4.830. **True/false questions about linear algebra.** Determine whether each of the following statements is true or false. In each case, give either a proof or a counterexample.

a) If $Q$ has orthonormal columns, then $\|Q^Tw\| \leq \|w\|$ for all vectors $w$.

b) Suppose $A \in \mathbb{R}^{m \times p}$ and $B \in \mathbb{R}^{m \times q}$. If $\text{null}(A) = \{0\}$ and $\text{range}(A) \subset \text{range}(B)$, then $p \leq q$.

c) If $V = [V_1 \ V_2]$ is invertible and $\text{range}(V_1) = \text{null}(A)$, then $\text{null}(AV_2) = \{0\}$.

d) If $\text{rank}([A \ B]) = \text{rank}(A) = \text{rank}(B)$, then $\text{range}(A) = \text{range}(B)$.

e) Suppose $A \in \mathbb{R}^{m \times n}$. Then, $x \in \text{null}(A^T)$ if and only if $x \notin \text{range}(A)$.

f) Suppose $A$ is invertible. Then, $AB$ is not full rank if and only if $B$ is not full rank.

g) If $A$ is not full rank, then there is a nonzero vector $x$ such that $Ax = 0$.

6.1240. **Iteratively reweighted least squares for 1-norm approximation.** In an ordinary least squares problem, we are given $A \in \mathbb{R}^{m \times n}$ (skinny and full rank) and $y \in \mathbb{R}^m$, and we choose $x \in \mathbb{R}^n$ in order to minimize

$$
\|Ax - y\|^2 = \sum_{i=1}^m (\tilde{a}_i^T x - y_i)^2.
$$

Note that the penalty that we assign to a measurement error does not depend on the sensor from which the measurement was taken. However, this is not always the right thing to do: if we believe that one sensor is more accurate than another, we might want to assign a larger penalty to an error in the measurement from the more accurate sensor. We can account for differences in the accuracies of our sensors by assigning sensor $i$ a weight $w_i > 0$, and then minimizing

$$
\sum_{i=1}^m w_i(\tilde{a}_i^T x - y_i)^2.
$$

By giving larger weights to more accurate sensors, we can account for differences in the precision of our sensors.

a) **Weighted least squares.** Explain how to choose $x$ in order to minimize

$$
\sum_{i=1}^m w_i(\tilde{a}_i^T x - y_i)^2,
$$

where the weights $w_1, \ldots, w_m > 0$ are given.
b) **Iteratively reweighted least squares for $\ell_1$-norm approximation.** Consider a cost function of the form

$$
\sum_{i=1}^{m} w_i(x)(\tilde{a}_i^T x - y_i)^2.
$$

(1)

One heuristic for minimizing a cost function of the form given in (1) is *iteratively reweighted least squares*, which works as follows. First, we choose an initial point $x^{(0)} \in \mathbb{R}^n$. Then, we generate a sequence of points $x^{(1)}, x^{(2)}, \ldots \in \mathbb{R}^n$ by choosing $x^{(k+1)}$ in order to minimize

$$
\sum_{i=1}^{m} w_i(x^{(k)})(\tilde{a}_i^T x^{(k+1)} - y_i)^2.
$$

Each step of this algorithm involves updating our weights, and solving a weighted least squares problem. Suppose we want to use this method to solve minimize the $\ell_1$-norm approximation error, which is defined to be

$$
\|Ax - y\|_1 = \sum_{i=1}^{m} |\tilde{a}_i^T x - y_i|,
$$

where the matrix $A \in \mathbb{R}^{m \times n}$ and the vector $y \in \mathbb{R}^m$ are given. How should we choose the weights $w_i(x)$ to make the cost function in (1) equal to the $\ell_1$-norm approximation error?

