Math221 Homework # 6 Solutions

Instructor: James W. Demmel

October 10, 2009

Prob. 3.3

(1) We have $r = b - Ax, A^T r = 0$. Therefore $A^T r = A^T b - A^T Ax = 0$, i.e. x satisfies the normal equations of the least squares problem.

(2) Write the SVD of $A = U\Sigma V^T$ where U is m-by-m and orthogonal, Σ is m-by-n with $\Sigma_{ii} = \sigma_i$ and its other entries zero, and V is n-by-n and orthogonal. Then the $m+n$-by-$m+n$ symmetric matrix H may be written

$$H = \begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix} = \begin{bmatrix} U & 0 \\ 0 & V \end{bmatrix} \begin{bmatrix} I & \Sigma \\ \Sigma & 0 \end{bmatrix} \begin{bmatrix} U^T & 0 \\ 0 & V^T \end{bmatrix} \equiv \begin{bmatrix} U & 0 \\ 0 & V \end{bmatrix} \cdot S \cdot \begin{bmatrix} U^T & 0 \\ 0 & V^T \end{bmatrix}$$

The eigenvalues of H and S are the same. By permuting the rows and columns of S in the order 1, $m+1$, 2, $m+2$, ... n, $m+n$, $n+1$, $n+2$,..., m we get the similar matrix below, which is block diagonal:

$$PSP^T = \begin{bmatrix} S_1 & & \\ & \ddots & \\ & & S_m \end{bmatrix}$$

Here P is the permutation just described, blocks S_1 through S_n are 2-by-2 with

$$S_i = \begin{bmatrix} 1 & \sigma_i \\ \sigma_i & 0 \end{bmatrix}$$

and blocks S_{n+1} through S_m (if $m > n$) are 1-by-1 and equal to 1. The eigenvalues of this block diagonal matrix are just the union of the eigenvalues of the diagonal blocks, namely $\lambda_{\pm i} \equiv \frac{1 \pm \sqrt{1+4\sigma_i^2}}{2}$ for $i = 1, ..., n$ and $\lambda_{n+1} = \cdots = \lambda_m = 1$. It is easy to see that $\lambda_{+i} \geq 1$ and $\lambda_{-i} \leq 0$, so the largest eigenvalue in absolute value is $\lambda_{+1} = \frac{1+\sqrt{1+4\sigma_1^2}}{2}$, and smallest eigenvalue in absolute value is $\lambda_{-n} = \frac{1-\sqrt{1+4\sigma_n^2}}{2}$. So if $m = n$, the condition number of H is $\frac{\lambda_{+1}}{-\lambda_{-n}} = \frac{\sqrt{1+4\sigma_1^2}+1}{\sqrt{1+4\sigma_n^2}-1}$, and if $m > n$ the condition number of H is $\frac{\lambda_{+1}}{\min(-\lambda_{-n},1)} = \frac{\sqrt{1+4\sigma_1^2}+1}{\min(\sqrt{1+4\sigma_n^2}-1,2)}$.

(3) Doing block Gaussian Elimination

$$H = \begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix} = \begin{bmatrix} I & 0 \\ A^T & I \end{bmatrix} \cdot \begin{bmatrix} I & A \\ 0 & -A^TA \end{bmatrix}$$

and inverting yields

$$\begin{aligned} H^{-1} &= \begin{bmatrix} I & A \\ 0 & -A^TA \end{bmatrix}^{-1} \begin{bmatrix} I & 0 \\ A^T & I \end{bmatrix}^{-1} \\ &= \begin{bmatrix} I & A(A^TA)^{-1} \\ 0 & -(A^TA)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -A^T & I \end{bmatrix} \\ &= \begin{bmatrix} I - A(A^TA)^{-1}A^T & A(A^TA)^{-1} \\ (A^TA)^{-1}A^T & -(A^TA)^{-1} \end{bmatrix} \end{aligned}$$

The (2,1) entry is the Moore-Penrose pseudo-inverse of A.

Prob. 3.4

$\|Ax-b\|_C^2 = (Ax-b)^TC(Ax-b) = (Ax-b)^TL^TL(Ax-b)$ where $C = L^TL$ is the Cholesky factorization of the s.p.d. matrix C.

So $\|Ax-b\|_C^2 = \|L(Ax-b)\|_2 = \|LAx-Lb\|_2$, the normal equations for which are $(LA)^T(LA)x = (LA)^TLb$ which is equivalent to $A^TCAx = A^TCb$.

The formulation analogous to problem 3.3 is to solve $\begin{bmatrix} I & A \\ A^TC & 0 \end{bmatrix} \begin{bmatrix} r \\ x \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix}$. This results in $r = b - Ax$, $A^TCr = 0$, i.e. $A^TCAx = A^TCb$ - the normal equations as above. We can make this matrix symmetric by multiplying the first block row by C to get $\begin{bmatrix} C & CA \\ A^TC & 0 \end{bmatrix} \begin{bmatrix} r \\ x \end{bmatrix} = \begin{bmatrix} Cb \\ 0 \end{bmatrix}$.

