STAT 30900: Homework 2

Caleb Derrickson

October 31, 2023

Collaborators: The TA's of the class, as well as Kevin Hefner, and Alexander Cram.

Contents

Problem 1

The files required for this problem can be found in the subfolder hw2 in Canvas. The matrix in processed.mat (Matlab format) or processed.txt (comma separated, plain text) is a 49 × 7 matrix where each row is indexed by a country in row.txt and each column is indexed by a demographic variable in column.txt, ordered as in the respective files. So for example, if we denote the matrix by

$$A = \begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \vdots \\ \mathbf{a}_{49}^T \end{bmatrix} = \begin{bmatrix} \alpha_1, & \ldots, & \alpha_7 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{17} \\ a_{21} & a_{22} & \ldots & a_{27} \\ \vdots & \vdots & \ddots & \vdots \\ a_{49,1} & a_{49,2} & \ldots & a_{49,7} \end{bmatrix} \in \mathbb{R}^{49\times 7},$$

then a_{23} = -0.2743 is Austria's population per square kilometers (row index 2 = Austria, column index 3 = population per square kilometers). As you probably notice, this matrix has been slightly preprocessed. If you want to see the raw data, you can find them in raw.txt (e.g. the actual value for Austria's population per square kilometers is 84) but you don't need the raw data for this problem.

Problem 1, part a

Show that to plot the projections of the row vectors (i.e., samples) $\mathbf{a}_1, ..., \mathbf{a}_{49} \in \mathbb{R}^7$ onto the two-dimensional subspace span $\{\mathbf{v}_j, \mathbf{v}_k \cong \mathbb{R}^2\}$, we may simply plot the n points

$$\{(\sigma_j u_{ij}, \sigma_k u_{ik}) \in \mathbb{R}^2 : i = 1, ..., 49\}$$

where $U = [u_{ij}] \in \mathbb{R}^{49\times 49}$ and $\Sigma = \text{diag}(\sigma_1, ..., \sigma_p) \in \mathbb{R}^{49\times 7}$ are the matrix of left singular vectors and matrix of singular values respectively.

Solution:

Denote the Singular Value Decomposition of A as $U\Sigma V^t$, where U, V are matrices constructed by orthonormal vectors. They are related to each other by $A\mathbf{v}_k = \sigma_k \mathbf{u}_k$, where $\mathbf{u}_k, \mathbf{v}_k$ is the $k-$th column vector of U and V, respectively, and σ_k is the k-th singular value of A. Similarly, $A^*\mathbf{u} = \sigma_k \mathbf{v}_k$. Thus, if there were some other $\mathbf{u}_i$, such that $A^*\mathbf{u}_i = \sigma_i \mathbf{v}_k$, then

$$\mathbf{v}_k^* \mathbf{v}_k = (\frac{1}{\sigma_i} A^* \mathbf{u}_i)^* A^* \mathbf{u}_k = \frac{1}{\sigma_i} \mathbf{u}_i^* A A^* \mathbf{u}_k = \frac{\sigma_k^2}{\sigma_i} \mathbf{u}_i^* \mathbf{u}_k = 0$$

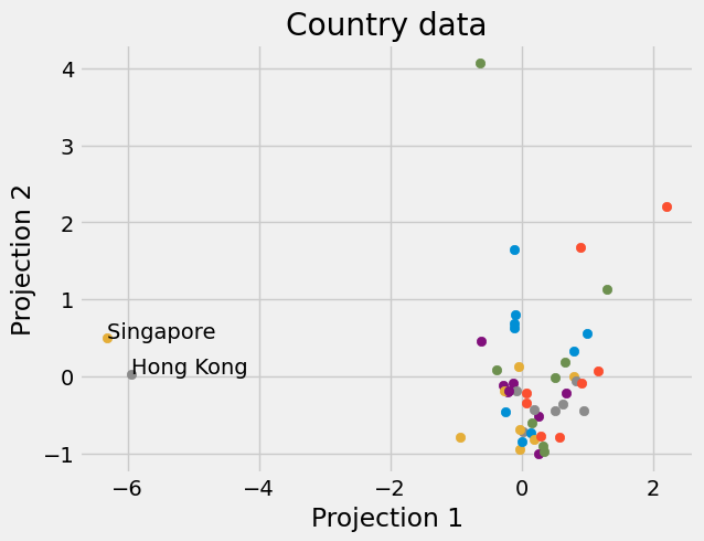
since the set of $\mathbf{u}$ vectors are orthonormal Thus, no other $\mathbf{u}$ contributes to $\mathbf{v}_k$, meaning if we want to project onto the given subspace, we only need to plot $\{(\sigma_j u_{ij}, \sigma_k u_{ik}) \in \mathbb{R}^2 : i = 1, ..., 49\}$.

Problem 1, part b

Find the first two right singular vectors of $A, \mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^7$. Project the data onto the two-dimensional space $\text{span}\{\mathbf{v}_1, \mathbf{v}_2\} \cong \mathbb{R}^2$. Plot this in a graph where the x- and y-axes correspond to $\mathbf{v}_1$ and $\mathbf{v}_2$ respectively and where the points correspond to the countries — label each point by the country it corresponds to. Identify the two obvious outliers.

Solution:

I will provide the plot below, followed by the code. It seems the two countries I found as outliers are Singapore and Hong Kong. Note that I did not plot the names for other countries, since the plot would be barely legible.



```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

#filepaths for country data, names
filepathmat = 'datafiles/processed.txt'
filepathcont = 'datafiles/row.txt'
filepathdemvar = 'datafiles/column.txt'

#read in country names, then data with index as names
cont = pd.read_csv(filepathcont, header=None)
demvar = pd.read_csv(filepathdemvar, header=None)
df = pd.read_csv(filepathmat, header=None)
df = df.set_axis(cont[0], axis='index')

#Write to numpy array to do SVD
```

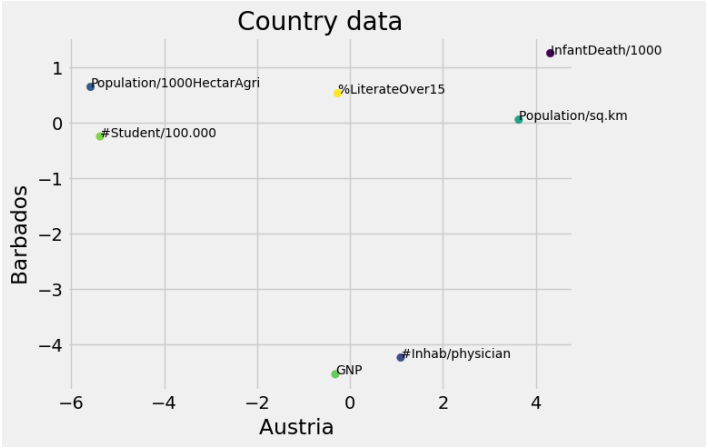
```
df_arr = df.to_numpy()
U, S, V = np.linalg.svd(df_arr)

testpts = []

j, k = 1, 2
for pts in range(U.shape[0]):
    testpts.append((S[j] *U[pts][j], S[k]*U[pts][k]))
pts = np.array(testpts)

#Finding outliers
outlier1 = np.argmin(pts[:, 0])
outpt1 = (outlier1, pts[outlier1, :])
outlier2 = np.argmin(np.delete(pts[:, 0], outlier1))
outpt2 = (outlier2, pts[outlier2, :])

outliers = np.array([df.index[outlier1], df.index[outlier2]])

for x, y in pts:
    plt.scatter(x, y)

plt.title("Country data")
plt.style.use('fivethirtyeight')
plt.gray
plt.ylabel(f'Projection {j}')
plt.xlabel(f'Projection {k}')
#Labelling outliers on plot
plt.annotate(f'{outliers[0]}', xy=(outpt1[1]) )
plt.annotate(f'{outliers[1]}', xy=(outpt2[1]) )
```

Problem 1, part c

Now do the same with the two left singular vectors of A, $\mathbf{u}_1, \mathbf{u}_2 \in \mathbb{R}^{49}$. Project the column vectors (i.e., variables) $\alpha_1, ..., \alpha_7 \in \mathbb{R}^{49}$ onto the two-dimensional space $\text{span}\{\mathbf{u}_1, \mathbf{u}_2\} \cong \mathbb{R}^2$ and plot this in a graph as before. Note that in this case, the points correspond to the demographic variables — label them accordingly.

Solution:

The plot will be provided below, as well as the code.



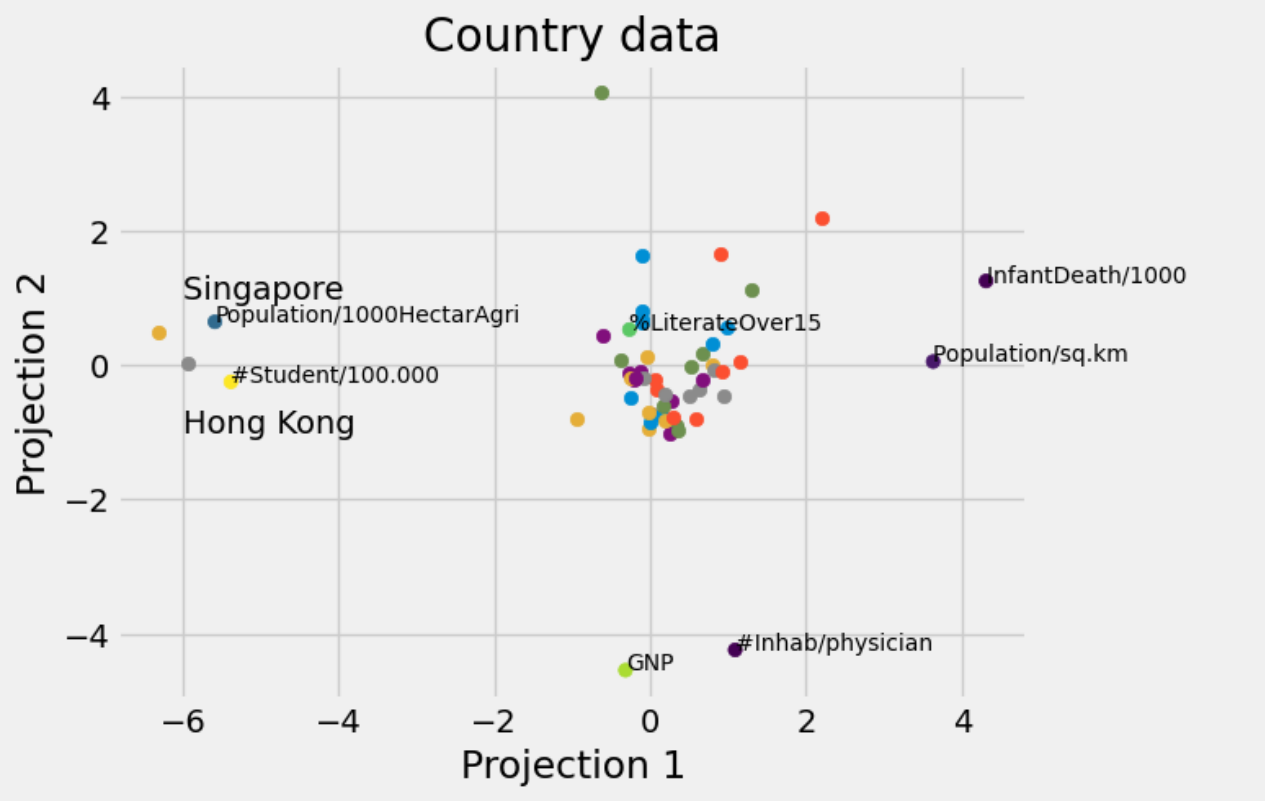
```
#Doing the same except for demographic variables
testpts = []

j, k = 1, 2
for pts in range(V.shape[0]):
    testpts.append((S[j] *V[pts][j], S[k]*V[pts][k]))
pts = np.array(testpts)


# Create the scatter plot
plt.scatter(pts[:, 0], pts[:, 1], c = [random.randint(0, x*10) for x in range(len(V))])

#Finding outliers
for i, (x, y) in enumerate(pts):
    plt.annotate(f'{demvar[0][i]}', xy=(x, y), fontsize=10)

plt.title("Country data")
plt.style.use('fivethirtyeight')
plt.gray
plt.ylabel(f'{df.index[k]}')
plt.xlabel(f'{df.index[j]}')
```

