1. Implement a matlab function which computes the Householder QR factorization of a full-rank matrix $A$. [You need to implement the “alternative” version which yields positive entries on the diagonal of the $R$ matrix.] The main function should be as follows:

$$[V, \text{bet}] = \text{housQR}(A) .$$

Here $V$ should contain the vectors $v_1, \ldots, v_n$ related to the successive Householder reflectors $P_j = I - \beta_j v_j v_j^T$ used to transform $A$ into upper triangular form and $\text{bet}$ is the vector of the coefficients $\beta_j$. The same array $V$ also contains $R$ matrix in its upper part. The basic scripts `house.m` and `house1.m` are provided in the class web-site (under matlab). The script `housQR` (a version of which was shown in class) does not generate the $Q$ matrix. Along with it you need to have another script which generates the $Q$ matrix from $V$ and the array $\text{bet}$. Show the two matlab scripts. Apply these scripts to compute the QR factorization of the same matrix as the one you used for Question 6 of HW4. You can now compare the result of the Householder QR obtained in this way with the three you already compared in HW4. Show the same error measures $\|A - Q^* R\|_2$, and $\|I - Q^T \ast Q\|_2$. Comment on what you observe relative to the results you had in HW4.

2. For this exercise, you can do all calculations by hand, and use matlab to verify or to help. Consider the matrix:

$$A = \begin{pmatrix}
1 & 0 & 1 \\
0 & -1 & -1 \\
1 & 0 & 1 \\
0 & 1 & 1
\end{pmatrix}$$

(a) What is the rank of $A$?

(b) Find orthonormal bases for the ranges of $A$ and of $A^T$.

(c) Using what you know about the URV decomposition and the answer to (b), show how you can get the nonzero singular values of $A$. Compute the singular values using this approach.

3. Consider the problem $\min \|b - Ax\|_2$ in the situation where $A$ is $m \times n$ and $m < n$ (called the ‘underdetermined’ case). Assume that $A$ is of full rank. Using what you learned from the URV decomposition, show that the set of solutions is of dimension $n - m$. Show that the least-squares solution $x_*$ of smallest norm must belong to $\text{Ran}(A^T)$. Find a method for computing $x_*$ which involves a form of normal equations. Find a method for computing $x_*$ which involves the QR factorization.

4. Consider the matrix

$$A = \begin{pmatrix}
2 & -1 & 2 & 1 \\
2 & 1 & 2 & -1
\end{pmatrix}$$

(a) Compute $AA^T$ and obtain the singular values and a set of left singular vectors of $A$.

(b) Find the “thin” SVD of $A$: $A = U \Sigma_1 V_1^T$, $U \in \mathbb{R}^{2 \times 2}$ unitary, $\Sigma_1 \in \mathbb{R}^{2 \times 2}$,...
(c) Find the matrix $B$ which is of rank 1 and which is the closest to the matrix $A$ in the 2-norm sense. What is the 2-norm distance between $A$ and $B$?

(d) Complete the columns of $V_1$ of question (b) into an orthonormal basis of $\mathbb{R}^4$ and find the (full) SVD of $A$.

(e) What is the null space of $A$? Find the vector of this null space that is the closest in the 2-norm sense to the vector $z = [0 \ 1 \ 1 \ 0]^T$.

(f) What are all the least-squares solutions $Ax = b$ where $b = [8 \ 8]^T$?

(g) Among the solutions found in (f) which one is the closest to the vector $u = [2 \ 1 \ -1 \ 0]^T$ in the 2-norm sense?

5. This problem is about information retrieval. You are given a matrix of term-documents which lists, as columns, for 15 given documents (book-titles - labeled D1, D2, · · · , D15), the most common terms occuring in the document, with a measure of how often the 10 terms occur on the books. The matrix, called $C$, is available as TDmat.mat in the class web-site. The terms correspond to the following index entries:


For example, you will see that Document D2 has the terms “Determinant” with a relative frequency of $\approx 0.56$, “linear system” with a relative frequency of $\approx 0.14$, and “eigenvalue” with a relative frequency of $\approx 0.45$. We would like to find the matching document for the query: \{ rank, matrix, QR \}.

a. Find the best 3 matches based on comparing angles between the columns and the query (the higher the cosine the better).

b. Compute the SVD of the term-document matrix called $C$.

c. Find the matrix $C_4$ which is the best rank 4 approximation to the matrix $C$, in the 2-morm sense.

d. Repeat the query in question (a) but use $C_4$ instead of $C$. 