

STAT 309: MATHEMATICAL COMPUTATIONS I
FALL 2014
PROBLEM SET 4

1. In this exercise, we will implement and compare Gram–Schmidt and Householder QR. Your implementation should be tailored to the program you are using for efficiency (e.g. vectorize your code in Matlab/Octave/Scilab). Assume in the following that the input is a matrix $A \in \mathbb{R}^{m \times n}$ with $\text{rank}(A) = n \leq m$ and we want to find its full QR decomposition $A = QR$ where $Q \in O(m)$ and $R \in \mathbb{R}^{m \times n}$ is upper-triangular.
 - (a) Implement the (classical) Gram–Schmidt algorithm to obtain Q and R .
 - (b) Implement the Householder QR algorithm to obtain Q and R . You should (i) store Q implicitly, taking advantage of the fact that it can be uniquely specified by a sequence of vectors of decreasing dimensions; (ii) choose α in your Householder matrices to have the opposite sign of x_1 to avoid cancellation in v_1 (cf. notations in lecture notes).
 - (c) Implement an algorithm for forming the product $Q\mathbf{x}$ and another for forming the product $Q^T\mathbf{y}$ when Q is stored implicitly as in (b).
 - (d) For increasing values of n , generate an upper triangular $R \in \mathbb{R}^{n \times n}$ and a $B \in \mathbb{R}^{n \times n}$, both with random standard normal entries. Use your program’s built-in function for QR factorization to obtain a random¹ $Q \in O(n)$ from the QR factorization of B . Now form $A = QR$ and apply your algorithms in (a) and (b) to find the QR factors of A — let these be \hat{Q} and \hat{R} . Tabulate (using graphs with appropriate scales) the relative errors

$$\frac{\|R - \hat{R}\|_F}{\|R\|_F}, \quad \|Q - \hat{Q}\|_F, \quad \frac{\|A - \hat{Q}\hat{R}\|_F}{\|A\|_F},$$
 for various values of n and for each method. Scale Q, R, \hat{Q}, \hat{R} appropriately so that R and \hat{R} have positive diagonal elements.
 - (i) Comment on the relative errors in \hat{Q} and \hat{R} (these are called forward errors) versus the relative error in $\hat{Q}\hat{R}$ (this is called backward error).
 - (ii) Comment on the relative error in $\hat{Q}\hat{R}$ computed with Gram–Schmidt versus that computed with Householder QR.
 - (e) Generate a *Vandermonde matrix* and a vector,

$$A = \begin{bmatrix} 1 & \alpha_0 & \alpha_0^2 & \cdots & \alpha_0^{n-1} \\ 1 & \alpha_1 & \alpha_1^2 & \cdots & \alpha_1^{n-1} \\ 1 & \alpha_2 & \alpha_2^2 & \cdots & \alpha_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_{m-1} & \alpha_{m-1}^2 & \cdots & \alpha_{m-1}^{n-1} \end{bmatrix} \in \mathbb{R}^{m \times n}, \quad \mathbf{b} = \begin{bmatrix} \exp(\sin 4\alpha_0) \\ \exp(\sin 4\alpha_1) \\ \exp(\sin 4\alpha_2) \\ \vdots \\ \exp(\sin 4\alpha_{m-1}) \end{bmatrix} \in \mathbb{R}^m,$$

where $\alpha_i = i/(m-1)$, $i = 0, 1, \dots, m-1$. This arises when we try to do polynomial fitting

$$e^{\sin 4x} \approx c_0 + c_1x + c_2x^2 + \cdots + c_{n-1}x^{n-1}$$

Date: November 1, 2014 (Version 1.0); due: November 14, 2014.

