Prof. Ming Gu, 861 Evans, tel: 2-3145
Office Hours: TuWTh 12:00-1:30PM
Email: mgu@math.berkeley.edu
http://www.math.berkeley.edu/~mgu/MA221

Math221: Matrix Computations
Selected Solutions to Homework #5

- **Problem 3.3:**
  - The system
    $$\begin{pmatrix} I & A \\ A^T & 0 \end{pmatrix} \begin{pmatrix} r \\ x \end{pmatrix} = \begin{pmatrix} b \\ 0 \end{pmatrix} \tag{1}$$
    is equivalent to
    $$r + Ax = b \quad , A^T r = 0,$$
    which is equivalent to $A^T (Ax - b) = 0$.
  - We let $A = U \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^T$ be the SVD of A, where $U \in \mathbf{R}^{m \times m}$, and Σ and $V \in \mathbf{R}^{n \times n}$. Then
    $$\begin{pmatrix} I & A \\ A^T & 0 \end{pmatrix} = \begin{pmatrix} U & \\ & V \end{pmatrix} \begin{pmatrix} I & 0 & \Sigma \\ 0 & I & 0 \\ \Sigma & 0 & 0 \end{pmatrix} \begin{pmatrix} U & \\ & V \end{pmatrix}^T . \tag{2}$$
    Hence 1 is a singular value of the coefficient matrix with multiplicty $m - n$. In addition, let σ_k be any singular value of A. Equation (2) suggests that each σ_k leads to two singular values of the coefficient matrix of (1), namely, the two singular values of the matrix $\begin{pmatrix} 1 & \sigma_k \\ \sigma_k & 0 \end{pmatrix}$, which are $\left(\sqrt{1 + \sigma_k^2} \pm 1 \right) /2$. The largest singular value is $\left(\sqrt{1 + \sigma_{\max}^2} + 1 \right) /2$, whereas the smallest singular value is $\min \left(1, \left(\sqrt{1 + \sigma_{\min}^2} - 1 \right) /2 \right)$. The condition number is the ratio
    $$\frac{\left(\sqrt{1 + \sigma_{\max}^2} + 1 \right) /2}{\min \left(1, \left(\sqrt{1 + \sigma_{\min}^2} - 1 \right) /2 \right)}.$$

– Since
$$\begin{pmatrix} I & A \\ A^T & 0 \end{pmatrix} = \begin{pmatrix} I & 0 \\ A^T & I \end{pmatrix} \begin{pmatrix} I & A \\ 0 & -A^TA \end{pmatrix}.$$
Hence
$$\begin{aligned} \begin{pmatrix} I & A \\ A^T & 0 \end{pmatrix}^{-1} &= \begin{pmatrix} I & A \\ 0 & -A^TA \end{pmatrix}^{-1} \begin{pmatrix} I & 0 \\ A^T & I \end{pmatrix}^{-1} = \begin{pmatrix} I & A\left(A^TA\right)^{-1} \\ 0 & -\left(A^TA\right)^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ -A^T & I \end{pmatrix} \\ &= \begin{pmatrix} I - A\left(A^TA\right)^{-1}A^T & A\left(A^TA\right)^{-1} \\ \left(A^TA\right)^{-1}A^T & -\left(A^TA\right)^{-1} \end{pmatrix}. \end{aligned}$$
The $(2,1)$ block is the solution operator for the least squares problem.

– As the least squares problem is equivalent to the linear system of equations (1), we can apply the iterative refinement techniques for linear equations to (1) in order to solve the least squares problem.

There are two key points: extra precision residual calculations and solution of (1) given QR factorization.

  * Extra precision residual calculations: Let $\begin{pmatrix} r \\ x \end{pmatrix}$ be the current solution to (1), we first compute the new residual as
$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} b \\ 0 \end{pmatrix} - \begin{pmatrix} I & A \\ A^T & 0 \end{pmatrix} \begin{pmatrix} r \\ x \end{pmatrix} = \begin{pmatrix} b - r - Ax \\ -A^Tr \end{pmatrix}.$$
For extra accuracy, we need to do the whole computation in doubled precision.

  * Solving equation (1) with right hand side $\begin{pmatrix} u \\ v \end{pmatrix}$
$$\begin{pmatrix} I & A \\ A^T & 0 \end{pmatrix} d = \begin{pmatrix} u \\ v \end{pmatrix}.$$
The key observation is that we can develop an orthogonal factorization of $\begin{pmatrix} I & A \\ A^T & 0 \end{pmatrix}$ given the full QR factorization of $A = Q\begin{pmatrix} R \\ 0 \end{pmatrix}$, where $Q \in \mathbf{R}^{m\times m}$ is orthogonal and $R \in \mathbf{R}^{n\times n}$ is upper triangular:
$$\begin{pmatrix} I & A \\ A^T & 0 \end{pmatrix} = \begin{pmatrix} Q & \\ & I_n \end{pmatrix} \begin{pmatrix} 0 & 0 & I_n \\ 0 & I_{m-n} & 0 \\ I_n & 0 & 0 \end{pmatrix} \begin{pmatrix} R^T & 0 & 0 \\ 0 & I_{m-n} & 0 \\ I_n & 0 & R \end{pmatrix} \begin{pmatrix} Q^T & \\ & I_n \end{pmatrix}.$$
On the right hand side, the first and the last matrices are orthogonal matrices, the inverses of which are obvious; and the second matrix is a permutation. Hence inversions with these matrices are simple. The third matrix is a 3×3 block lower triangular matrix, with each of its diagonal blocks being either identity or triangular. Hence the inversion involving this matrix is also trivial.

As in linear systems iterative refinement, the new iteration is $\begin{pmatrix} r \\ x \end{pmatrix} - d$.
For more details, see the Tech Report by Demmel, Hida, Li, and Riedy at
`http://techreports.lib.berkeley.edu/accessPages/EECS-2007-77.html`

- **Problem 3.4:** To derive the equation, we let $C = LL^T$ be the Cholesky factorization of C (it need not be computed.) Then

$$\|Ax - b\|_C = \|L^T (Ax - b)\|_2.$$

This is a regular LS problem with coefficient matrix $L^T A$ and right hand side $L^T b$. Hence the normal equation is

$$\left(L^T A\right)^T \left(L^T A x - L^T b\right) = 0,$$

or

$$A^T C (Ax - b) = 0.$$

This can be rewritten as

$$\begin{pmatrix} I & L^T A \\ A^T L & 0 \end{pmatrix} \begin{pmatrix} r \\ x \end{pmatrix} = \begin{pmatrix} L^T b \\ 0 \end{pmatrix},$$

or

$$\begin{pmatrix} C^{-1} & A \\ A^T & 0 \end{pmatrix} \begin{pmatrix} r \\ x \end{pmatrix} = \begin{pmatrix} b \\ 0 \end{pmatrix}.$$

- Let $\mathcal{S}$ be a subspace spanned by a set of basis vectors $v_1, \cdots, v_k$. Let $A = LL^T$ be the Cholesky factorization of A. Now let $QR = L^T (v_1, \cdots, v_k)$ be the QR factorization and $\widehat{Q} = L^{-T} Q$. It follows that $\widehat{Q} R = (v_1, \cdots, v_k)$. This means columns of $\widehat{Q}$ are a basis for $\mathcal{S}$ as well. But $\widehat{Q}^T A \widehat{Q} = Q^T Q = I$. Hence $\widehat{Q}$ is an A-orthogonal basis.