Problem 1, part d

Overlay the two graphs in (b) and (c). Identify the two demographic variables near the two outlier countries — these explain why the two countries are outliers.

Solution:

Here is my plot. It seems like the two demographic variables very near to Singapore and Hong Kong are Population per 1000 Hectacr Agri and number of students per 100,000. These seem very reasonable, since these two countries are tightly packed, huge populations, and an emphasis on rigorous childhood education.



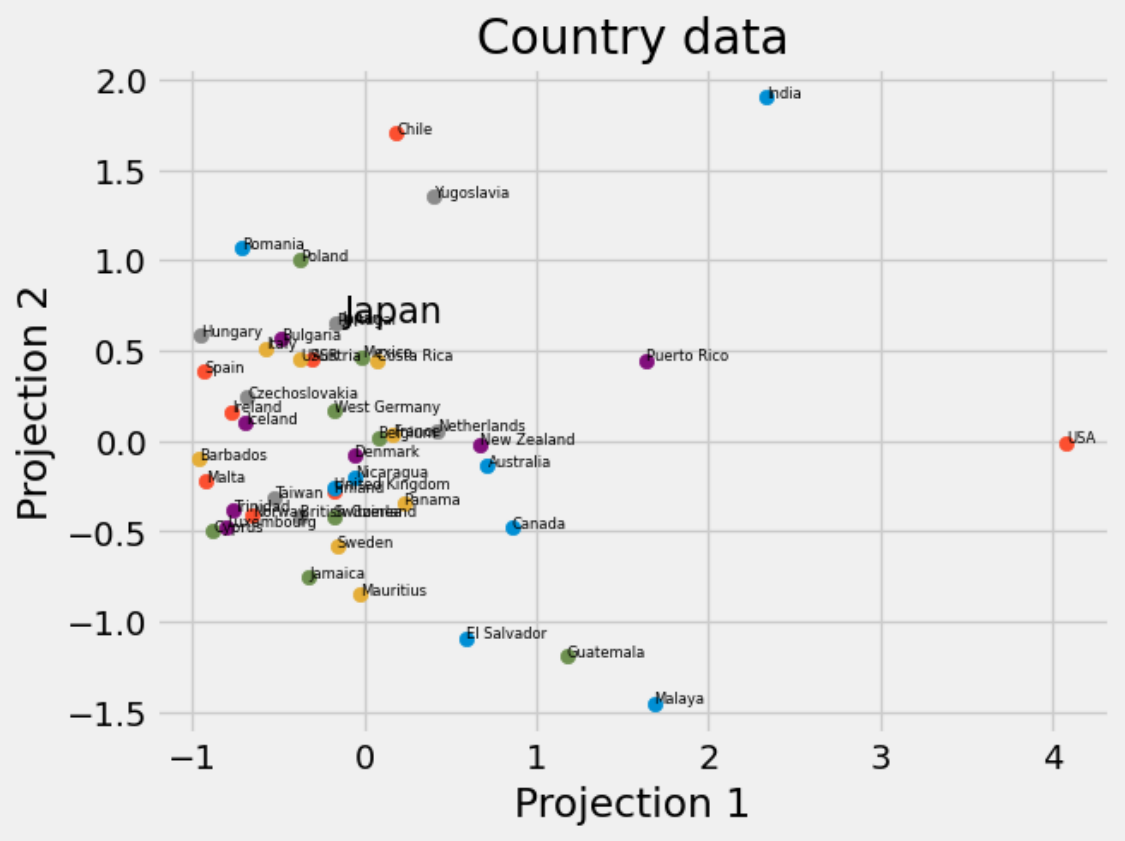
Problem 1, part e

Remove the two outlier countries and redo (b) with this 47 × 7 matrix. This allows you to see features that were earlier obscured by the outliers. Which two European countries are most alike Japan?

Solution: Since you can't see the name too well on the map, I made some code to find the first few countries nearest to japan. Here's the code, then the plot. It seems that Portugal and Austria are most like japan in these projections (that are in Europe!).

```
dist = np.array([np.linalg.norm((x - pts[22][0], y - pts[22][1])) for x, y in pts])
k = 5
indices = np.argpartition(dist, k)[:k]

# Get the 3 minimum values and their indices
min_values = dist[indices]

print("Closest countries to Japan:", df.index[indices])
```

```
Closest countries to Japan: Index(['Portugal ', 'Japan    ', 'Austria    ', 'Mexico     ',
       'Costa Rica  '],
      dtype='object', name=0)
```



Problem 2

Let $A, B \in \mathbb{R}^{m \times n}$ where A has full column rank.

Problem 2, part a

Show that

$$\min_{X \in \mathbb{R}^{n \times m}} \|AX - \mathbb{I}_m\|_F$$

has a unique solution. What is the minimum length solution, i.e., where $\|X\|_F$ is minimum?

Solution:

Suppose false, that is, there exists $X, X' \in \mathbb{R}^{n \times m}$ with $X \neq X'$ that achieve minimum. Then, via SVD, we can deduce a contradiction. Note the norm we are working with is the Frobenius norm, i will omit this in writing. We can also write $\mathbb{I}_m = UU^T$.

$$\|AX' - \mathbb{I}_m\| = \|AX - \mathbb{I}_m\| \qquad \text{(Both in argmin.)}$$

$$\left\|U\Sigma V^t X' - UU^T\right\| = \left\|U\Sigma V^t X - UU^T\right\| \qquad \text{(SVD of } A\text{.)}$$

$$\left\|\Sigma V^T X' - U^T\right\| = \left\|\Sigma V^T X - U^T\right\| \qquad \text{(Unitary invariance.)}$$

$$\text{tr}\left[(\Sigma V^T X' - U^T)^T(\Sigma V^T X' - U^T)\right] = \text{tr}\left[(\Sigma V^T X' - U^T)^T(\Sigma V^T X' - U^T)\right] \qquad \text{(Frobenius trace.)}$$

$$\text{tr}\left[(\Sigma V^T X')^T\Sigma V^T X' - (\Sigma V^T X')^T U^T - U\Sigma V^T X' + UU^T\right] = \qquad (1)$$

$$\text{tr}\left[(\Sigma V^T X)^T\Sigma V^T X - (\Sigma V^T X')^T U^T - U\Sigma V^T X + UU^T\right] \qquad (2)$$

$$1) : \text{tr}\left[X'^T V\Sigma\Sigma V^T X' - X'^T V\Sigma U^T - U\Sigma V^T X' + UU^T\right] \qquad \text{(By symmetry 1 = 2.)}$$

$$\text{tr}\left[X'^T V\Sigma\Sigma V^T X'\right] - \text{tr}\left[X'^T V\Sigma U^T\right] - \text{tr}\left[U\Sigma V^T X'\right] + \text{tr}\left[UU^T\right] \qquad \text{(Trace is linear.)}$$

$$\text{tr}\left[X'^T V\Sigma\Sigma V^T X'\right] - \text{tr}\left[X'^T A^\dagger\right] - \text{tr}\left[AX'\right] + \text{tr}\left[UU^T\right] \qquad (A \text{ and } A\dagger.)$$

$$\text{tr}\left[V\Sigma\Sigma V^T X' X'^T\right] - \text{tr}\left[X'^T A^\dagger\right] - \text{tr}\left[AX'\right] + \text{tr}\left[UU^T\right] \qquad \text{(Permutation in trace.)}$$

$$\text{tr}\left[V\Sigma\Sigma V^T(X'X'^T - XX^T)\right] + \text{tr}\left[(X^T - X'^T)A^\dagger\right] + \text{tr}\left[A(X - X')\right] = 0 \qquad \text{(Combining both.)}$$

$$\left\|V\Sigma\Sigma V^T(X'X'^T - XX^T)\right\| + \left\|(X^T - X'^T)A\right\| + \|A(X - X')\| = 0 \qquad \text{(Frobenius definition.)}$$

$$\implies \|A(X - X')\| = 0 \qquad \text{(All norms} \geq 0\text{, so all are 0.)}$$

$$\iff X - X' = 0 \qquad \text{(Norm definition.)}$$

$$\iff X = X'$$

Thus we found $X = X'$, which is a contradiction, so X is unique. Aplologies for the bad formatting, this is the only way I could get it to fit.

Problem 2, part b

Show that the following method produces a symmetric matrix $X \in \mathbb{R}^{n \times n}$ that solves

$$\min_{X^T = X} \|AX - B\|_F$$

i)

show that the SVD of A takes the form

$$A = U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^T$$

Solution:

We are given that A is full column rank, thus full rank. Thus all singular values of $A > 0$. We then have two choices for Σ'. Either Σ' is "fat", where the zero matrix appears to the right of the singular values, or Σ' is "skinny," meaning the zero matrix appear below. In either case, Σ' is $m \times n$, meaning it retains the same shape as A. Since A is full *column* rank, that means Σ' spans the entirety of n, so we are left with Σ' being the skinny case, which is the one shown.

ii)

Show that

$$\|AX - B\|_F^2 = \|\Sigma Y - C_1\|_F^2 + \|C_2\|_F^2$$

where $Y = V^T X V$ and $C = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} = U^T B V$.

