AE 370 — solutions to homework #1

1. Consider the 'first-order difference' matrix $\boldsymbol{A} \in \mathbb{R}^{n-1 \times n}$ defined as

$$\boldsymbol{A} = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 & 1 & 0 & 0 \\ 0 & 0 & \cdots & 0 & -1 & 1 & 0 \\ 0 & 0 & \cdots & 0 & 0 & -1 & 1 \end{bmatrix} \tag{1}$$

We will encounter this matrix (and ones like it) later in class, and we already have tools to say a lot about the properties of this matrix. You will demonstrate that below.

(a) Does the problem $\boldsymbol{Ax} = \boldsymbol{b}$, where $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{b} \in \mathbb{R}^{n-1}$, have a solution? How do you know? *Hint*: it may be helpful to inspect this matrix for $n = 3$ or 4 and determine the answer for that case before considering the general n case.
Let's consider the case where $n = 4$, for which $\boldsymbol{A}$ looks like

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

A solution does exist, since the columns of $\boldsymbol{A}$ span all of $\mathbb{R}^3$; *i.e.*, $\mathcal{R}(\boldsymbol{A}) = \mathbb{R}^{n-1}$. We can see this by noting that any $\boldsymbol{b} = [b_1, b_2, b_3]^T$ can be written as $b_1\boldsymbol{a}_2 + (b_2 + b_1)\boldsymbol{a}_3 + (b_3 + b_2 + b_1)\boldsymbol{a}_4$ (Try it for yourself!).
Now, we turn to the general case. It is clear from the $n = 4$ example that for any n, $\mathcal{R}(\boldsymbol{A}) = \mathbb{R}^{n-1}$, as any $\boldsymbol{b}$ will continue to be expressable recursively in terms of the last $n-1$ columns of $\boldsymbol{A}$. Thus, a solution exists for any (finite) value of n.

(b) If there is a solution, is it unique? Why or why not? If it is not unique, what vector can be added to the solution without affecting the solution. *Hint*: is there a vector that gets annihilated by taking differences of its components and their neighbors?)?
No, the solution is not unique because $\boldsymbol{A}$ has a nontrivial null space: multiplying $\boldsymbol{A}$ by any constant vector will result in zero. Thus, a constant vector with each entry equal to $\alpha(\in \mathbb{R})$ may be added to the solution without affecting the solution.

2. We derived the linear system for approximating a solution to $\boldsymbol{Ax} = \boldsymbol{b}$ using the fact that if $\boldsymbol{y}^* \in \mathcal{R}(\boldsymbol{A})$, then $(\boldsymbol{y}^* - \boldsymbol{b}, \boldsymbol{r}) = 0$ for *any* $\boldsymbol{r} \in \mathcal{R}(\boldsymbol{A})$.

(a) Show that this is indeed true. *Hint*: any $\boldsymbol{w} \in \mathcal{R}(\boldsymbol{A})$ can be written as $\boldsymbol{w} = \boldsymbol{y}^* + \beta\boldsymbol{r}$ for some $\beta \in \mathbb{R}$ and $\boldsymbol{r} \in \mathcal{R}(\boldsymbol{A})$. Use this in conjunction with the fact that $||\boldsymbol{y}^* - \boldsymbol{b}|| \leq ||\boldsymbol{w} - \boldsymbol{b}||$.
Letting $\boldsymbol{w} \in \mathcal{R}(\boldsymbol{A})$ be written as as $\boldsymbol{w} = \boldsymbol{y}^* + \beta\boldsymbol{r}$, the inequality becomes $||\boldsymbol{y}^* - \boldsymbol{b}|| \leq ||(\boldsymbol{y}^* + \beta\boldsymbol{r}) - \boldsymbol{b}||$. Writing the norms in terms of the inner product, we get

$$\begin{aligned} (\boldsymbol{y}^* - \boldsymbol{b}, \boldsymbol{y}^* - \boldsymbol{b}) &\leq (\boldsymbol{y}^* + \beta\boldsymbol{r} - \boldsymbol{b}, \boldsymbol{y}^* + \beta\boldsymbol{r} - \boldsymbol{b},) && (3) \\ &= ((\boldsymbol{y}^* - \boldsymbol{b}) + \beta\boldsymbol{r}, (\boldsymbol{y}^* - \boldsymbol{b}) + \beta\boldsymbol{r},) && (4) \\ &= (\boldsymbol{y}^* - \boldsymbol{b}, \boldsymbol{y}^* - \boldsymbol{b}) + 2\beta(\boldsymbol{r}, (\boldsymbol{y}^* - b)) + \beta^2(\boldsymbol{r}, \boldsymbol{r}) && (5) \end{aligned}$$

Subtracting $(\boldsymbol{y}^* - \boldsymbol{b}, \boldsymbol{y}^* - \boldsymbol{b})$ from both sides of the equation gives

$$2\beta(\boldsymbol{r}, (\boldsymbol{y}^* - b)) + \beta^2(\boldsymbol{r}, \boldsymbol{r}) \geq 0 \tag{6}$$

The second term always satisfies the inequality, so the only way the inequality holds *for any* $\boldsymbol{w}$ (*i.e.*, for any β and $\boldsymbol{r}$) is if $(\boldsymbol{r}, (\boldsymbol{y}^* - b)) = 0$.

