Note: 1) The due date is firm because we need to grade this HW asap to determine who is exempted from the final. 2) For the same reason not all questions will be graded – however a full solution will be posted in preparation for the final.

1. Find the least-squares solution to the system \( \min \| b - Ax \| \) where

\[
A = \begin{pmatrix}
2 & 1 \\
1 & 0 \\
0 & -1 \\
-1 & 1
\end{pmatrix}, \quad b = \begin{pmatrix}
3 \\
1 \\
2 \\
-1
\end{pmatrix}
\]

Find the QR factorization of the matrix \( A \). See the lecture notes (pages 14-14 and pages 14-15) and the text (theorem 15 of text) to find out how to solve the least-squares system using the QR factorization you just found. Find the solution to the least squares problem using this approach. [Note: as it turns out, because of rounding errors, this approach is numerically more reliable than the one based on normal equations.]

2. The yearly temperature cycle in Fairbanks, Alaska is given in the next table.

<table>
<thead>
<tr>
<th>Date</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
<th>g</th>
<th>h</th>
<th>i</th>
<th>j</th>
<th>k</th>
<th>l</th>
<th>m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degrees</td>
<td>-14</td>
<td>-9</td>
<td>2</td>
<td>15</td>
<td>35</td>
<td>52</td>
<td>62</td>
<td>63</td>
<td>58</td>
<td>50</td>
<td>34</td>
<td>12</td>
<td>-5</td>
</tr>
</tbody>
</table>

There are 13 equally spaced data points which correspond to a measurement every 28 days ((a) corresponds to Jan. 1, (b) to Jan 29, ..., and (m) to Dec. 3).

(a) Find the trigonometric polynomial \( T_3(t) \) of the form

\[
a_0 + a_1 \cos\left(\frac{2\pi t}{13}\right) + b_1 \sin\left(\frac{2\pi t}{13}\right) + a_2 \cos\left(\frac{4\pi t}{13}\right) + b_2 \sin\left(\frac{4\pi t}{13}\right) + a_3 \cos\left(\frac{6\pi t}{13}\right) + b_3 \sin\left(\frac{6\pi t}{13}\right)
\]

which approximates the above data in the least-squares sense. Any important observations about the matrix \( A^T A \) of the normal equations in this case?

(b) Graph \( T_3(t) \) along with the 13 data points on the same coordinate system. Use a total at least 200 equally spaced points to plot \( T_3 \).

3. Consider the matrix

\[
A = \begin{pmatrix}
1 & 0 & 2 \\
0 & -1 & -2 \\
2 & -2 & 0
\end{pmatrix}
\]

(a) Find the eigenvalues and associated eigenvectors of \( A \);

(b) Is \( A \) similar to a diagonal matrix? If so find a matrix \( M \) such that \( M^{-1}AM \) is diagonal. Is \( M \) unique? Explain.
(c) Find the eigenvalues and associated eigenvectors of $A^2$.

(d) Show that of $A + I$ is invertible and find the eigenvalues and associated eigenvectors of $(A + I)^{-1}$.

(e) Let $B = A/3$. Determine $9B^{213}$ and $9B^{214}$ (computing the results with matlab in a brute force manner will be discarded as incorrect)

4. The eigenspace associated with an eigenvalue $\lambda$ of a matrix $A$ is the null space of $(A - \lambda I)$ [see text, Sec. 5.1 for details]. What are the eigenvalues and eigenspaces of the matrix:

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{bmatrix}$$

Is $A$ diagonalizable?

5. The following table shows the U.S. per capita healthcare expenditure for the last 50 years starting in 1960 [Source: http://www/cdc.gov/..]

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Expenditure</td>
<td>125</td>
<td>300</td>
<td>943</td>
<td>2431</td>
<td>4129</td>
<td>7101</td>
</tr>
</tbody>
</table>

(a) Find a simple least-squares linear fit to the data.

(b) Find a quadratic fit to the model, i.e., a function of the form $dp(t) = \alpha_0 + \alpha_1(t - t_0) + \alpha_2(t - t_0)^2$ (take $t_0 = 1960$).

(c) Now assume that the expenditures evolve exponentially, i.e., that they are of the form $d(t) = \beta e^{\alpha(t-t_0)}$ (take again $t_0 = 1960$). Find the scalars $\alpha$ and $\beta$ of this model by doing a linear least-squares fit with the logs of the expenditures.

(d) Which of your 3 models yields the best match to the original data? Plot the three curves for the predictions (along the original data) on the same figure.

(e) Using your most accurate model, predict the US per capita healthcare expenditure for the years 2020 and 2030.

6. Let $A = \frac{1}{2} \begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix}$. Compute $A^{16}$ by successive squaring (i.e., set $B = A^2$, then apply the recurrence $B := B^2$ a few times to get $A^{16}$). Then compute $A^{16}$ by using the eigenvalue decomposition of $A$. (hint: exploit the fact that $(1/2)^{16}$ is very small).

7. 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, \cdots, D15), the most common terms occurring in the document, with a measure of how often the 10 terms occur on the books. The matrix, called $C$, is available as $\text{TDmat.mat}$ 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 Document which is the best match for the query: \{ echelon, linear system, rank, matrix \}.

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 \( C \). The command is \([U, S, V] = \text{svd}(C)\). What are the singular values of \( C \)?

c. The SVD gives you the relation \( C = USV^T \), where \( S \) is diagonal. Replace all diagonal entries of \( S \) except the first 4 by zeros. Call \( S_1 \) the resulting matrix. You now get a new matrix \( C_1 = US_1V^T \). What is the rank of \( C_1 \)? Repeat the query in a), in which \( C \) is replaced by \( C_1 \).

8. You will find in the matlab section of the course website a matrix called \( X \) and stored in a file called \( \text{X.mat} \) (load with the command \( \text{load('X')\} \)). Using matlab compute the Singular Value Decomposition of the matrix \( X \). What is the approximate rank \( r \) of the matrix?

**Bonus points: (3pts)** Find the subspace of dimension \( r \) that represents the column space associated with this rank.