{pmatrix} C & D \end{pmatrix}^T$, and $\mathbf{b} = (b_1)$, which obviously does not have full column rank: there are more columns than rows in A!

Alternatively, you could be fitting a line $C + Dt$ to multiple points (t_1, b_1), (t_1, b_2), (t_1, b_3) etcetera with the same t coordinate—this is enough information to determine C or D but not both. In this case you get a matrix equation of the form:

$$\begin{pmatrix} 1 & t_1 \\ 1 & t_1 \\ \vdots & \vdots \\ 1 & t_1 \end{pmatrix} \begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix},$$

which obviously does not have full column rank: (column 2) $= t_1$(column 1) in A.

You could also construct an example where your fit parameters are not really independent, regardless of the data. For example, if you are fitting to $C + Dt + E(3t - 1)$, in which case E does not add any information because $3t - 1$ is a linear combination of C and Dt. Correspondingly, the third column of A will be three times the second column minus the first column.

Perhaps the most trivial example of all is where you have no data points whatsoever, in which case there is no information to constrain the fit. In terms of matrices, though, this is a bit too weird because it would correspond to a matrix A with zero rows, and we usually consider only matrices with positive sizes.

4 (28 pts.) Let $A = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 2 & 2 \\ 0 & 3 & 3 & 3 \\ 4 & 4 & 4 & 4 \end{pmatrix}$. Ordinary Gram-Schmidt would take the columns of A and produce an orthonormal basis Q with $C(A) = C(Q)$. In this problem, we will modify that process and see what happens.

In particular, suppose that we proceed as in Gram-Schmidt, but we omit the normalization steps—we construct a basis of *orthogonal* vectors spanning $C(A)$ but with lengths $\neq 1$, by subtracting the projections as in Gram-Schmidt but skipping the division by the lengths. Let's call this "unnormalized Gram-Schmidt."

(a) Do "unnormalized Gram-Schmidt" on A to get an orthogonal but not orthonormal basis B for $C(A)$.

(b) Compute two the (4-dimensional) volumes of the two parallelepipeds with edges given by the columns of A and the columns of B.

Recall how ordinary Gram-Schmidt corresponded to multiplying A by a sequence of matrices, leading to the QR decomposition. Now, we want to look at *unnormalized* Gram-Schmidt in the same way, in order to see what it does to the volume (determinant). The next two parts refer to an *arbitrary* $n \times n$ matrix A with independent columns, *not* the 4×4 matrix from parts (a) and (b).

(c) For an arbitrary $n \times n$ matrix A with (independent) columns $\mathbf{a}_1$, $\mathbf{a}_2$, etcetera, write down the matrix M_2 that you would multiply by A in the *first step* of unnormalized Gram-Schmidt to make the *second* column orthogonal to the first. What is $\det M_2$?

(d) Argue that the matrices M_3, M_4, etcetera that you would multiply by in subsequent steps of unnormalized Gram-Schmidt all have the same determinant as M_2. Therefore, the determinant of the final matrix B after unnormalized Gram-Schmidt is __ $\boxed{\det(A)}$ __?

Solution:

(a) Unnormalized Gram-Schmidt:

$$\begin{aligned}
\mathbf{b}_1 &= \mathbf{a}_1 = \begin{pmatrix} 0 & 0 & 0 & 4 \end{pmatrix}^T, \\
\mathbf{b}_2 &= \mathbf{a}_2 - \frac{\mathbf{b}_1^T \mathbf{a}_2}{\mathbf{b}_1^T \mathbf{b}_1} \mathbf{b}_1 = \begin{pmatrix} 0 & 0 & 3 & 0 \end{pmatrix}^T, \\
\mathbf{b}_3 &= \mathbf{a}_3 - \frac{\mathbf{b}_1^T \mathbf{a}_3}{\mathbf{b}_1^T \mathbf{b}_1} \mathbf{b}_1 - \frac{\mathbf{b}_2^T \mathbf{a}_3}{\mathbf{b}_2^T \mathbf{b}_2} \mathbf{b}_2 = \begin{pmatrix} 0 & 2 & 0 & 0 \end{pmatrix}^T, \\
\mathbf{b}_4 &= \mathbf{a}_4 - \frac{\mathbf{b}_1^T \mathbf{a}_4}{\mathbf{b}_1^T \mathbf{b}_1} \mathbf{b}_1 - \frac{\mathbf{b}_2^T \mathbf{a}_4}{\mathbf{b}_2^T \mathbf{b}_2} \mathbf{b}_2 - \frac{\mathbf{b}_3^T \mathbf{a}_4}{\mathbf{b}_3^T \mathbf{b}_3} \mathbf{b}_3 = \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix}^T.
\end{aligned}$$