(b) Consider the matrix $\hat{\boldsymbol{A}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. Determine the best approximation to the problem $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ onto $\mathcal{R}(A)$ for $\boldsymbol{b} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, and demonstrate the this solution indeed satisfies the required orthogonality relation.

A basis for $\mathcal{R}(\hat{\boldsymbol{A}})$ is $\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. We look for the optimal $\boldsymbol{y}^* = \alpha_1 \boldsymbol{v}_1$ that best approximates $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$. To solve for α_1 we use the orthogonality relation we just derived:

$$\alpha_1(\boldsymbol{v}_1, \boldsymbol{v}_1) = (\boldsymbol{v}_1, \boldsymbol{b}) \tag{7}$$

$$\implies \qquad \alpha_1 = \frac{(\boldsymbol{v}_1, \boldsymbol{b})}{(\boldsymbol{v}_1, \boldsymbol{v}_1)} = \frac{3}{5} \tag{8}$$

The best approximation is then $\boldsymbol{y}^* = 3/5 \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

As we demonstrated must be true in a), we see that

$$(\boldsymbol{y}^* - \boldsymbol{b}, \eta \boldsymbol{v}_1) = \left(\frac{3}{5} \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \eta \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right) \tag{9}$$

$$= \left(\begin{bmatrix} -2/5 \\ 1/5 \end{bmatrix}, \eta \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right) \tag{10}$$

$$= -2\eta/5 + 2\eta/5 \tag{11}$$

$$= 0 \quad \text{for any } \eta \in \mathbb{R} \tag{12}$$

3. An alien spacecraft has crash landed on Earth and biologists were able to measure concentrations of an unknown chemical as a function of radius from the crash-sight. The values of concentration ($\boldsymbol{y}$) and radius ($\boldsymbol{x}$) are as follows:

$$\boldsymbol{x} = \begin{bmatrix} 0.02 \\ 0.10 \\ 0.32 \\ 0.66 \\ 0.94 \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} 1.93 \\ 2.60 \\ 3.03 \\ 3.97 \\ 4.89 \end{bmatrix} \tag{13}$$

The biologists are turning to you to ask you to determine a line that best fits this data. That is, they want a best approximation to the problem

$$\boldsymbol{y} = a\boldsymbol{x} + b\boldsymbol{e} \tag{14}$$

where $a, b \in \mathbb{R}$ are the constants to be solved for and $\boldsymbol{e} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$. Can you help Planet Earth?!

Once you have solved for a, b, provide a plot that includes both the raw data (with circles or some comparable marker) as well as the best fit line through the data.

We are looking for a $\boldsymbol{y}$ which is a linear combination of $\boldsymbol{x}$ and $\boldsymbol{e}$; *i.e.*, it is in the span of $\{\boldsymbol{x}, \boldsymbol{e}\}$. One way to see that we want our solution to be in the span of $\{\boldsymbol{x}, \boldsymbol{e}\}$ is to recast the problem as

$$\begin{bmatrix} | & | \\ \boldsymbol{x} & \boldsymbol{e} \\ | & | \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} | \\ \boldsymbol{y} \\ | \end{bmatrix} \tag{15}$$

We note that $\boldsymbol{x}$ and $\boldsymbol{e}$ are linearly independent and therefore form a basis for $\{\boldsymbol{x}, \boldsymbol{e}\}$. So we can write our minimizer $\boldsymbol{y}^*$ as $\boldsymbol{y}^* = \alpha \boldsymbol{x} + \beta \boldsymbol{e}$, where α and β are as yet unknown and are solved for from the linear system resulting from the orthogonality condition:

$$\begin{bmatrix} (\boldsymbol{x}, \boldsymbol{x}) & (\boldsymbol{x}, \boldsymbol{e}) \\ (\boldsymbol{e}, \boldsymbol{x}) & (\boldsymbol{e}, \boldsymbol{e}) \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} (\boldsymbol{x}, \boldsymbol{y}) \\ (\boldsymbol{e}, \boldsymbol{y}) \end{bmatrix} \tag{16}$$

Plugging everything into Matlab (see code at the end of the problem), we get that $\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 2.977 \\ 2.0691 \end{bmatrix}$. The plots of the raw data and the best-fit line are provided in figure 1.