Prob. 3.9

(1) $(A^TA)^{-1} = V\Sigma^{-2}V^T$

(2) $(A^TA)^{-1}A^T = V\Sigma^{-1}U^T$

(3) $A(A^TA)^{-1} = U\Sigma^{-1}V^T$

(4) $A(A^TA)^{-1}A^T = UU^T$ (note that $UU^T \neq I$ unless $m = n$).

Prob. 3.12

We can write the SVD of A in two ways: $A = U_A \Sigma_A V_A^T$ where U_A and V_A are both square and orthogonal and Σ_A has the same dimensions as A, and as

$$A = U_A \Sigma_A V_A^T = [U_{A1}, U_{A2}] \cdot \begin{bmatrix} \Sigma_{A1} & 0 \\ 0 & 0 \end{bmatrix} \cdot [V_{A1}, V_{A2}]^T = U_{A1} \Sigma_{A1} V_{A1}^T$$

where Σ_{A1} is square and nonsingular. Depending the dimensions of A and whether it is full rank or not, some blocks in the above factorization may be missing. These factorizations make sense for any dimensions of A. This lets us write the pseudo-inverse of A as $A^+ = V_{A1} \Sigma_{A1}^{-1} U_{A1}^T$, whether A has full rank or not. We use similar notation for the SVD of B. Now we can write

$$\begin{aligned}
\|AXB - C\|_F^2 &= \|U_A \Sigma_A V_A^T X U_B \Sigma_B V_B^T - C\|_F^2 \\
&= \|\Sigma_A (V_A^T X U_B) \Sigma_B - U_A^T C V_B\|_F^2 \\
&\equiv \|\Sigma_A (\hat{X}) \Sigma_B - \hat{C}\|_F^2 \\
&= \|\Sigma_A (\begin{bmatrix} \hat{X}_{11} & \hat{X}_{12} \\ \hat{X}_{21} & \hat{X}_{22} \end{bmatrix}) \Sigma_B - \begin{bmatrix} \hat{C}_{11} & \hat{C}_{12} \\ \hat{C}_{21} & \hat{C}_{22} \end{bmatrix} \|_F^2 \\
& \quad \text{where } \hat{X}_{ij} = V_{Ai}^T X U_{Bj} \text{ and } \hat{C}_{ij} = U_{Ai}^T X V_{Bj} \\
&= \| \begin{bmatrix} \Sigma_{A1} \hat{X}_{11} \Sigma_{B1} & 0 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} \hat{C}_{11} & \hat{C}_{12} \\ \hat{C}_{21} & \hat{C}_{22} \end{bmatrix} \|_F^2 \\
&= \|\Sigma_{A1} \hat{X}_{11} \Sigma_{B1} - \hat{C}_{11}\|_F^2 + \|\hat{C}_{12}\|_F^2 + \|\hat{C}_{21}\|_F^2 + \|\hat{C}_{22}\|_F^2
\end{aligned}$$

This is clearly minimized when the first term is zero, i.e. $\hat{X}_{11} = \Sigma_{A1}^{-1} \hat{C}_{11} \Sigma_{B1}^{-1}$. The other blocks of $\hat{X}$ ($\hat{X}_{12}$, $\hat{X}_{21}$ and $\hat{X}_{22}$) are arbitrary, but since

$$\|X\|_F^2 = \|\hat{X}\|_F^2 = \|\hat{X}_{11}\|_F^2 + \|\hat{X}_{12}\|_F^2 + \|\hat{X}_{21}\|_F^2 + \|\hat{X}_{22}\|_F^2$$

the norm of X is minimized by choosing $\hat{X}_{12} = 0$, $\hat{X}_{21} = 0$, and $\hat{X}_{22} = 0$. Call this minimal norm solution $\hat{X}_0$. Then

$$\begin{aligned}
X_0 &= V_A \hat{X}_0 U_B^T \\
&= V_A \begin{bmatrix} \Sigma_{A1}^{-1} \hat{C}_{11} \Sigma_{B1}^{-1} & 0 \\ 0 & 0 \end{bmatrix} U_B^T \\
&= V_{A1} \Sigma_{A1}^{-1} \hat{C}_{11} \Sigma_{B1}^{-1} U_{B1}^T \\
&= V_{A1} \Sigma_{A1}^{-1} U_{A1}^T C V_{B1} \Sigma_{B1}^{-1} U_{B1}^T \\
&= A^+ C B^+
\end{aligned}$$

has the same minimal norm as $\hat{X}_0$ as desired.

Prob. 3.13 When A is full rank, so that $A^+ = (A^TA)^{-1}A^T$, we can write

- $A \cdot A^+ \cdot A = A \cdot (A^TA)^{-1}A^T \cdot A = A$
- $A^+ \cdot A \cdot A^+ = (A^TA)^{-1}A^T \cdot A \cdot (A^TA)^{-1}A^T = (A^TA)^{-1} \cdot (A^TA) \cdot (A^TA)^{-1}A^T = (A^TA)^{-1}A^T = A^+$
- $A^+ \cdot A = (A^TA)^{-1}A^T\dot{A} = I = I^T$
- $(A \cdot A^+)^T = (A \cdot (A^TA)^{-1}A^T)^T = (A^T)^T(A^TA)^{-T}A^T = A(A^TA)^{-1}A^T = A \cdot A^+$.