c) **Numerical example.** The file *l1_irwls_data.json* contains data $(t_1, y_1), \ldots, (t_m, y_m)$. We want to fit an affine model to this data:

$$
y_i = x_1 + x_2 t_i, \quad i = 1, \ldots, m.
$$

Choose $x^{(0)}$ to be the vector of least-squares parameter estimates: that is, choose $x^{(0)}$ in order to minimize

$$
\sum_{i=1}^{m} ((x_1^{(0)} + x_2^{(0)} t_i) - y_i)^2.
$$

Generate $x^{(1)}, x^{(2)}, \ldots$ using iteratively reweighted least squares for $\ell_1$-norm approximation. You can stop generating iterates when $\|x^{(k+1)} - x^{(k)}\| < 10^{-6}$. Report your values of $x^{(0)}$ and the final $x^{(k)}$ in your sequence of points. Draw a scatterplot of the data points $(t_i, y_i)$. Add the fitted lines corresponding to $x^{(0)}$ and the final $x^{(k)}$ to your scatterplot. What do you observe?

**Remark.** Suppose we fit the least-squares line to some data. Then, a point that is very far from the least-squares line may be an outlier: that is, a point that does not seem to follow the same model as the rest of the data. Because such points may not follow the same model as the rest of data, it may make sense to give such points less weight. This idea is the intuition behind iteratively reweighted least squares for $\ell_1$-norm approximation.
8.160. **Designing an equalizer for backwards-compatible wireless transceivers.** You want to design the equalizer for a new line of wireless handheld transceivers (more commonly called walkie-talkies). The transmitter for the new line of transceivers has already been designed (and cannot be changed) – if the input signal is \( x \in \mathbb{R}^n \), then the transmitted signal is \( y = A_{\text{new}}x \in \mathbb{R}^m \), where \( A_{\text{new}} \in \mathbb{R}^{m \times n} \) is known. An equalizer for \( A_{\text{new}} \) is a matrix \( B \in \mathbb{R}^{n \times m} \) such that \( By = x \) for every \( x \in \mathbb{R}^n \).

The new line of transceivers will replace an older model. Given an input signal \( x \in \mathbb{R}^n \), the old line of transceivers transmit a signal \( y_{\text{old}} = A_{\text{old}}x \in \mathbb{R}^m \), where \( A_{\text{old}} \in \mathbb{R}^{m \times n} \) is known. In addition to providing exact equalization for the new line of transceivers, you want your equalizer to be able to at least partially equalize signals transmitted using the old line of transceivers. In other words, to the extent that it is possible, you want the new line of transceivers to be backwards compatible with the old line of transceivers.

a) Explain how to find an equalizer \( B \) that minimizes

\[
J = \|BA_{\text{old}} - I\|_F^2 = \sum_{i=1}^n \sum_{j=1}^n (BA_{\text{old}} - I)_{ij}^2
\]

among all \( B \) that exactly equalize \( A_{\text{new}} \). Such a \( B \) is an exact equalizer for \( A_{\text{new}} \), and an approximate equalizer for \( A_{\text{old}} \). State any assumptions that are needed for your method to work.

b) The file `backwards_compatible_transceiver_data.json` defines the following variables.

- \( A_{\text{new}} \), the \( m \times n \) matrix that describes the transmitter used in the new line of transceivers
- \( A_{\text{old}} \), the \( m \times n \) matrix that describes the transmitter used in the old line of transceivers
- \( x \), a vector of length \( n \) that serves as an example input signal

Apply your method to this example data. Report the optimal value of \( J \). The pseudoinverse \( A_{\text{new}}^\dagger \) is another exact equalizer for \( A_{\text{new}} \). Compare the optimal value of \( J \), and the value of \( J \) achieved by \( A_{\text{new}}^\dagger \).

c) The example signal \( x \) defined in the data file is a binary signal. Form the signal \( y_{\text{old}} = A_{\text{old}}x \) transmitted by the old line of transceivers, and construct an estimate of \( x \) by equalizing \( y_{\text{old}} \) using \( B \), and then rounding the result to a binary signal. More concretely, compute the estimate \( \hat{x} \in \mathbb{R}^n \), where

\[
\hat{x}_i = \begin{cases} 
1 & (By_{\text{old}})_i > \frac{1}{2}, \\
0 & \text{otherwise.}
\end{cases}
\]

Report the bit error rate of your estimate, which is defined as

\[
\frac{1}{n} \sum_{i=1}^n I(x_i \neq \hat{x}_i),
\]
where $I(x_i \neq \hat{x}_i)$ is an indicator function:

$$I(x_i \neq \hat{x}_i) = \begin{cases} 
1 & x_i \neq \hat{x}_i, \\
0 & \text{otherwise}
\end{cases}$$

Similarly, report the bit error rate if $A_{\text{new}}^\dagger$ is used as the equalizer.