¹This is usually how one would generate a random orthogonal matrix.

over the interval $[0, 1]$ at discrete points $x = 0, \frac{1}{m-1}, \frac{2}{m-1}, \dots, \frac{m-2}{m-1}, 1$. For $n = 15$ and $m = 100$, solve the least squares problem $\min \|A\mathbf{x} - \mathbf{b}\|_2$ and state your value of c_{14} using each of the following methods:

- (i) Applying QR factorization to A .
- (ii) Applying QR factorization to the augmented matrix $[A, \mathbf{b}] \in \mathbb{R}^{m \times (n+1)}$.
- (iii) Solving the normal equations $A^T A \mathbf{x} = A^T \mathbf{b}$.

For (i) and (ii), your code should show how the respective QR factors are used in obtaining a solution of the least squares problem. You are free to use your program's built-in functions (e.g. `A\b` in Matlab/Octave/Scilab) for solving linear systems but for other things, use what you have implemented in (a), (b), (c). The true value of c_{14} is 2006.787453080206... Comment on the accuracy of each method and algorithm.

- 2.** Let $A \in \mathbb{R}^{m \times n}$ where $m \geq n$ and $\text{rank}(A) = n$. Suppose GECP is performed on A to get

$$\Pi_1 A \Pi_2 = LU$$

where $L \in \mathbb{R}^{m \times n}$ is unit lower triangular, $U \in \mathbb{R}^{n \times n}$ is upper triangular, and $\Pi_1 \in \mathbb{R}^{m \times m}$, $\Pi_2 \in \mathbb{R}^{n \times n}$ are permutation matrices.

- (a) Show that U is nonsingular and that L is of the form

$$L = \begin{bmatrix} L_1 \\ L_2 \end{bmatrix}$$

where $L_1 \in \mathbb{R}^{n \times n}$ is nonsingular.

- (b) We will see how the LU factorization may be used to solve the least squares problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|A\mathbf{x} - \mathbf{b}\|_2.$$

- (i) Show that the problem may be solved via

$$U\tilde{\mathbf{x}} = \mathbf{y}, \quad L^T L \mathbf{y} = L^T \tilde{\mathbf{b}},$$

where $\tilde{\mathbf{b}} = \Pi_1 \mathbf{b}$ and $\tilde{\mathbf{x}} = \Pi_2^T \mathbf{x}$.

- (ii) Describe how you would compute the solution \mathbf{y} in

$$L^T L \mathbf{y} = L^T \tilde{\mathbf{b}}.$$

- 3.** Let $\varepsilon > 0$. Consider the matrix

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 + \varepsilon \\ 1 & 1 - \varepsilon \end{bmatrix}.$$

- (a) Why is it a bad idea to solve the normal equation associated with A , i.e.

$$A^T A \mathbf{x} = A^T \mathbf{b}$$

when ε is small?

- (b) Show that the LU factorization of A is

$$A = LU = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & \varepsilon \end{bmatrix}.$$

- (c) Why is it a much better idea to solve the normal equation associated with L , i.e.

$$L^T L \mathbf{y} = L^T \tilde{\mathbf{b}}?$$

This shows that the method in Problem 2 is a more stable method than using the normal equation in (a) directly.

(d) Show that the Moore–Penrose pseudoinverse of A is

$$A^\dagger = \frac{1}{6} \begin{bmatrix} 2 & 2 - 3\varepsilon^{-1} & 2 + 3\varepsilon^{-1} \\ 0 & 3\varepsilon^{-1} & -3\varepsilon^{-1} \end{bmatrix}.$$

(e) Describe a method to compute A^\dagger given L and U . Verify that your method is correct by checking it against the expression in (d).

4. We will now discuss an alternative method to solve the least squares problem in Problem 2 that is more efficient when $m - n < n$.

(a) Show that the least squares problem in Problem 2 is equivalent to

$$\min_{\mathbf{z} \in \mathbb{R}^n} \left\| \begin{bmatrix} I_n \\ S \end{bmatrix} \mathbf{z} - \tilde{\mathbf{b}} \right\|_2$$

where $S = L_2 L_1^{-1}$ and $L_1 \mathbf{y} = \mathbf{z}$. Here and below, I_n denotes the $n \times n$ identity matrix.

