MTH 210, Spring 2016

HW 6 Solutions

Problem 1: Let A be any $m \times n$ matrix and assume that the $n \times n$ matrix $A^T A$ is invertible. Then we will define the $m \times m$ "projection matrix"

$$P := A(A^TA)^{-1}A^T.$$

(a) Using the general rules $(XY)^T = X^T Y^T$ and $(X^{-1})^T = (X^T)^{-1}$ shows us that

$$\begin{aligned}
P^T &= \left(A(A^TA)^{-1}A^T\right)^T \\
&= (A^T)^T\left((A^TA)^{-1}\right)^T(A)^T \\
&= A\left((A^TA)^T\right)^{-1}A^T \\
&= A\left((A)^T(A^T)^T\right)^{-1}A^T \\
&= A(A^TA)^{-1}A^T \\
&= P
\end{aligned}$$

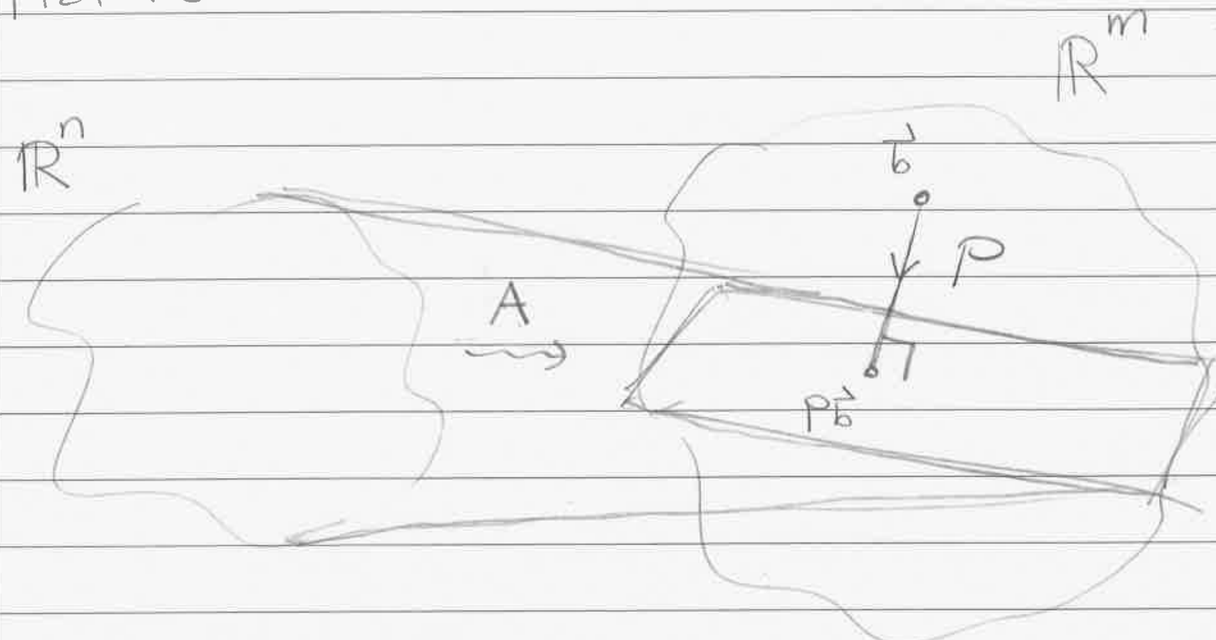
and

$$\begin{aligned}
P^2 &= \left(A(A^TA)^{-1}A^T\right)\left(A(A^TA)^{-1}A^T\right) \\
&= A(A^TA)^{-1}(A^T\cancel{A)(A^T}A)^{-1}A^T \\
&= A(A^TA)^{-1}I_n A^T \\
&= A(A^TA)^{-1}A^T = P
\end{aligned}$$

, as desired.

We saw in class that P is the matrix that projects orthogonally onto the column space of A.

Picture:



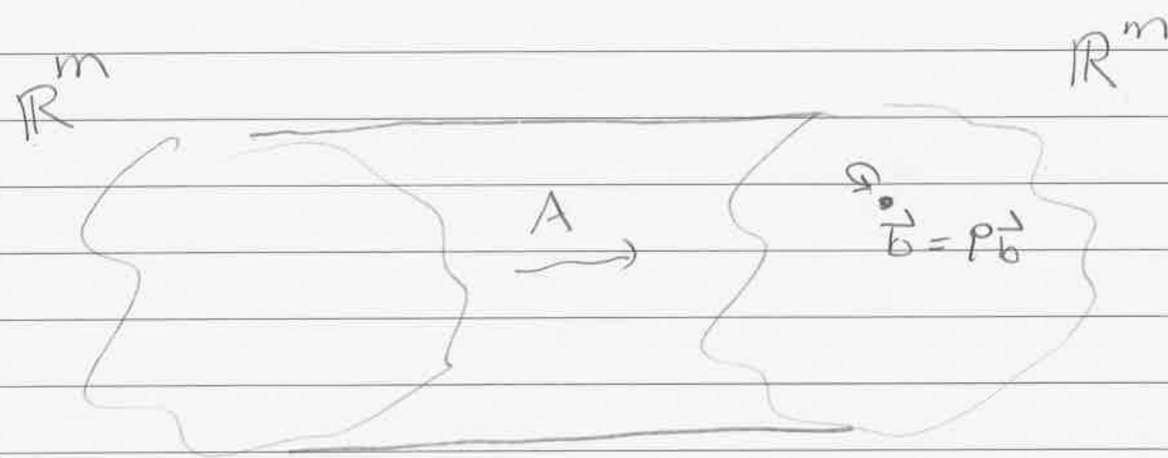
(b) But if A itself is a square $m \times m$ invertible matrix then we get