Solution:

We first note that the Frobenius norm is unitary invariant. That is, multiplying the argument inside by a unitary matrix does nothing to the result. If we multiply by V, the right eigenvectors of A, we get the following:

$$
\begin{aligned}
\|AX - B\|_F^2 &= \|AXV - BV\|_F^2 && \text{(Justified above.)} \\
&= \left\|U\Sigma' V^T XV - BV\right\|_F^2 && \text{(SVD of } A\text{.)} \\
&= \mathrm{tr}\left[(U\Sigma' Y - BV)^T (U\Sigma' Y - BV)\right] && \text{(Frobenius trace and } Y \text{ definition.)} \\
&= \mathrm{tr}\left[Y^T \Sigma'^T U^T U\Sigma' Y - Y^T \Sigma'^T U^T BV\right] && \text{(Distribution.)} \\
&= \mathrm{tr}\left[Y^T \Sigma'^T \Sigma^T Y - Y^T \Sigma'^T C - C^T \Sigma' Y + V^T B^T UU^T BV\right] && \text{(Definition of } C\text{.)} \\
&= \mathrm{tr}\left[Y^T \Sigma^T \Sigma^T Y - Y^T \Sigma^T C_1 - C_1^T \Sigma^T Y + C^T C\right] && \text{(Form of } \Sigma' \text{ and } C\text{.)} \\
&= \mathrm{tr}\left[(\Sigma Y)^T \Sigma Y - (\Sigma Y)^T C_1 - C_1^T \Sigma Y + C_1^T C_1 + C_2^T C_2\right] && \text{(Rewriting and form of } C\text{.)} \\
&= \mathrm{tr}\left[(\Sigma Y - C_1)^T (\Sigma Y - C_1)\right] + \mathrm{tr}\left[C_2^T C_2\right] && \text{(Trace is linear.)} \\
&= \|\Sigma Y - C_1\|_F^2 + \|C_2\|_F^2 && \text{(Frobenius trace.)}
\end{aligned}
$$

iii)

Note that Y must be symmetric if X is. Show that

$$\|\Sigma Y - C_1\|_F^2 = \sum_{i=1}^{n} |\Sigma_i y_{ii} - c_{ii}|^2 + \sum_{j>i} (|\sigma_i y_{ij} - c_{ij}|^2 + |\sigma_j y_{ij} - c_{ji}|^2$$

and deduce that the minimum value above is attained when

$$y_{ij} = \frac{\sigma_i c_{ij} + \sigma_j c_{ji}}{\sigma_i^2 + \sigma_j^2}$$

for all $i, j = 1, ..., n$.

Solution:

This can be shown by the Frobenius norm and symmetry of Y.

$$\|\Sigma Y - C_1\|_F^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} |\sigma_i y_{ij} - c_{ij}|^2 \qquad \text{(Definition of Frobenius norm.)}$$

$$= \sum_{i=1}^{n} \left[\sum_{j=1}^{i-1} |\sigma_i y_{ij} - c_{ij}|^2 + |\sigma_i y_{ii} - c_{ii}|^2 + \sum_{j=i+1}^{n} |\sigma_i y_{ij} - c_{ij}|^2 \right] \qquad \text{(Breaking } j \text{ sum up.)}$$

$$= \sum_{i=1}^{n} |\sigma_i y_{ii} - c_{ii}|^2 + \sum_{i=1}^{n} \left[\sum_{j=1}^{i-1} |\sigma_i y_{ij} - c_{ij}|^2 + \sum_{j=i+1}^{n} |\sigma_i y_{ij} - c_{ij}|^2 \right] \qquad \text{(Distribution.)}$$

$$= \sum_{i=1}^{n} |\sigma_i y_{ii} - c_{ii}|^2 + \sum_{i=1}^{n} \left[\sum_{j=i+1}^{n} |\sigma_j y_{ji} - c_{ji}|^2 + \sum_{j=i+1}^{n} |\sigma_i y_{ij} - c_{ij}|^2 \right] \qquad \text{(Transposition.)}$$

$$= \sum_{i=1}^{n} |\sigma_i y_{ii} - c_{ii}|^2 + \sum_{j>i} \left[|\sigma_j y_{ij} - c_{ji}|^2 + |\sigma_i y_{ij} - c_{ij}|^2 \right] \qquad \text{(Symmetry of } Y \text{ and simplifying.)}$$

We should expect the minimal symmetric X to have only diagonal terms. Thus, we set the terms of the second sum to be equal zero. Thus, the following can be shown:

$$
\begin{aligned}
0 &= (\sigma_j y_{ij} - c_{ji})^2 + (\sigma_i y_{ij} - c_{ij})^2 \\
&= \sigma_j^2 y_{ji}^2 + c_{ji}^2 - 2\sigma_j y_{ij} c_{ji} + \sigma_i^2 y_{ij}^2 + c_{ij}^2 - 2\sigma_i y_{ij} c_{ij} && \text{(Expanding.)} \\
&= y_{ij}^2(\sigma_j^2 + \sigma_i^2) - 2(\sigma_j c_{ji} + \sigma_i c_{ij}) y_{ij} + (c_{ij}^2 + c_{ji}^2) && \text{(Grouping, polynomial in } y_{ij}.)
\end{aligned}
$$

We can note that this polynomial is strictly positive, as this was how it was constructed. Then the minimum of the polynomial is at the vertex, $-b/2a$, meaning,

$$y_{ij} = \frac{\sigma_j c_{ji} + \sigma_i c_{ij}}{\sigma_i^2 + \sigma_j^2}$$

which is what we wanted to show.

Problem 2, part c

Now emulate the previous part to find the rank-r matrix $X \in \mathbb{R}^{m \times n}$ that solves

$$\min_{\text{rank}(X) \leq r} \|AX - B\|_F.$$

Solution:

Sorry, I did not have time this week to get to this part.

Problem 3

Let $A \in \mathbb{C}^{m\times n}$ and $\mathbf{b} \in \mathbb{C}^m$. We will discuss a variant of $A\mathbf{x} \approx \mathbf{b}$ where the error occurs only in A. Note that in ordinary least squares we assume that the error occurs only in $\mathbf{b}$ while in total least squares we assume that it occurs in both A and $\mathbf{b}$.

Problem 3, part a

Show that if $0 \neq \mathbf{x} \in \mathbb{C}^n$, then

$$\left\| A\left(\mathbb{I} - \frac{\mathbf{x}\mathbf{x}^*}{\mathbf{x}^*\mathbf{x}}\right)\right\|_F^2 = \|A\|_F^2 - \frac{\|A\mathbf{x}\|_2^2}{\mathbf{x}^*\mathbf{x}}$$

Solution:

Taking the left side, the following steps can be shown:

$$\left\| A\left(\mathbb{I} - \frac{\mathbf{x}\mathbf{x}^*}{\mathbf{x}^*\mathbf{x}}\right)\right\|_F^2 = \operatorname{tr}\left[\left(A - \frac{1}{\mathbf{x}^*\mathbf{x}}A\mathbf{x}\mathbf{x}^*\right)^*\left(A - \frac{1}{\mathbf{x}^*\mathbf{x}}A\mathbf{x}\mathbf{x}^*\right)\right] \qquad \text{(Frobenius trace.)}$$

$$= \operatorname{tr}\left[A^*A\right] - \frac{1}{\mathbf{x}^*\mathbf{x}}\operatorname{tr}\left[A^*A\mathbf{x}\mathbf{x}^*\right] - \frac{1}{\mathbf{x}^*\mathbf{x}}\operatorname{tr}\left[(A\mathbf{x}\mathbf{x}^*)^*A\right] + \frac{1}{(\mathbf{x}^*\mathbf{x})^2}\operatorname{tr}\left[(A\mathbf{x}\mathbf{x}^*)^*A\mathbf{x}\mathbf{x}^*\right] \qquad \text{(Expanding.)}$$

$$= \|A\|_F^2 - \frac{1}{\mathbf{x}^*\mathbf{x}}\operatorname{tr}\left[A^*A\mathbf{x}\mathbf{x}^*\right] - \frac{1}{\mathbf{x}^*\mathbf{x}}\operatorname{tr}\left[\mathbf{x}\mathbf{x}^*A^*A\right] + \frac{1}{(\mathbf{x}^*\mathbf{x})^2}\operatorname{tr}\left[\mathbf{x}\mathbf{x}^*A^*A\mathbf{x}\mathbf{x}^*\right] \qquad \text{(Hermiting(?).)}$$

$$= \|A\|_F^2 - \frac{1}{\mathbf{x}^*\mathbf{x}}\operatorname{tr}\left[\mathbf{x}^*A^*A\mathbf{x}\right] - \frac{1}{\mathbf{x}^*\mathbf{x}}\operatorname{tr}\left[\mathbf{x}^*A^*A\mathbf{x}\right] + \frac{1}{(\mathbf{x}^*\mathbf{x})^2}\operatorname{tr}\left[A^*A\mathbf{x}\mathbf{x}^*\mathbf{x}\mathbf{x}^*\right] \qquad \text{(Trace permutating.)}$$

$$= \|A\|_F^2 - \frac{1}{\mathbf{x}^*\mathbf{x}}\operatorname{tr}\left[(A\mathbf{x})^*A\mathbf{x}\right] - \frac{1}{\mathbf{x}^*\mathbf{x}}\operatorname{tr}\left[(A\mathbf{x})^*A\mathbf{x}\right] + \frac{1}{(\mathbf{x}^*\mathbf{x})^2}\operatorname{tr}\left[A^*A\mathbf{x}(\mathbf{x}^*\mathbf{x})\mathbf{x}^*\right] \qquad \text{(Hermiting, associating.)}$$

$$= \|A\|_F^2 - \frac{1}{\mathbf{x}^*\mathbf{x}}\|A\mathbf{x}\|_2^2 - \frac{1}{\mathbf{x}^*\mathbf{x}}\|A\mathbf{x}\|_2^2 + \frac{1}{\mathbf{x}^*\mathbf{x}}\operatorname{tr}\left[A^*A\mathbf{x}\mathbf{x}^*\right] \qquad \text{(Trace of scalar is scalar.)}$$

$$= \|A\|_F^2 - \frac{1}{\mathbf{x}^*\mathbf{x}}\|A\mathbf{x}\|_2^2 - \frac{1}{\mathbf{x}^*\mathbf{x}}\|A\mathbf{x}\|_2^2 + \frac{1}{\mathbf{x}^*\mathbf{x}}\operatorname{tr}\left[\mathbf{x}^*A^*A\mathbf{x}\right] \qquad \text{(Permutating.)}$$

$$= \|A\|_F^2 - \frac{1}{\mathbf{x}^*\mathbf{x}}\|A\mathbf{x}\|_2^2 - \frac{1}{\mathbf{x}^*\mathbf{x}}\|A\mathbf{x}\|_2^2 + \frac{1}{\mathbf{x}^*\mathbf{x}}\|A\mathbf{x}\|_2^2 \qquad \text{(Trace of scalar is scalar.)}$$

$$= \|A\|_F^2 - \frac{\|A\mathbf{x}\|_2^2}{\mathbf{x}^*\mathbf{x}} \qquad \text{(Simplifying.)}$$