(b) Write

$$\tilde{\mathbf{b}} = \begin{bmatrix} \tilde{\mathbf{b}}_1 \\ \tilde{\mathbf{b}}_2 \end{bmatrix}$$

where $\tilde{\mathbf{b}}_1 \in \mathbb{R}^n$ and $\tilde{\mathbf{b}}_2 \in \mathbb{R}^{m-n}$. Show that the solution \mathbf{z} is given by

$$\mathbf{z} = \tilde{\mathbf{b}}_1 + S^\top (I_{m-n} + SS^\top)^{-1} (\tilde{\mathbf{b}}_2 - S\tilde{\mathbf{b}}_1).$$

(c) Explain why when $m - n < n$, the method in (a) is much more efficient than the method in Problem 2. For example, what happens when $m = n + 1$?

5. Let $\mathbf{c} \in \mathbb{R}^n$ and consider the linearly constrained least squares problem

$$\min \|\mathbf{w}\|_2 \quad \text{s.t.} \quad A^\top \mathbf{w} = \mathbf{c}.$$

(a) If we write $\tilde{\mathbf{c}} = \Pi_2^\top \mathbf{c}$ and $\tilde{\mathbf{w}} = \Pi_1 \mathbf{w}$, show that

$$\tilde{\mathbf{w}} = L(L^\top L)^{-1} U^{-\top} \tilde{\mathbf{c}}$$

where $U^{-\top} = (U^{-1})^\top = (U^\top)^{-1}$, a standard notation that we will also use below. (*Hint:* You'd need to use something that you've already determined in an earlier part).

(b) Write

$$\tilde{\mathbf{w}} = \begin{bmatrix} \tilde{\mathbf{w}}_1 \\ \tilde{\mathbf{w}}_2 \end{bmatrix}$$

where $\tilde{\mathbf{w}}_1 \in \mathbb{R}^n$ and $\tilde{\mathbf{w}}_2 \in \mathbb{R}^{m-n}$. Show that

$$\tilde{\mathbf{w}}_1 = L_1^{-\top} U^{-\top} \tilde{\mathbf{c}} - S^\top \tilde{\mathbf{w}}_2.$$

(c) Write $\mathbf{d} = L_1^{-\top} U^{-\top} \tilde{\mathbf{c}}$. Deduce that $\tilde{\mathbf{w}}_2$ may be obtained either as a solution to

$$\min_{\tilde{\mathbf{w}}_2 \in \mathbb{R}^{m-n}} \left\| \begin{bmatrix} S^\top \\ I_{m-n} \end{bmatrix} \tilde{\mathbf{w}}_2 - \begin{bmatrix} \mathbf{d} \\ \mathbf{0} \end{bmatrix} \right\|_2$$

or as

$$\tilde{\mathbf{w}}_2 = (I_{m-n} + SS^\top)^{-1} S\mathbf{d}.$$

Note that when $m - n < n$, this method is advantageous for the same reason in Problem 4.

6. So far we have assumed that A has full column rank. Suppose now that $\text{rank}(A) = r < \min\{m, n\}$.

- (a) Show that the LU factorization obtained using GECP is of the form

$$\Pi_1 A \Pi_2 = LU = \begin{bmatrix} L_1 \\ L_2 \end{bmatrix} \begin{bmatrix} U_1 & U_2 \end{bmatrix}$$

where $L_1, U_1 \in \mathbb{R}^{r \times r}$ are triangular and nonsingular.

- (b) Show that the above equation may be rewritten in the form

$$\Pi_1 A \Pi_2 = \begin{bmatrix} I_r \\ S_1 \end{bmatrix} L_1 U_1 \begin{bmatrix} I_r & S_2^T \end{bmatrix}$$

for some matrices S_1 and S_2 .

- (c) Hence show that the Moore–Penrose inverse of A is given by

$$A^\dagger = \Pi_2 \begin{bmatrix} I_r & S_2^T \end{bmatrix}^\dagger U_1^{-1} L_1^{-1} \begin{bmatrix} I_r \\ S_1 \end{bmatrix}^\dagger \Pi_1.$$

- (d) Using the general formula (derived in the lectures) for the Moore–Penrose inverse of a rank-retaining factorization, what do you get for A^\dagger ?