$$\begin{aligned} P &= A(A^TA)^{-1}A^T \\ &= \cancel{A(A)^{-1}}\cancel{(A^T)^{-1}A^T} \\ &= I_m I_m \\ &= I_m \end{aligned}$$

What's going on here?

In this case the column space of A is $\underline{\text{all}}$ of $\mathbb{R}^m$, so that "project onto the column space" means the same thing as "do nothing".

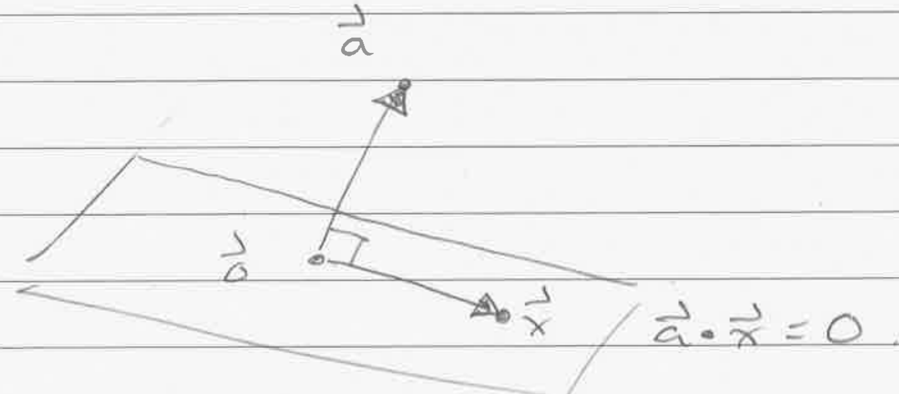
Picture:



Problem 2: Consider the following vector and its orthogonal plane.

$$\vec{a} = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \quad \& \quad x + 2y + 2z = 0.$$

Picture:



(a) Note that the line $t\vec{a}$ is the same as the column space of the matrix $A = \vec{a}$. So the projection matrix onto the line is

$$P_1 = \vec{a}\left(\vec{a}^T\vec{a}\right)^{-1}\vec{a}^T$$

$$= \begin{pmatrix}1\\2\\2\end{pmatrix}\left(\begin{pmatrix}1 & 2 & 2\end{pmatrix}\begin{pmatrix}1\\2\\2\end{pmatrix}\right)^{-1}\begin{pmatrix}1 & 2 & 2\end{pmatrix}$$

$$= \begin{pmatrix}1\\2\\2\end{pmatrix}\left(1+4+4\right)^{-1}\begin{pmatrix}1 & 2 & 2\end{pmatrix}$$

$$= \frac{1}{9}\begin{pmatrix}1\\2\\2\end{pmatrix}\begin{pmatrix}1 & 2 & 2\end{pmatrix}$$

$$= \frac{1}{9}\begin{pmatrix}1 & 2 & 2\\2 & 4 & 4\\2 & 4 & 4\end{pmatrix}.$$

(b) The projection of the point $\vec{b} = (1,-1,1)$ onto the line $t\vec{a}$ is

$$P_1\vec{b} = \frac{1}{9}\begin{pmatrix}1 & 2 & 2\\2 & 4 & 4\\2 & 4 & 4\end{pmatrix}\begin{pmatrix}1\\-1\\1\end{pmatrix} = \frac{1}{9}\begin{pmatrix}1\\2\\2\end{pmatrix}.$$

[Note: We could also have computed this with the formula

$$\left(\frac{\vec{a}^T\vec{b}}{\vec{a}^T\vec{a}}\right)\vec{a} = \left(\frac{(1\ 2\ 2)\begin{pmatrix}1\\-1\\1\end{pmatrix}}{(1\ 2\ 2)\begin{pmatrix}1\\2\\2\end{pmatrix}}\right)\begin{pmatrix}1\\2\\2\end{pmatrix}$$

$$= \left(\frac{1-2+2}{1+4+4}\right)\begin{pmatrix}1\\2\\2\end{pmatrix} = \frac{1}{9}\begin{pmatrix}1\\2\\2\end{pmatrix}.$$]

(c) To compute the projection onto the plane $x+2y+2z=0$ we first find two vectors in the plane. There are many ways to do this. Here's one way: Think of

$$\left\{\ x+2y+2z=0\right.$$

as a system of one linear equation in RREF. Then y & z are free variables so the general solution is

$$\begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{pmatrix}-2y-2z\\y\\z\end{pmatrix} = y\begin{pmatrix}-2\\1\\0\end{pmatrix} + z\begin{pmatrix}-2\\0\\1\end{pmatrix}.$$

We see that $(-2,1,0)$ & $(-2,0,1)$ are two vectors in the plane.