Problem 3, part b

Show that the matrix

$$E = \frac{(\mathbf{b} - A\mathbf{x})\mathbf{x}^*}{\mathbf{x}^*\mathbf{x}} \in \mathbb{C}^{m\times n}$$

has the smallest 2-norm among all $E \in \mathbb{C}^{m\times n}$ that satisfy

$$(A + E)\mathbf{x} = \mathbf{b}$$

Solution:

Note that all E's satisfy $E\mathbf{x} = \mathbf{b} - A\mathbf{x}$. To distinguish the E matrices, we'll denote E' as a matrix that satisfies $(A + E')\mathbf{x} = \mathbf{b}$. Thus, the following can be shown:

$$\begin{aligned}
\|E\|_2^2 &= \left\| \frac{(\mathbf{b} - A\mathbf{x})\mathbf{x}^*}{\mathbf{x}^*\mathbf{x}} \right\|_2^2 && \text{(Given.)} \\
&= \left| \frac{1}{\mathbf{x}^*\mathbf{x}} \right|^2 \|(\mathbf{b} - A\mathbf{x})\mathbf{x}^*\|_2^2 && \text{(Denominator is real.)} \\
&\leq \left| \frac{1}{\mathbf{x}^*\mathbf{x}} \right|^2 \|\mathbf{b} - A\mathbf{x}\|_2^2 \|\mathbf{x}^*\|_2^2 && \text{(Cauchy Schwartz.)} \\
&= \frac{\|\mathbf{b} - A\mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2} && \text{(Simplifying.)} \\
&= \frac{\|E'\mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2} && \text{(Equality above.)} \\
&\leq \|E'\|_2^2 \frac{\|\mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2} && \text{(Consistency of 2-norm.)} \\
\implies \|E\|_2^2 &\leq \|E'\|_2^2 && \text{(Simplifying.)} \\
\iff \|E\|_2 &\leq \|E'\|_2
\end{aligned}$$

Problem 3, part c

Let $A, \mathbf{b}, \mathbf{x}$ be given and fixed. What are the solutions of

$$\min_{(A+E)\mathbf{x}=\mathbf{b}} \|E\|_2 \quad \text{and} \quad \min_{(A+E)\mathbf{x}=\mathbf{b}} \|E\|_F$$

where the minimum is taken over all $E \in \mathbb{C}^{m\times n}$ such that $(A+E)\mathbf{x} = \mathbf{b}$?

Solution:

We just showed in the previous part that when

$$E = \frac{(\mathbf{b} - A\mathbf{x})\mathbf{x}^*}{\mathbf{x}^*\mathbf{x}} \in \mathbb{C}^{m\times n}$$

Then E is the smallest of all E's satisfying $(A+E)\mathbf{x} = \mathbf{b}$. We can then just plug this value for E in to get the minimum value for each.

$$\begin{aligned}
\min_{(A+E)\mathbf{x}=\mathbf{b}} \|E\|_2 &= \left\| \frac{(\mathbf{b} - A\mathbf{x})\mathbf{x}^*}{\mathbf{x}^*\mathbf{x}} \right\|_2 && \text{(Plugging in.)} \\
&= \frac{1}{\|\mathbf{x}\|_2^2} \|\mathbf{b}\mathbf{x}^* - A\mathbf{x}\mathbf{x}^*\|_2 && \text{(Simplifying.)} \\
\min_{(A+E)\mathbf{x}=\mathbf{b}} \|E\|_F &= \frac{1}{\|\mathbf{x}\|_2^2} \mathrm{tr}\Big[(\mathbf{b}\mathbf{x}^* - A\mathbf{x}\mathbf{x}^*)^* (\mathbf{b}\mathbf{x}^* - A\mathbf{x}\mathbf{x}^*) \Big] && \text{(Frobenius trace.)} \\
&= \frac{1}{\|\mathbf{x}\|_2^2} \mathrm{tr}\Big[\mathbf{x}\mathbf{b}^*\mathbf{b}\mathbf{x}^* + \mathbf{x}\mathbf{x}^*A^*A\mathbf{x}\mathbf{x}^* - \mathbf{x}\mathbf{b}^*A\mathbf{x}\mathbf{x}^* - \mathbf{x}\mathbf{x}^*A^*\mathbf{b}\mathbf{x}^* \Big] && \text{(Hermiting.)} \\
&= \frac{1}{\|\mathbf{x}\|_2^2} \mathrm{tr}\Big[\|\mathbf{b}\|_2^2 \mathbf{x}\mathbf{x}^* + A^*A\mathbf{x}\mathbf{x}^*\mathbf{x}\mathbf{x}^* - \mathbf{b}^*A\mathbf{x}\mathbf{x}^*\mathbf{x} - \mathbf{x}^*A^*\mathbf{b}\mathbf{x}^*\mathbf{x} \Big] && \text{(Permuting.)} \\
&= \frac{1}{\|\mathbf{x}\|_2^2} \mathrm{tr}\Big[\|\mathbf{b}\|_2^2\|\mathbf{x}\|_2^2 + \|\mathbf{x}\|_2^2\|A\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2\mathbf{b}^*A\mathbf{x} - \|\mathbf{x}\|_2^2\mathbf{x}^*A^*\mathbf{b} \Big] && \text{(Taking norms.)} \\
&= \mathrm{tr}\Big[\|\mathbf{b}\|_2^2 + \|A\mathbf{x}\|_2^2 - \mathbf{b}^*A\mathbf{x} - \mathbf{x}^*A^*\mathbf{b} \Big] && \text{(Simplifying.)} \\
&= \|\mathbf{b}\|_2^2 + \|A\mathbf{x}\|_2^2 - 2\mathrm{tr}\Big[\mathbf{b}^*A\mathbf{x} \Big] && \text{(Simplifying.)}
\end{aligned}$$

Problem 3, part d

Given $\mathbf{a} \in \mathbb{C}^n$, $\mathbf{b} \in \mathbb{C}^m$, and $\delta > 0$. Show how to solve the problems

$$\min_{\|E\|_F \leq \delta} \|E\mathbf{a} - \mathbf{b}\|_2 \quad \text{and} \quad \max_{\|E\|_F \leq \delta} \|E\mathbf{a} - \mathbf{b}\|_2$$

Solution:

Sorry, I didn't get to this part in time, but I will try to explain my thoughts. I would imagine you would want E to map $\mathbf{a}$ the closest to $\mathbf{b}$ in the minimize problem, but E can only stretch $\mathbf{a}$ at most size δ. If The distance from $\mathbf{a}$ to $\mathbf{b}$ is less than δ, then the minimum can be achieved for some E, so our minimum value would be 0. If not, then you would just take the distance from zero to the origin. If this distance is smaller than the distance to $\mathbf{b}$, then just let $E = 0$.

For the maximizer, you would just want E to map $\mathbf{a}$ the furthest away from $\mathbf{b}$. If $\mathbf{a}$ is greater than $\mathbf{b}$ by distance δ, then you would just map it further away.

Problem 4

In the following, $\kappa(A) := \|A\|\left\|A^\dagger\right\|$ for $A \in \mathbb{C}^{m\times n}$ where $\|\cdot\|$ denotes a submultiplicative matrix norm. We will write $\kappa_p(A)$ if the norm involved is a matrix p-norm.

Problem 4, part a

Show that for any nonzero $A \in \mathbb{C}^{m\times n}$, $\kappa(A) \geq 1$.

Solution:

We are given that $\|\cdot\|$ is submultiplicative, thus $\|A\|\left\|A^\dagger\right\| \geq \left\|AA^\dagger\right\|$. If A is invertible, then $AA^\dagger = AA^{-1} = \mathbb{I}$. Note that $\|\mathbb{I}\| = \left\|\mathbb{I}^2\right\| \leq \|\mathbb{I}\|^2$, thus $\|\mathbb{I}\| \geq 1$ for any submultiplicative norm. Then $\kappa(A) = \|A\|\left\|A^\dagger\right\| \geq 1$. If A is not invertible, then we can use SVD. We will write $A = U\Sigma V^*$, so $A^\dagger = V\Sigma^\dagger U^*$. Then,

---

$$
\begin{aligned}
\left\|AA^\dagger\right\| &= \left\|U\Sigma V^* V\Sigma^\dagger U^*\right\| && \text{(By SVD.)} \\
&= \left\|U\Sigma\Sigma^\dagger U^*\right\| && (V \text{ is hermitian.}) \\
&= \left\|U\begin{bmatrix}\mathbb{I}_r \\ 0\end{bmatrix}U^*\right\|, \left\|U\begin{bmatrix}\mathbb{I}_r & 0\end{bmatrix}U*\right\| && \text{(Depends if } m \geq n, \text{ or reverse.)} \\
&= \|U_r U^*\| && \text{(In either case, we take the first r vectors of } U.) \\
&= \|\mathbb{I}_r\| && \text{(Simplifying.)} \\
&\geq 1 && \text{(Identity is } \geq 1.)
\end{aligned}
$$

Note in these steps, I took the first r rows *or* columns from U or U^*. Without loss of generality, since U is hermitian, the above is given. Note that $\mathbb{I}_r$ in the last step is an $m \times n$ matrix, which I take its norm as in the $r \times r$ case.

Problem 4, part b

Show that for any $A \in \mathbb{C}^{m \times n}$,

$$\kappa_2(A^*A) = \kappa_2(A)^2$$

but that in general,

$$\kappa(A^*A) \neq \kappa(A)^2$$

Solution:

We know that A^*A will have eigenvalues $|\lambda|^2$ for eigenvalue λ of A. Thus, $\|A^*A\|_2 = \max\{|\lambda|\} = (\max\{|\lambda|\})^2$. Similarly, $(A^*A)^{-1}$ has eigenvalues $1/|\lambda|^2$. Note that A^*A is positive *semi*-definite, so any eigenvalue of A^*A is ≥ 0. If there is an eigenvalue equal zero, then $\kappa_2(A^*A) = \infty$. This also happens for $\kappa(A)$, since $\max\{1/|\lambda|\} = \infty$. If we still take the SVD of A^*A, just to be safe, then

$$\kappa_2(A^*A) = \|A^*A\|_2 \left\|(A^*A)^{-1}\right\|_2 = \frac{(\max|\lambda|)^2}{(\max 1/|\sigma^2|)} = (\max|\lambda|)^2(\min|\sigma|)^2$$

Here we cannot assume that A is invertible, but by SVD, we can say what the 2-norm of A^{-1} is. This will then be the largest singular value of $\Sigma^\dagger$, which is the smallest singular value of Σ. Thus when squaring, we get

$$\kappa_2(A)^2 = (\|A\|_2 \left\|A^\dagger\right\|_2)^2 = (\max\{|\lambda|\})^2(\min\{\sigma\})^2$$

Thus, these two are equivalent. This should hold true for any norm, by the equivalence of norms over a finite dimensional vector space.