7. Consider the block matrix

$$X = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \in \mathbb{R}^{n \times n}$$

where $A \in \mathbb{R}^{p \times p}$, $B \in \mathbb{R}^{p \times q}$, $C \in \mathbb{R}^{q \times p}$, $D \in \mathbb{R}^{q \times q}$ and $n = p + q$. The Schur complements of A and D are

$$S = D - CA^\dagger B \quad \text{and} \quad T = A - BD^\dagger C$$

respectively.

- (a) Verify that if A and S are nonsingular, then

$$X^{-1} = \begin{bmatrix} A^{-1} + A^{-1}BS^{-1}CA^{-1} & -A^{-1}BS^{-1} \\ -S^{-1}CA^{-1} & S^{-1} \end{bmatrix}$$

and if D and T are nonsingular, then

$$X^{-1} = \begin{bmatrix} T^{-1} & -T^{-1}BD^{-1} \\ -D^{-1}CT^{-1} & D^{-1} + D^{-1}CT^{-1}BD^{-1} \end{bmatrix}.$$

- (b) Show that

$$\det X = \begin{cases} \det(A) \det(D - CA^{-1}B) & \text{if } A \text{ nonsingular,} \\ \det(D) \det(A - BD^{-1}C) & \text{if } D \text{ nonsingular.} \end{cases}$$

Deduce that

$$\det(A + BC) = \det(A) \det(I + CA^{-1}B)$$

and use it to find the determinants of the following matrices

$$\begin{bmatrix} \frac{1+\lambda_1}{\lambda_1} & 1 & \cdots & 1 \\ 1 & \frac{1+\lambda_2}{\lambda_2} & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & \frac{1+\lambda_n}{\lambda_n} \end{bmatrix}, \quad \begin{bmatrix} 1+\lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \lambda_1 & 1+\lambda_2 & \cdots & \lambda_n \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1 & \lambda_2 & \cdots & 1+\lambda_n \end{bmatrix}, \quad \begin{bmatrix} \lambda & \mu & \mu & \cdots & \mu \\ \mu & \lambda & \mu & \cdots & \mu \\ \mu & \mu & \lambda & \cdots & \mu \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu & \mu & \mu & \cdots & \lambda \end{bmatrix}.$$

- (c) Show that if A has all principal matrices nonsingular so that we may perform Gaussian elimination without pivoting to A , then applying the first p steps of that to X yields

$$X = \begin{bmatrix} L_{11} & 0 \\ L_{21} & I_q \end{bmatrix} \begin{bmatrix} I_p & 0 \\ 0 & S \end{bmatrix} \begin{bmatrix} U_{11} & U_{12} \\ 0 & I_q \end{bmatrix}$$

where $A = L_{11}U_{11}$ is the LU factorization of A . What are L_{21} and U_{12} in terms of L_{11}, U_{11} and the blocks of X ?

- (d) Suppose X is symmetric (so $C = B^\top$) and A is positive definite. Show that applying the first p steps of Cholesky factorization to X yields

$$X = \begin{bmatrix} R_{11}^\top \\ R_{12}^\top \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & S \end{bmatrix}$$

where $A = R_{11}^\top R_{11}$ is the Cholesky factorization. What is R_{12} in terms of R_{11} and the blocks of X ?

- (e) Suppose the coefficient matrix of a nonsingular system $A\mathbf{x} = \mathbf{b}$ is updated to produce another nonsingular system $(A + \mathbf{u}\mathbf{v}^\top)\mathbf{z} = \mathbf{b}$. Show that

$$\mathbf{z} = \mathbf{x} - \frac{\mathbf{v}^\top \mathbf{x}}{1 + \mathbf{v}^\top \mathbf{y}} \mathbf{y}$$

where $A\mathbf{y} = \mathbf{u}$.