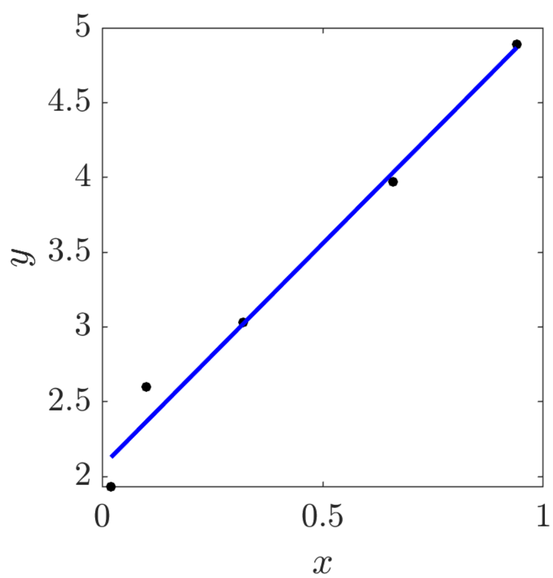

Figure 1: A plot of the raw data (black markers) and the best-fit solution (blue line).

```
clear all, close all, clc
```

```
%raw data
x = [ 0.02; 0.1; 0.32; 0.66; 0.94];
y = [1.93; 2.60; 3.03; 3.97; 4.89];
e = ones( size( x ) );

%plot raw data and hold on
plot( x, y, 'k.', 'markersize', 16 ), hold on


%set up and solve approximation problem:
A = [ x'*x x'*e; x'*e e'*e ];
b = [ x'*y; e'*y ];
soln = A \ b %always use \ instead of inv()

plot( x, soln(1)*x + soln(2)*e, 'b-', 'linewidth', 2  )

%make plot pretty
xlabel( '$x$', 'fontsize', 16, 'interpreter', 'latex' )
ylabel( '$y$', 'fontsize', 16, 'interpreter', 'latex' )
set( gca, 'fontsize', 16, 'ticklabelinterpreter', 'latex' )


%format and save
set(gcf, 'PaperPositionMode', 'manual')
set(gcf, 'Color', [1 1 1])
set(gca, 'Color', [1 1 1])
set(gcf, 'PaperUnits', 'centimeters')
set(gcf, 'PaperSize', [10 10])
set(gcf, 'Units', 'centimeters' )
set(gcf, 'Position', [0 0 10 10])
set(gcf, 'PaperPosition', [0 0 10 10])

dirsv = '~/Desktop/';
svnm = 'best_fit';
print( '-dpng', [dirsv, svnm], '-r200' )
```

4. We showed in class that when the problem $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ (with $\boldsymbol{A} \in \mathbb{R}^{m\times n}, \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{b} \in \mathbb{R}^m$) does not have a solution, then we can approximate this solution using a basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k\}$ for $\mathcal{R}(\boldsymbol{A})$. Re-derive these equations assuming that the basis is orthogonal; *i.e.*, when $(\boldsymbol{v}_i, \boldsymbol{v}_j) = 0$ for $i \neq j$. What simplifications arise, and what are some benefits of these simplifications?

   We write the approximation problem as

$$\min_{y \in \mathcal{R}(\boldsymbol{A})} ||\boldsymbol{y} - \boldsymbol{b}|| \tag{17}$$

   If $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k\}$ forms a basis for $\mathcal{R}(\boldsymbol{A})$ then we may write the minimizer (let's call it $\boldsymbol{u}$) as $\boldsymbol{u} = \sum_{i=1}^k \alpha_i \boldsymbol{v}_i$ for some as-yet undetermined coefficients α_i. Now we may use the fact that, as we just showed in problem 2, $(\boldsymbol{u} - \boldsymbol{b}, \boldsymbol{r}) = 0$ for *any* $\boldsymbol{r} \in \mathcal{R}(\boldsymbol{A})$. Since this holds for *any* $\boldsymbol{r} \in \mathcal{R}(\boldsymbol{A})$, we may use choose k different $\boldsymbol{r}$: $\boldsymbol{r} = \boldsymbol{v}_j$, $j = 1, \cdots, k$. This leads us with k

different equations:

$$
\begin{aligned}
(\boldsymbol{u}-\boldsymbol{b}, \boldsymbol{v}_j) &= 0, \quad j = 1, \dots, k && (18)\\
\implies \quad (\boldsymbol{u}, \boldsymbol{v}_j) &= (\boldsymbol{b}, \boldsymbol{v}_j), \quad j = 1, \dots, k && (19)\\
\implies \quad \left(\sum_{i=1}^{k} \alpha_i \boldsymbol{v}_i, \boldsymbol{v}_j \right) &= (\boldsymbol{b}, \boldsymbol{v}_j), \quad j = 1, \dots, k && (20)\\
\implies \sum_{i=1}^{k} \alpha_i (\boldsymbol{v}_i, \boldsymbol{v}_j) &= (\boldsymbol{b}, \boldsymbol{v}_j), \quad j = 1, \dots, k && (21)\\
\implies \alpha_j (\boldsymbol{v}_j, \boldsymbol{v}_j) &= (\boldsymbol{b}, \boldsymbol{v}_j), \quad j = 1, \dots, k && (22)\\
\implies \alpha_j &= \frac{(\boldsymbol{b}, \boldsymbol{v}_j)}{(\boldsymbol{v}_j, \boldsymbol{v}_j)}, \quad j = 1, \dots, k && (23)
\end{aligned}
$$

This is a huge simplification—it turns a full linear system of equations into a set of scalar equations for the unknown coefficients!