Problem 4, part c

Show that for nonsingular $A, B \in \mathbb{C}^{n\times n}$,

$$\kappa(AB) \leq \kappa(A)\kappa(B).$$

It this true in general without the nonsingular condition?

Solution:

For nonsingular A, B, this property is easy to show. Note the pseudoinverse turns into a normal inverse for the product.

$$\begin{aligned}
\kappa(AB) &= \|AB\|\left\|(AB)^{-1}\right\| && \text{(Given.)} \\
&= \|AB\|\left\|B^{-1}A^{-1}\right\| && \text{(Inverse of } AB\text{.)} \\
&\leq \|A\|\|B\|\left\|B^{-1}\right\|\left\|A^{-1}\right\| && \text{(Submultiplication.)} \\
\implies \kappa(AB) &\leq \kappa(A)\kappa(B)
\end{aligned}$$

Problem 4, part d

Let $Q \in \mathbb{C}^{n\times n}$ be a matrix with orthonormal columns. Show that

$$\kappa_2(Q) = 1.$$

Is this true if Q has orthonormal rows instead? Is this true with κ_1 or κ_∞ in place of κ_2?

Solution: First, note that the 2-norm is unitary invariant, Since $Q^* = Q^{-1}$, then

$$\kappa_2(Q) = \|Q\|_2\left\|Q^{-1}\right\|_2 = \|Q\|_2\left\|Q^T\right\|_2 = \|\mathbb{I}\|_2\|\mathbb{I}\|_2 = 1$$

I wrote the above equality to emphasize that this will work is Q has orthonormal rows instead, since the Q and Q^T terms will just swap around. Since κ_1 is taken with the one matrix norm, it will take the max absolute row sum of Q, which is 1 since Q has orthonormal columns. This is the same for κ_∞, as again the Q and Q^T terms will just swap.

Problem 4, part e

Let $R \in \mathbb{C}^{n\times n}$ be a nonsingular upper-triangular matrix. Show that

$$\kappa_\infty(R) \geq \frac{\max_{i=1,\ldots,n}|r_{ii}|}{\min_{i=1,\ldots,n}|r_{ii}|}$$

Note for an upper triangular matrix, its eigenvalues are on the diagonal, so we can note the numerator of the right hand side is just $\|R\|_2$. If we can accept that the diagonal of the inverse of R is just the inverted diagonal of R, then

$$\min|r_{ii}| = \max 1/|r_{ii}| = \max 1/|\lambda_i| = \left\|R^{-1}\right\|_2$$

That is, since the upper triangular matrices form a closed subgroup of $GL(n)$, then we can say the diagonal or R^{-1} is the inverse of the diagonal of R. We already know that R^{-1} will have eigenvaules $1/\lambda_i$ for λ_i of R. They are all nonzero, so this is valid. So the right hand side is just $\kappa_2(R)$. We just need to show that

$$\kappa_\infty(R) \geq \kappa_2(R).$$

By the equivalence of norms, $\|R\|_2 \leq \sqrt{n}\|R\|_\infty$ for any positive n, so $\|R\|_2 \leq \|R\|_\infty$. This argument also holds for R^{-1}. Thus, $\kappa_2(R) \leq \kappa_\infty(R)$.

Problem 4, part f

Show that for any nonsingular $A \in \mathbb{C}^{n\times n}$,

$$\kappa(A) \leq \max\left\{\frac{\|AX-\mathbb{I}\|}{\|XA-\mathbb{I}\|}, \frac{\|XA-\mathbb{I}\|}{\|AX-\mathbb{I}\|}\right\}$$

Solution:

We are given $\|\cdot\|$ is submultiplicative, so we can note, by the hint, that

$$\|AX-\mathbb{I}\| = \left\|A(XA-\mathbb{I})A^{-1}\right\| \leq \|A\|\|XA-\mathbb{I}\|\left\|A^{-1}\right\|$$

By rearranging, we get

$$\frac{\|AX-\mathbb{I}\|}{\|XA-\mathbb{I}\|} \leq \|A\|\left\|A^{-1}\right\| = \kappa(A)$$

This shows the first inequality. Next note that we could have also don't the same for $XA-\mathbb{I}$, that is, $A^{-1}(AX-\mathbb{I})A = XA-\mathbb{I}$. We can then write the following

$$\|XA-\mathbb{I}\| = \left\|A^{-1}(AX-\mathbb{I})A\right\| \leq \left\|A^{-1}\right\|\|AX-\mathbb{I}\|\|A\|.$$

Which can then be rewritten as

$$\frac{\|XA-\mathbb{I}\|}{\|AX-\mathbb{I}\|} \leq \left\|A^{-1}\right\|\|A\| = \kappa(A),$$

thus the second inequality has been shown, so $\kappa(A)$ is greater than both, so it is greater than the max.

Problem 5

We will examine the effect of various parameters on the accuracy of a computed solution to a nonsingular linear system. Relevant commands in Matlab syntax are given in brackets.

Problem 5, part a

Generate $A = [a_{ij}] \in \mathbb{R}^{n \times n}$ as follows:

1. a_{ij} randomly generated from a standard normal distribution $[randn(n)]$.
2. A Hilbert matrix, i.e. $a_{ij} = 1/(i + j - 1)$
3. A Pascal matrix
4. A magic square

Solution: I will provide my definitions below. Note that I had some external help creating the magic square.

```
import numpy as np
import random
import math
def random_matrix(n: int):
    """
    Returns a numpy matrix of size n by n filled with random numbers
    """
    res = np.zeros((n, n))
    for i in range(res.shape[0]):
        for j in range(res.shape[1]):
            res[i][j] = random.gauss(0, 1)
    return res

def hilbert_matrix(n: int):
    """
    Returns a Hilbert matrix of size n by n where
    a_{ij} = 1/ (i + j - 1)
    """
    res = np.zeros((n , n))
    for i in range(n):
        for j in range(n):
            if i== 0 and j ==0:
                res[i][j] = 1
            else:
```

```
            res[i][j] = float(1 / ((i+1) + (j+1) - 1))
    return res

def pascal_matrix(n: int):
    """
    Returns a Pascal matrix of size n with entries
    a_{ij} = {i + j \choose i}
    """
    res = np.zeros((n, n))
    for i in range(n):
        for j in range(n):
            res[i][j] = math.comb(i+j, i)
    return res

def magic_matrix(N: int):
    """
    Returns a magic square of size N by N.
    Courtesy of https://scipython.com/book/chapter-6-numpy/examples/creating-a-magic-square/
    """

    magic_square = np.zeros((N,N), dtype=int)

    n = 1
    i, j = 0, N//2

    while n <= N**2:
        magic_square[i, j] = n
        n += 1
        newi, newj = (i-1) % N, (j+1)% N
        if magic_square[newi, newj]:
            i += 1
        else:
            i, j = newi, newj

    return magic_square
```

Problem 5, part b

Generate $\mathbf{x}$ and $\mathbf{b} \in \mathbb{R}^n$ as follows:

1. $\mathbf{x} = [1, ..., 1]^T$
2. $\mathbf{b} = A\mathbf{x}$

Solution:

Here is my code:

```
def one_vect(n: int):
    """
    Returns a column vector of size n x 1 filled with ones
    """
    return np.ones((n, 1))


def mat_vec_mult(A: np.ndarray, x: np.array):
    """
    Returns the multiplication of Ax. Has no protection on it
    so be careful.
    """
    return np.matmul(A, x)
```

Problem 5, part c

For each A generated as above, perform the following for $n = 5, 10, 15, ..., 500$.

1. Solve $A\mathbf{x} = \mathbf{b}$ using your program to get $\hat{\mathbf{x}}$. Note that in general the result computed by your program will not exactly be the true solution $\mathbf{x} = A^{-1}\mathbf{b}$ because of roundoff errors that occurred during computations.
2. Compute $\Delta b = A\hat{\mathbf{x}} - \mathbf{b}$ and record the values of $\|\mathbf{x} - \hat{\mathbf{x}}\|/\|\mathbf{x}\|, \kappa(A) and \kappa(A)\|\Delta\mathbf{b}\|/\|\mathbf{b}\|$ for $\|\cdot\| = \|\cdot\|_1, \|\cdot\|_2, \|\cdot\|_\infty$
3. Present everything for the $n = 5$ case but only tabulate the relevant trend for general $n > 5 in a graph$.