**8.1460. Filling-in missing data.** In this problem we have a signal, $y_i \in \mathbb{R}$ for $i = 1, \ldots, n$, which we view as $y \in \mathbb{R}^n$. We will have $n = 100$. The signal $y$ comes from measurements of a physical system, and so $y_{i+1}$ is measured a short time interval after $y_i$. Unfortunately, during the data acquisition process some of the data was lost and so the signal we have has gaps in it. Specifically, we have a known set $K \subset \mathbb{Z}$ and we know $y_i$ only for values $i \in K$.

The data for this problem is in the file `missing.json`. The supplied vector `known` contains the list of known points $K$, and the vector `yknown` is the list of values of $y$ at the points in $K$. The length of `yknown` is therefore $|K|$.

a) For a signal $z \in \mathbb{R}^n$, we define the discrete derivative $z_{\text{der}} \in \mathbb{R}^{n-1}$ by

$$z_{i,\text{der}} = z_{i+1} - z_i \quad \text{for } i = 1, \ldots, n-1$$

Find the matrix $G$ such that $z_{\text{der}} = Gz$

b) Our first approach will be to find the signal $z$ which minimizes $\|z_{\text{der}}\|$ and satisfies

$$z_i = y_i \quad \text{if } i \in K$$

Give a method finding the optimal $z$.

c) Find the optimal $z$ in the previous part and plot $z_i$ against $i$. Be sure to plot the points $(i, z_i)$, not just a line joining them.

d) One way to do a better job at filling in the missing data is to put additional criteria on our estimate. Here we will do this by additionally penalizing the second derivative of $z$. Define the discrete second derivative $z_{\text{hes}} \in \mathbb{R}^{n-2}$ by

$$z_{i,\text{hes}} = z_{i+2} - 2z_{i+1} + z_i \quad \text{for } i = 1, \ldots, n-2$$

Find the matrix $H$ such that $z_{\text{hes}} = Hz$

e) Define the two objective functions

$$J_1 = \|Gz\|^2 \quad J_2 = \|Hz\|^2$$

We would like to find the signal $z$ that minimizes

$$J_1 + \mu J_2$$

and satisfies

$$z_i = y_i \quad \text{if } i \in K$$

Give a method for finding the optimal $z$.

f) Plot the trade-off curve of $J_2$ (on the vertical axis) versus $J_1$ (on the horizontal axis).

Give the interpretation of the endpoints of this curve.

g) Find the optimal $z$ for the three different cases $\mu = 5, 20, 100$. 

4
9.1360. **A simple population model.** We consider a certain population of fish (say) each (yearly) season. $x(t) \in \mathbb{R}^3$ will describe the population of fish at year $t \in \mathbb{Z}$, as follows:

- $x_1(t)$ denotes the number of fish less than one year old
- $x_2(t)$ denotes the number of fish between one and two years old
- $x_3(t)$ denotes the number of fish between two and three years

(We will ignore the fact that these numbers are integers.) The population evolves from year $t$ to year $t + 1$ as follows.

- The number of fish less than one year old in the next year $(t + 1)$ is equal to the total number of offspring born during the current year. Fish that are less than one year old in the current year $(t)$ bear no offspring. Fish that are between one and two years old in the current year $(t)$ bear an average of 2 offspring each. Fish that are between two and three years old in the current year $(t)$ bear an average of 1 offspring each.
- 40% of the fish less than one year old in the current year $(t)$ die; the remaining 60% live on to be between one and two years old in the next year $(t + 1)$.
- 30% of the one-to-two year old fish in the current year die, and 70% live on to be two-to-three year old fish in the next year.
- All of the two-to-three year old fish in the current year die.

Express the population dynamics as an autonomous linear system with state $x(t)$, i.e., in the form $x(t + 1) = Ax(t)$. **Remark:** this example is silly, but more sophisticated population dynamics models are very useful and widely used.