1. Let $A \in \mathbb{R}^{m \times n}$ where $m \geq n$ and 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

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

(i) Show that the problem may be solved via

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

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

(ii) Describe how you would compute the solution $y$ in

$$L^T Ly = L^T \tilde{b}.$$

2. 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.

$$A^T Ax = A^T b$$

when $\varepsilon$ 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.

$$L^T Ly = L^T \tilde{b}?$$

This shows that the method in Problem 1 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).

3. We will now discuss an alternative method to solve the least squares problem in Problem 1 that is more efficient when $m - n < n$.

(a) Show that the least squares problem in Problem 1 is equivalent to

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

where $S = L_2L_1^{-1}$ and $L_1y = z$. Here and below, $I_n$ denotes the $n \times n$ identity matrix.

(b) Write

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

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

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

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

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

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

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

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

where $U^{-T} = (U^{-1})^T = (U^T)^{-1}$, 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

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

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

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

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

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

or as

$$\tilde{w}_2 = (I_{m-n} + SS^T)^{-1} Sd.$$ 

Note that when $m - n < n$, this method is advantageous for the same reason in Problem 3.

5. 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

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

for some matrices $S_1$ and $S_2$.

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

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

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

6. 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

$$S = D - CA^{-1}B \quad \text{and} \quad T = A - BD^{-1}C$$ 

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

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

(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 $C = B^T$) and $A$ is positive definite. Show that applying the first $p$ steps of Cholesky factorization to $X$ yields

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

where $A = R_{11}^T R_{11}$ 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 $A x = b$ is updated to produce another nonsingular system $(A + uv^T) z = b$. Show that

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

where $A y = u$. 