Solution: Here is my code (don't judge me, I only had like 2 days to work on this homework)

```
xhats = []
deltab = []

#x - xhat / mag x
xmag1 = []
xmag2 = []
xmaginf = []

#cond numbers
con1 = []
con2 = []
coninf = []

#max b * cond
relcon1 = []
relcon2 = []
relconinf = []

for inc in range(5, 505, 5):
    vec = one_vect(inc)
    ran = random_matrix(inc)
    hil = hilbert_matrix(inc)
    pas = pascal_matrix(inc)
    mag = magic_matrix(inc)

    ransol = np.matmul(np.linalg.pinv(ran), vec)
    hilsol = np.matmul(np.linalg.pinv(hil), vec)
    passol = np.matmul(np.linalg.pinv(pas), vec)
```

```
magsol = np.matmul(np.linalg.pinv(mag), vec)

dbran = np.matmul(ran, ransol) - mat_vec_mult(ran, vec)
dbhil = np.matmul(hil, hilsol) - mat_vec_mult(hil, vec)
dbpas = np.matmul(pas, passol) - mat_vec_mult(pas, vec)
dbmag = np.matmul(mag, magsol) - mat_vec_mult(mag, vec)

magran1 = np.linalg.norm(vec - ransol, ord= 1) / np.linalg.norm(vec, ord = 1)
maghil1 = np.linalg.norm(vec - hilsol, ord= 1) / np.linalg.norm(vec, ord = 1)
magpas1 = np.linalg.norm(vec - passol, ord= 1) / np.linalg.norm(vec, ord = 1)
magmag1 = np.linalg.norm(vec - magsol, ord= 1) / np.linalg.norm(vec, ord = 1)


magran2 = np.linalg.norm(vec - ransol, ord= 2) / np.linalg.norm(vec, ord = 2)
maghil2 = np.linalg.norm(vec - hilsol, ord= 2) / np.linalg.norm(vec, ord = 2)
magpas2 = np.linalg.norm(vec - passol, ord= 2) / np.linalg.norm(vec, ord = 2)
magmag2 = np.linalg.norm(vec - magsol, ord= 2) / np.linalg.norm(vec, ord = 2)


magraninf = np.linalg.norm(vec - ransol, ord= np.inf) / np.linalg.norm(vec, ord =
    np.inf)
maghilinf = np.linalg.norm(vec - hilsol, ord= np.inf) / np.linalg.norm(vec, ord =
    np.inf)
magpasinf = np.linalg.norm(vec - passol, ord= np.inf) / np.linalg.norm(vec, ord =
    np.inf)
magmaginf = np.linalg.norm(vec - magsol, ord= np.inf) / np.linalg.norm(vec, ord =
    np.inf)

rancon1 = np.linalg.cond(ran, p=1)
hilcon1 = np.linalg.cond(hil, p=1)
pascon1 = np.linalg.cond(pas, p=1)
magcon1 = np.linalg.cond(mag, p=1)

rancon2 = np.linalg.cond(ran, p=2)
hilcon2 = np.linalg.cond(hil, p=2)
pascon2 = np.linalg.cond(pas, p=2)
magcon2 = np.linalg.cond(mag, p=2)

ranconinf = np.linalg.cond(ran, p=np.inf)
hilconinf = np.linalg.cond(hil, p=np.inf)
pasconinf = np.linalg.cond(pas, p=np.inf)
magconinf = np.linalg.cond(mag, p=np.inf)
```

```
relconran1 = rancon1 * np.linalg.norm(dbran, ord = 1) / np.linalg.norm(vec, ord = 1)
relconhil1 = hilcon1 * np.linalg.norm(dbhil, ord = 1) / np.linalg.norm(vec, ord = 1)
relconpas1 = pascon1 * np.linalg.norm(dbpas, ord = 1) / np.linalg.norm(vec, ord = 1)
relconmag1 = magcon1 * np.linalg.norm(dbmag, ord = 1) / np.linalg.norm(vec, ord = 1)

relconran2 = rancon2 * np.linalg.norm(dbran, ord = 2) / np.linalg.norm(vec, ord = 2)
relconhil2 = hilcon2 * np.linalg.norm(dbhil, ord = 2) / np.linalg.norm(vec, ord = 2)
relconpas2 = pascon2 * np.linalg.norm(dbpas, ord = 2) / np.linalg.norm(vec, ord = 2)
relconmag2 = magcon2 * np.linalg.norm(dbmag, ord = 2) / np.linalg.norm(vec, ord = 2)

relconraninf = ranconinf * np.linalg.norm(dbran, ord = np.inf) / np.linalg.norm(vec,
    ord = np.inf)
relconhilinf = hilconinf * np.linalg.norm(dbhil, ord = np.inf) / np.linalg.norm(vec,
    ord = np.inf)
relconpasinf = pasconinf * np.linalg.norm(dbpas, ord = np.inf) / np.linalg.norm(vec,
    ord = np.inf)
relconmaginf = magconinf * np.linalg.norm(dbmag, ord = np.inf) / np.linalg.norm(vec,
    ord = np.inf)

temp11 = [relconraninf, relconhilinf, relconpasinf, relconmaginf]
temp10 = [relconran2, relconhil2, relconpas2, relconmag2]
temp9 = [relconran1, relconhil1, relconpas1, relconmag1]
temp8 = [ranconinf, hilconinf, pasconinf, magconinf]
temp7 = [rancon2, hilcon2, pascon2, magcon2]
temp6 = [rancon1, hilcon1, pascon1, magcon1]
temp5 = [magran1, maghil1, magpas1, magmag1]
temp4 = [magran2, maghil2, magpas2, magmag2]
temp3 = [magraninf, maghilinf, magpasinf, magmaginf]
temp2 = [dbran, dbhil, dbpas, dbmag]
temp1 = [ransol, hilsol, passol, magsol]

xhats.append(temp1)
deltab.append(temp2)
xmag1.append(temp5)
xmag2.append(temp4)
xmaginf.append(temp3)
con1.append(temp6)
con2.append(temp7)
coninf.append(temp8)
relcon1.append(temp9)
relcon2.append(temp10)
```

```
    relconinf.append(temp11)
```

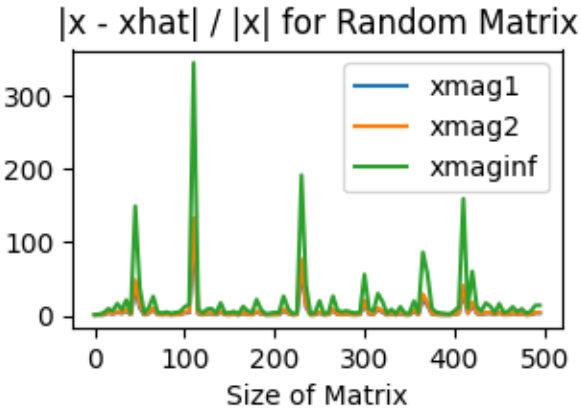
This disgusting abomination will record everything for us. Here is the $n = 5$ case wrapped in a numpy array:

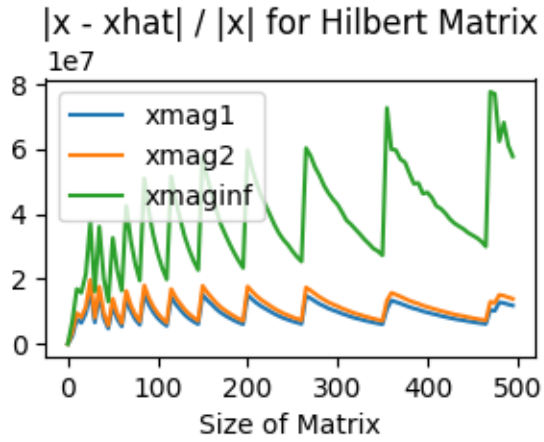
```
import pandas as pd
```

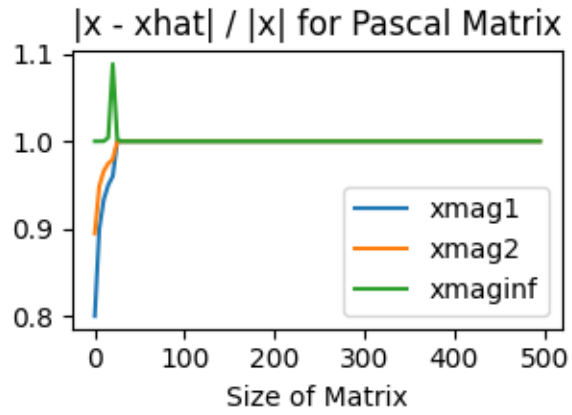
```
data = {
    'xmag1': xmag1[0],
    'xmag2': xmag2[0],
    'xmaginf': xmaginf[0],
    'condition1': con1[0],
    'condition2': con2[0],
    'conditioninf': coninf[0],
    'relcon1': relcon1[0],
    'relcon2': relcon2[0],
    'relconinf': relconinf[0]
}

row_names = ['random', 'hilbert', 'pascal', 'magic']

df = pd.DataFrame(data, index=row_names)
df
```

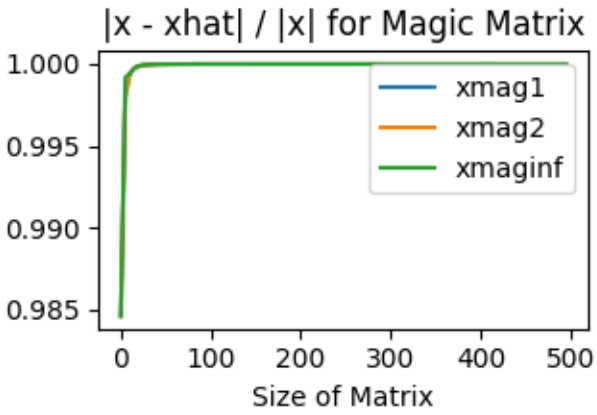
	xmag1	xmag2	xmaginf	condition1	condition2	conditioninf	relcon1	relcon2	relconinf
random	0.925476	1.030888	1.350415	12.553461	5.618668	15.132394	26.790832	13.463444	4.497819e+01
hilbert	500.800000	642.274085	1121.000000	943656.000006	476607.250242	943656.000006	414459.706669	296578.659617	1.211025e+06
pascal	0.800000	0.894427	1.000000	15624.000000	8517.524361	15624.000000	768700.800000	561820.968799	1.953000e+06
magic	0.984615	0.984615	0.984615	6.850000	5.461822	6.850000	438.400000	349.556639	4.384000e+02

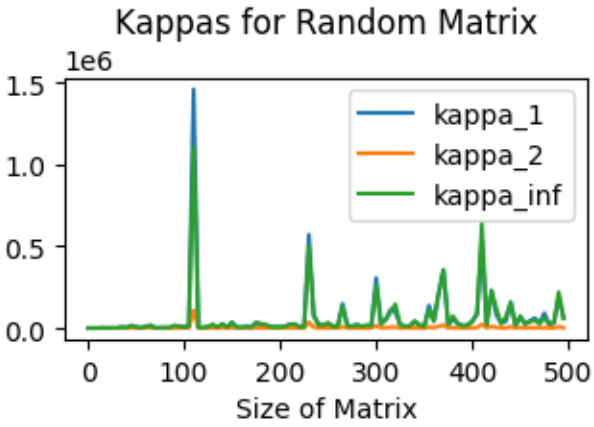
Here are the trends for the three different values to consider:

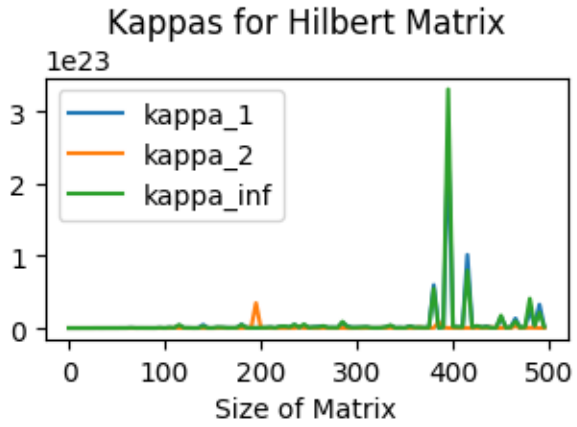

|x - xhat| / |x| for Random Matrix
xmag1
xmag2
xmaginf
Size of Matrix

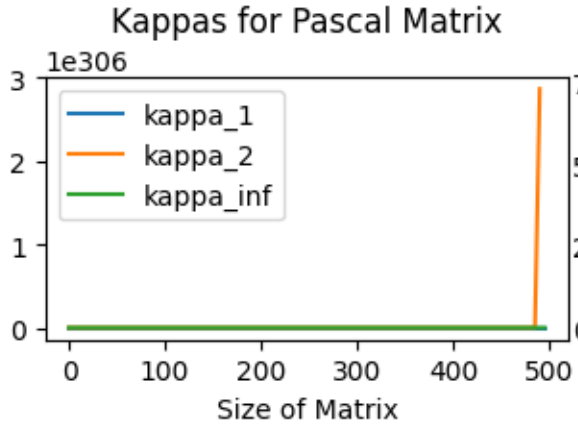

|x - xhat| / |x| for Hilbert Matrix
1e7
xmag1
xmag2
xmaginf
Size of Matrix

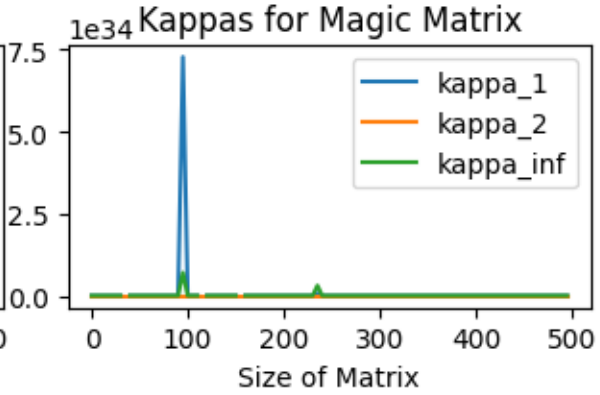

|x - xhat| / |x| for Pascal Matrix
xmag1
xmag2
xmaginf
Size of Matrix

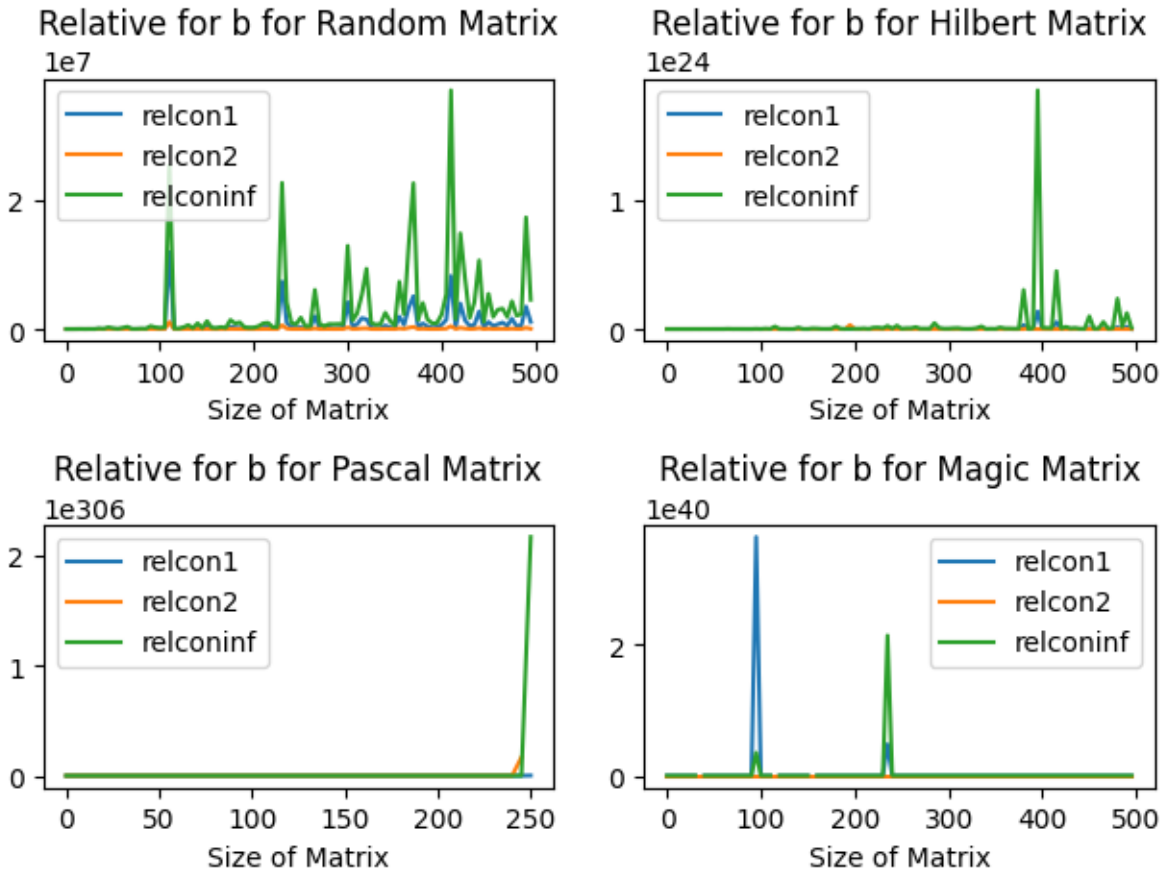

|x - xhat| / |x| for Magic Matrix
xmag1
xmag2
xmaginf
Size of Matrix


Kappas for Random Matrix
1e6
kappa_1
kappa_2
kappa_inf
Size of Matrix


Kappas for Hilbert Matrix
1e23
kappa_1
kappa_2
kappa_inf
Size of Matrix


Kappas for Pascal Matrix
1e306
kappa_1
kappa_2
kappa_inf
Size of Matrix


1e34
Kappas for Magic Matrix
kappa_1
kappa_2
kappa_inf
Size of Matrix



Problem 5, part d

Discuss an explain the effects of different choices of A, b, $\|\cdot\|$, and n have on the accuracy of the computed solution $\hat{\mathbf{x}}$.

Solution:

For the relative magnitudes for x_hat, we see wildly different plots for the random matrix, the Hilbert matrix, and Pascal Matrix. It looks like the Magic matrix has nearly the same shape as the pascal matrix. There seems to be some periodicity in the Hilbert matrix, and the relative magnitude difference is quite massive compared to other plots. It also seems that the mags stayed relatively the same shape for each matrix, aside from the Hilbert matrix.

For the condition number of these matrices, they seem relatively stable, aside from a few rogue points on the plot. Admittedly, I probably should have removed these points to get a far more interesting plot.

It seems as though the relative b suffered the same problem as the conditional number plots. The only "clear" plot we have is the random matrix, which seems to show random spikes for any size of the matrix.

Problem 5, part e

Instead of solving the linear system directly, compute A^{-1} and then define $\hat{\mathbf{x}} = A^{-1}\mathbf{b}$. Comment on the accuracy of the approach. Provide numerical evidence to support your conclusion.

Solution:

Since I was having issues solving the system for some sizes of the matrices (they were reported singular by numpy), I had to resort to computing the pseudo inverses for all systems, then solving it from there, as shown in my code. I would imagine my approach would incur more error in the result, as obtaining the pseudo inverses for these matrices are far more costly than solving the system directly.

Problem 5, part f

Write a program that computes the (1, 1)-entry of the matrix A^{-1} that does not involve computing A^{-1} , i.e., if $A^{-1}= b_{ij}$, you want the value b_{11} but you are not allowed to compute A^{-1}.

Solution: I can imagine a solution would be to use the fact that the inverse A^{-1} can be computed element wise, as the b_{11} element would be a_{11} times the determinant of the minor of a_{11}. There is no restriction on using determinants, so this solution is valid. Here is my code:

```
def det_11 (A: np.array):
    """
    Computes the a_11 entry of the inverse of A, assuming A meets the
    conditions for being invertible.
    This will take advantage of an element wise defintion of the inverse.
    """
    minor = A[1:, 1:]
    detminor = np.linalg.det(minor)
    return float(A[0][0] / detminor)
```

Problem 6

Let $A \in \mathbb{R}^{n\times n}$ be nonsingular and let $\mathbf{0} \neq \mathbf{b} \in \mathbb{R}^n$. Let $\mathbf{x} = A^{-1}\mathbf{b} \in \mathbb{R}^n$. In the following, $\Delta A \in \mathbb{R}^{n\times n}$ and $\Delta\mathbf{b} \in \mathbb{R}^n$ are some arbitrary matrix and vector. We assume that the norm on A satisfies $\|A\mathbf{x}\| \leq \|A\|\|\mathbf{x}\|$ for all $A \in \mathbb{R}^{n\times n}$ and all $\mathbf{x} \in \mathbb{R}^n$.

Problem 6, part a

Show that if $\Delta A \in \mathbb{R}^{n\times n}$ is any matrix satisfying

$$\frac{\|\Delta A\|}{\|A\|} < \frac{1}{\kappa(A)},$$

the $\Delta A + A$ must be nonsingular.

Solution:

Suppose false, that is, $\Delta A + A$ is singular. This means that there exists some nonzero $\mathbf{v} \in \mathbb{R}^n$ where $(A + \Delta A)\mathbf{v} = 0$. This would then mean $\|(A + \Delta A)\mathbf{v}\| = 0$. By the property of the norm given, this means $0 \leq \|A + \Delta A\|\|\mathbf{v}\|$, which means $\|A + \Delta A\| \geq 0$. Since this norm must satisfy the properties for being a norm, this means the triangle inequality can be applied, thus $0 \leq \|A\| + \|\Delta A\|$, so $-\|A\| \leq \|\Delta A\|$. This can then be plugged into the given inequality to get

$$\frac{-\|A\|}{\|A\|} \leq \frac{\|\Delta A\|}{\|A\|} < \frac{1}{\kappa(A)}$$

After reciprocating both sides, we get $-1 > \kappa(A)$, which should never happen, since $\kappa(A)$ is the product of two positive numbers. Thus $A + \Delta A$ cannot be singular.

Problem 6, part b

Suppose $(A + \Delta A)(\mathbf{x} + \Delta\mathbf{x}) = \mathbf{b}$ and $\hat{\mathbf{x}} = \mathbf{x} + \Delta\mathbf{x}$. Show that

$$\frac{\|\Delta\mathbf{x}\|}{\|\hat{\mathbf{x}}\|} \leq \kappa(A)\frac{\|\Delta A\|}{\|A\|}$$

Solution:

We are given $(A+\Delta A)(\mathbf{x}+\Delta\mathbf{x}) = \mathbf{b} = A\mathbf{x}$. Cancelling out the $A\mathbf{x}$ on each side gives us $A\Delta\mathbf{x}+\Delta A(\mathbf{x}+\Delta\mathbf{x} = 0/$ Thus the following steps can be justified:

$$-A\Delta\mathbf{x} = \Delta A(\mathbf{x} + \Delta\mathbf{x}) \qquad \text{(Rearranging.)}$$

$$\implies -\Delta\mathbf{x} = A^{-1}\Delta A\hat{\mathbf{x}} \qquad \text{(Inverting } A\text{.)}$$

$$\implies \|\Delta\mathbf{x}\| = \left\|A^{-1}\Delta A\hat{\mathbf{x}}\right\| \qquad \text{(Negative lost in norm.)}$$

$$\|\Delta\mathbf{x}\| \leq \left\|A^{-1}\Delta A\right\|\|\hat{\mathbf{x}}\| \qquad \text{(Submultiplicating.)}$$

$$\frac{\|\Delta\mathbf{x}\|}{\|\hat{\mathbf{x}}\|} \leq \left\|A^{-1}\Delta A\right\| \qquad \text{(Simplifying.)}$$

$$\implies \frac{\|\Delta\mathbf{x}\|}{\|\hat{\mathbf{x}}\|} \leq \left\|A^{-1}\right\|\|\Delta A\| \qquad \text{(Submultplicating.)}$$

$$\iff \frac{\|\Delta\mathbf{x}\|}{\|\hat{\mathbf{x}}\|} \leq \left\|A^{-1}\right\|\|A\|\frac{\|\Delta A\|}{\|A\|} \qquad \text{(Multiplying by 1.)}$$

$$\therefore \frac{\|\Delta\mathbf{x}\|}{\|\hat{\mathbf{x}}\|} \leq \kappa(A)\frac{\|\Delta A\|}{\|A\|} \qquad \text{(Definition of } \kappa(A)\text{.)}$$