When A is not necessarily full rank, we have to write $A = U\Sigma V^T$ and then $A^+ = V\Sigma^+U^T$ where $\Sigma^+ = \text{diag}(\sigma_1^+, ..., \sigma_n^+)$ and $\sigma_i^+ = \sigma_i^{-1}$ if $\sigma_i > 0$ and $0^+ = 0$. Plugging in we get

- $A \cdot A^+ \cdot A = U\Sigma V^T \cdot V\Sigma^+U^T \cdot U\Sigma V^T == U\Sigma\Sigma^+\Sigma V^T = U\Sigma V^T = A$, since $(\Sigma \cdot \Sigma^+ \cdot \Sigma)_{ii} = \sigma_i \cdot \sigma_i^+ \cdot \sigma_i = \sigma_i\sigma_i^{-1}\sigma_i = \sigma_i$ if $\sigma_i \neq 0$, and $0 \cdot 0 \cdot 0 = 0$ if $\sigma_i = 0$.
- $A^+ \cdot A \cdot A^+ = V\Sigma^+U^T \cdot U\Sigma V^T \cdot V\Sigma^+U^T = V\Sigma^+\Sigma\Sigma^+U^T = V\Sigma^+U^T = A^+$, since $\Sigma^+\Sigma\Sigma^+ = \Sigma^+$ for analogous reasons as before.
- $A^+ \cdot A = V\Sigma^+U^T \cdot U\Sigma V^T = V\Sigma^+\Sigma V^T = VDV^T$ where D is diagonal with $D_{ii} = \sigma_i^+\sigma_i = 1$ if $\sigma_i \neq 0$ and $D_{ii} = 0$ if $\sigma_i = 0$. This is clearly a symmetric matrix as desired.
- $A \cdot A^+ = U\Sigma V^T \cdot V\Sigma^+U^T = U\Sigma\Sigma^+U^T = UDU^T$, with D as before. This is also clearly symmetric as desired.

Prob. 3.15

Solution via QR: Let $A^T = QR$, Q n-by-m and orthogonal, R m-by-m and upper triangular. A is of full row rank so A^T is of full column rank and therefore R is nonsingular. Write $A^T = [Q\ ,\ Q'] \begin{bmatrix} R \\ 0 \end{bmatrix} \equiv \hat{Q} \begin{bmatrix} R \\ 0 \end{bmatrix}$ where $\hat{Q}$ is square and orthogonal, so $A = \begin{bmatrix} R^T & 0 \end{bmatrix} \hat{Q}^T$. Then $Ax = b$ is equivalent to $[R^T \quad 0]\hat{Q}^Tx = b$. Let $y = \hat{Q}^Tx = \begin{bmatrix} Q^Tx \\ Q'^Tx \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$. So we solve $Ax = R^Ty_1 = b$ and y_2 can be chosen arbitrarily. Therefore the solution $x = \hat{Q}y = Qy_1 + Q'y_2 = QR^{-T}b + Q'y_2$ can be anywhere in the $(n-m)$-dimensional set

$$P = \left\{QR^{-T}b + Q'y_2 \text{ such that } y_2 \in \mathbf{R}^{n-m}\right\}$$

Since $\|x\|_2^2 = \|y\|_2^2 = \|y_1\|_2^2 + \|y_2\|_2^2 = \|R^{-T}b\|_2^2 + \|y_2\|_2^2$, the minimum norm solution is obtained when $y_2 = 0$, namely $x = QR^{-T}b$.

Solution via SVD: Write $A^T = U\Sigma V^T = [U, U'] \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^T$, where $[U, U']$ is square and orthogonal. Then

$$\|Ax - b\|_2 = \|V\Sigma U^Tx - b\|_2 = \|\Sigma U^Tx - V^Tb\|_2$$

is 0 when $U^Tx = \Sigma^{-1}V^Tb$. Write $x = [U, U']y = [U, U'] \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = Uy_1 + U'y_2$. Then $U^Tx = y_1 = \Sigma^{-1}V^Tb$ and $\|x\|_2^2 = \|y\|_2^2 = \|y_1\|_2^2 + \|y_2\|_2^2 = \|\Sigma^{-1}V^Tb\|_2^2 + \|y_2\|_2^2$ is minimized by choosing $y_2 = 0$, i.e. $x = Uy_1 = U\Sigma^{-1}V^Tb$.

Solution via normal equations: Write $x = x_1 + x_2$ where x_1 is in the space spanned by the rows of A and x_2 is orthogonal to this space. Then $x_1 = A^Ty$ for some y, and

$Ax = A(x_1 + x_2) = Ax_1 = AA^T y = b$ exactly when $y = (AA^T)^{-1}b$ and $x_1 = A^T(AA^T)^{-1}b$. Furthermmore, $\|x\|_2^2 = \|x_1\|_2^2 + \|x_2\|_2^2$ is minimized when $x_2 = 0$, so $x = x_1 = A^T(AA^T)^{-1}b$ is the minimum norm solution.