Note that we must subtract off the projections onto the **b** vectors, not the **a** vectors—the **b** vectors span the same space, but are much simpler to project onto because they are orthogonal.

(Why does this work? When creating $\mathbf{b}_3$, for example, we should subtract off the projection of $\mathbf{a}_3$ onto the span of $\mathbf{a}_1$ and $\mathbf{a}_2$, which is some projection matrix. For ordinary Gram-Schmidt, where we have orthonormal vectors $\mathbf{q}$, the projection matrix simplifies to QQ^T and we can just subtract of projections onto each $\mathbf{q}$ individually. Here, the projection onto the span of the previous **b** vectors simplifies similarly. One way to think of this is just to realize that $\mathbf{q}_k = \mathbf{b}_k/\|\mathbf{b}_k\|$, and therefore we can apply the ordinary Gram-Schmidt step with this substitution, skipping the normalization. A more complicated way is to realize that $B^T B$, while not the identity as for Q, is a diagonal matrix, and this leads to the same simplified result.)

Thus, an orthogonal but not orthonormal basis B for $C(A)$ is

$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 2 & 0 \\ 0 & 3 & 0 & 0 \\ 4 & 0 & 0 & 0 \end{pmatrix}.$$

(b) The volume of the parallelepiped with edges given by the columns of A is just the determinant (or rather, its absolute value, but here the determinant is positive anyway):

$$\det(A) = \det \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 2 & 2 \\ 0 & 3 & 3 & 3 \\ 4 & 4 & 4 & 4 \end{pmatrix} = (-1)^2 \det \begin{pmatrix} 4 & 4 & 4 & 4 \\ 0 & 3 & 3 & 3 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix} = 4 \cdot 3 \cdot 2 \cdot 1 = 24,$$

where we have rearranged A into an upper-triangular matrix via two row swaps, and the determinant is then the product of the diagonals. The volume of the parallelepiped with edges given by the columns of B is similarly:

$$\det(B) = \det \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 2 & 0 \\ 0 & 3 & 0 & 0 \\ 4 & 0 & 0 & 0 \end{pmatrix} = (-1)^2 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 24 = \det(A).$$

(c) We should have

$$\begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{a}_3 & \cdots & \mathbf{a}_n \end{pmatrix} = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \cdots & \mathbf{a}_n \end{pmatrix} M_2.$$

Notice that we must multiply M_2 on the *right* since we are manipulating *columns* of A. Since $\mathbf{b}_1 = \mathbf{a}_1, \mathbf{b}_2 = \mathbf{a}_2 - \frac{\mathbf{b}_1^T \mathbf{a}_2}{\mathbf{b}_1^T \mathbf{b}_1} \mathbf{b}_1$, we must have

$$M_2 = \begin{pmatrix} 1 & -\frac{\mathbf{b}_1^T \mathbf{a}_2}{\mathbf{b}_1^T \mathbf{b}_1} & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

Since this is an upper-triangular matrix with 1's on the diagonal, we have $\det M_2 = 1$.

(d) In step k, we only change the vector $\mathbf{a}_k$ to $\mathbf{b}_k$, which is

$$\mathbf{b}_k = \mathbf{a}_k - \text{linear combination of } \mathbf{b}_1, \cdots, \mathbf{b}_{k-1},$$

thus the matrix M_k is an upper triangular matrix with diagonal entries 1 (and only nonzero off-diagonal entries are in the k^{th} column). Therefore, $\det(M_k) = 1$ for all k. This implies, finally, that $\det(B) = \det(AM_2M_3 \cdots M_n) = \det(A)\det(M_1) \cdots \det(M_n) = \det(A)$.