Problem 6, part c

Suppose that $(A + \Delta A)(\mathbf{x} + \Delta\mathbf{x}) = \mathbf{b}$ and $\hat{\mathbf{x}} = \mathbf{x} + \Delta\mathbf{x}$ and the statement in the first part is satisfied. Show that

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \frac{\kappa(A)\frac{\|\Delta A\|}{\|A\|}}{1 - \kappa(A)\frac{\|\Delta A\|}{\|A\|}}$$

Solution:

We will follow the given steps:

1. Note that $(A + \Delta A)(\mathbf{x} + \Delta\mathbf{x}) = \mathbf{b} = A\mathbf{x}$. We can see that $\Delta A(\mathbf{x} + \Delta\mathbf{x}) = -A\Delta\mathbf{x}$, so $\Delta\mathbf{x} = -A^{-1}\Delta A\hat{\mathbf{x}}$. The second part can be found by bringing over the $\hat{\mathbf{x}}$ on the last part, to get $\|\Delta\mathbf{x}\| \leq \kappa(A)\frac{\|\Delta A\|}{\|A\|}\|\hat{\mathbf{x}}\|$. By the triangle inequality, this is says $\|\Delta\mathbf{x}\| \leq \kappa(A)\frac{\|\Delta A\|}{\|A\|}(\|\mathbf{x}\| + \|\Delta\mathbf{x}\|)$.

2. Rearranging, we see that $(1 - \kappa(A)(\|\Delta A\|/\|A\|))\|\mathbf{x}\| \leq \kappa(A)(\|\Delta A\|/\|A\|)\|\mathbf{x}\|$. Swapping the $\|\mathbf{x}\|$ and the factor of $\|\Delta\mathbf{x}\|$, we get

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \frac{\kappa(A)\frac{\|\Delta A\|}{\|A\|}}{1 - \kappa(A)\frac{\|\Delta A\|}{\|A\|}}$$

which is what we wanted to find.

Problem 6, part d

Suppose $(A + \Delta A)\hat{\mathbf{x}} = \mathbf{b} + \Delta\mathbf{b}$, where $\hat{\mathbf{b}} \neq 0$ and $\hat{\mathbf{x}} \neq 0$. Show that

$$\frac{\|\Delta\mathbf{x}\|}{\|\hat{\mathbf{x}}\|} \leq \kappa(A)\Big(\frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta\mathbf{b}\|}{\|\hat{\mathbf{b}}\|} + \frac{\|\Delta A\|}{\|A\|}\frac{\|\Delta\mathbf{b}\|}{\|\hat{\mathbf{b}}\|}\Big)$$

Solution:

We will follow the given steps:

1. Since $(A+\Delta A)\hat{\mathbf{x}} = \mathbf{b}+\Delta\mathbf{b}$, and $A\mathbf{x} = \mathbf{b}$, we can eliminate the $A\mathbf{x} = \mathbf{b}$ on both sides to get $\Delta A\hat{\mathbf{x}}+A\Delta\mathbf{x} = \Delta\mathbf{b}$. Then $\Delta\mathbf{x} = A^{-1}(\Delta\mathbf{b} - \Delta A\hat{\mathbf{x}})$. The following steps can then be taken

---

$$\|\Delta\mathbf{x}\| = \kappa(A)\frac{\|\Delta\mathbf{b} - \Delta A\hat{\mathbf{x}}\|}{\|A\|} \qquad \text{(Multiplying through by } A.\text{)}$$

$$\frac{\|\Delta\mathbf{x}\|}{\|\hat{\mathbf{x}}\|} = \kappa(A)\frac{\|\Delta\mathbf{b} - \Delta A\hat{\mathbf{x}}\|}{\|A\|\|\hat{\mathbf{x}}\|} \qquad \text{(Dividing through by } \|\hat{\mathbf{x}}\|.\text{)}$$

$$\implies \frac{\|\Delta\mathbf{x}\|}{\|\hat{\mathbf{x}}\|} \leq \kappa(A)\left(\frac{\|\Delta\mathbf{b}\|}{\|A\|\|\hat{\mathbf{x}}\|} + \frac{\|\Delta A\|}{\|A\|}\right) \qquad \text{(Triangle inequality and simplifying.)} \qquad (3)$$

---

2. By $(A + \Delta A)\hat{\mathbf{x}} = \mathbf{b} + \Delta\mathbf{b}$, we can write the following:

$$(A + \Delta A)\hat{\mathbf{x}} = \hat{\mathbf{b}} \qquad \text{(Given.)}$$

$$\|(A + \Delta A)\hat{\mathbf{x}}\| = \|\hat{\mathbf{b}}\| \qquad \text{(Taking norm.)}$$

$$\|\hat{\mathbf{b}}\| \leq (\|A\| + \|\Delta A\|)\hat{\mathbf{x}} \qquad \text{(Triangle Inequality.)}$$

$$\frac{1}{\|\hat{\mathbf{x}}\|} \leq \frac{\|A\| + \|\Delta A\|}{\|\hat{\mathbf{b}}\|} \qquad \text{(Rearranging.)} \qquad (4)$$

---

3. Then we can take the result from 4 and plug it into the right hand side of 3.

$$\frac{\|\Delta \mathbf{x}\|}{\|\hat{\mathbf{x}}\|} \leq \kappa(A)\left(\frac{\|\Delta \mathbf{b}\|}{\|A\|\|\hat{\mathbf{x}}\|} + \frac{\|\Delta A\|}{\|A\|}\right) \qquad \text{(Given, from 3.)}$$

$$\leq \kappa(A)\left(\frac{\|\Delta \mathbf{b}\|(\|A\| + \|\Delta A\|)}{\|\hat{\mathbf{b}}\|\|A\|} + \frac{\|\Delta A\|}{\|A\|}\right) \qquad \text{(From 4.)}$$

$$= \kappa(A)\left(\frac{\|\Delta \mathbf{b}\|}{\|\hat{\mathbf{b}}\|} + \frac{\|\Delta \mathbf{b}\|\|\Delta A\|}{\|\hat{\mathbf{b}}\|\|A\|} + \frac{\|\Delta A\|}{\|A\|}\right) \qquad \text{(Simplifying.)}$$

Thus, the statement above has been shown.

Problem 6, part e

Suppose $(A + \Delta A)\hat{\mathbf{x}} = \mathbf{b} + \Delta \mathbf{b}$ where $\hat{\mathbf{b}} \neq 0$ and $\hat{\mathbf{x}} \neq 0$ and the statement in part a is satisfied. Use the same ideas in part b to deduce that

$$\frac{\|\Delta \mathbf{x}\|}{\|\mathbf{x}\|} \leq \frac{\kappa(A)\left(\frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta \mathbf{b}\|}{\|\mathbf{b}\|}\right)}{1 - \kappa(A)\frac{\|\Delta A\|}{\|A\|}}.$$

Solution:

To save some time, we can note that we will have the same $\Delta \mathbf{x}$ as in the last part, so we can write

$$\Delta \mathbf{x} = -(A^{-1}\Delta A\hat{\mathbf{x}} - A^{-1}\Delta \mathbf{b})$$

The following steps can then be justified:

$$
\begin{aligned}
\|\Delta\mathbf{x}\| &= \left\|A^{-1}\Delta A\hat{\mathbf{x}} - \Delta\mathbf{b}\right\| && \text{(Taking Norm.)}\\
&= \frac{\left\|A^{-1}\Delta A\hat{\mathbf{x}} - \Delta\mathbf{b}\right\|}{\|A\|}\|A\| && \text{(Multiplying through by } \|A\|.)\\
&\leq \left\|A^{-1}\right\|\frac{\Delta A\hat{\mathbf{x}} - \Delta\mathbf{b}}{\|A\|}\|A\| && \text{(Submultiplicity.)}\\
&\leq \kappa(A)\left(\frac{\|\Delta A\hat{\mathbf{x}}\|}{\|A\|} + \frac{\|\Delta\mathbf{b}\|}{\|A\|}\right) && \text{(Triangle Inequality and definition of } \kappa(A).)\\
&\leq \kappa(A)\left(\frac{\|\Delta A\|}{\|A\|}\|\mathbf{x}\| + \frac{\|\Delta A\|}{\|A\|}\|\Delta\mathbf{x}\| + \frac{\|\Delta\mathbf{b}\|}{\|A\|}\right) && \text{(Triangle Inequality.)}\\
\implies \left(1 - \kappa(A)\frac{\|\Delta A\|}{\|A\|}\right)\|\Delta\mathbf{x}\| &\leq \kappa(A)\left(\frac{\|\Delta A\|}{\|A\|}\|\mathbf{x}\| + \frac{\|\Delta\mathbf{b}\|}{\|A\|}\right) && \text{(Grouping.)}\\
\|\Delta\mathbf{x}\| &\leq \frac{\kappa(A)\left(\frac{\|\Delta A\|}{\|A\|}\|\mathbf{x}\| + \frac{\|\Delta\mathbf{b}\|}{\|A\|}\right)}{1 - \kappa(A)\frac{\|\Delta A\|}{\|A\|}} && \text{(Rearranging.)}\\
\|\Delta\mathbf{x}\| &\leq \frac{\kappa(A)\left(\frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta\mathbf{b}\|}{\|A\|\|\mathbf{x}\|}\right)\|\mathbf{x}\|}{1 - \kappa(A)\frac{\|\Delta A\|}{\|A\|}} && \text{(Rearranging.)}\\
\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} &\leq \frac{\kappa(A)\left(\frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta\mathbf{b}\|}{\|A\|\|\mathbf{x}\|}\right)}{1 - \kappa(A)\frac{\|\Delta A\|}{\|A\|}} && \text{(Rearranging.)}\\
\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} &\leq \frac{\kappa(A)\left(\frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta\mathbf{b}\|}{\|A\mathbf{x}\|}\right)}{1 - \kappa(A)\frac{\|\Delta A\|}{\|A\|}} && (\|A\mathbf{x}\| \leq \|A\|\|\mathbf{x}\|.)\\
\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} &\leq \frac{\kappa(A)\left(\frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|}\right)}{1 - \kappa(A)\frac{\|\Delta A\|}{\|A\|}} && (A\mathbf{x} = \mathbf{b}.)
\end{aligned}
$$

Thus, the statement has been shown.