In other words, the plane $x+2y+2z=0$ is the same as the column space of the matrix

$$A = \begin{pmatrix} -2 & -2 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Now we can compute the projection matrix $P_2 = A(A^TA)^{-1}A^T$. First we compute

$$A^TA = \begin{pmatrix} -2 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix} \begin{pmatrix} -2 & -2 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}$$

$$\Rightarrow (A^TA)^{-1} = \frac{1}{5\cdot5-4\cdot4}\begin{pmatrix} 5 & -4 \\ -4 & 5 \end{pmatrix} = \frac{1}{9}\begin{pmatrix} 5 & -4 \\ -4 & 5 \end{pmatrix},$$

and then we have

$$P_2 = \begin{pmatrix} -2 & -2 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \frac{1}{9}\begin{pmatrix} 5 & -4 \\ -4 & 5 \end{pmatrix} \begin{pmatrix} -2 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix}$$

$$= \frac{1}{9}\begin{pmatrix} -2 & -2 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -2 & 5 & -4 \\ -2 & -4 & 5 \end{pmatrix}$$

$$= \frac{1}{9}\begin{pmatrix} 8 & -2 & -2 \\ -2 & 5 & -4 \\ -2 & -4 & 5 \end{pmatrix}.$$

(c) Next we compute the projection of the point $\vec{b} = (1, -1, 1)$ onto the plane:

$$P_2 \vec{b} = \frac{1}{9}\begin{pmatrix} 8 & -2 & -2 \\ -2 & 5 & -4 \\ -2 & -4 & 5 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$$

$$= \frac{1}{9}\begin{pmatrix} 8+2-2 \\ -2-5-4 \\ -2+4+5 \end{pmatrix} = \frac{1}{9}\begin{pmatrix} 8 \\ -11 \\ 7 \end{pmatrix}.$$

(d) Finally, we observe that

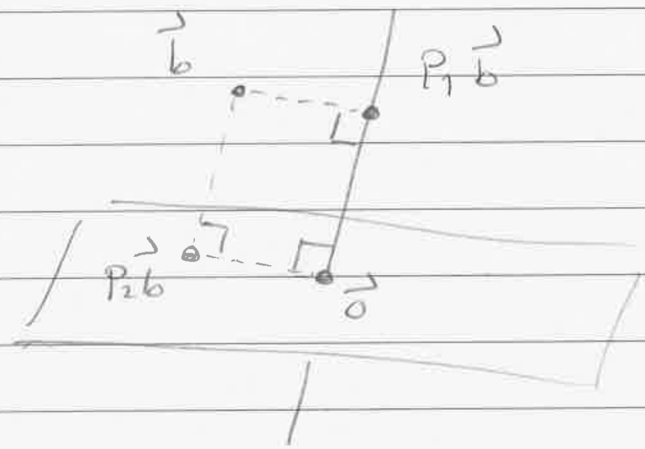
$$P_1 + P_2 = \frac{1}{9}\begin{pmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{pmatrix} + \frac{1}{9}\begin{pmatrix} 8 & -2 & -2 \\ -2 & 5 & -4 \\ -2 & -4 & 5 \end{pmatrix}$$

$$= \frac{1}{9}\begin{pmatrix} 1+8 & 2-2 & 2-2 \\ 2-2 & 4+5 & 4-4 \\ 2-2 & 4-4 & 4+5 \end{pmatrix}$$

$$= \frac{1}{9}\begin{pmatrix} 9 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Should we be surprised by this?

No. Given any point $\vec{b}$ we see that the points $\vec{0}, P_1\vec{b}, P_2\vec{b}$ & $\vec{b}$ form a 2D rectangle livining in $\mathbb{R}^3$:



Then the "Parallelogram Law" tells us that

$$\vec{b} = P_1\vec{b} + P_2\vec{b} = (P_1 + P_2)\vec{b}.$$

Since this is true for all $\vec{b}$ we must have $P_1 + P_2 = I_3$.

[This works more generally for the projection matrices onto any two "orthgonal subspaces".]

In summary, we really only needed to compute <u>one</u> of P_1 & P_2. Then we could have gotten the other one for free.

Problem 3: What linear combination of $(1,2,-1)$ & $(1,0,1)$ is closest to $(3,-1,-1)$?

Instead of using the same method as Problem 2 parts (c) & (d) I'll show you a quicker way to do this. The linear combinations of $(1,2,-1)$ & $(1,0,1)$ form the column space of

$$A = \begin{pmatrix} 1 & 1 \\ 2 & 0 \\ -1 & 1 \end{pmatrix}.$$

We want to find $\vec{x}$ such that $A\vec{x} = \vec{b}$ where $\vec{b} = (3,-1,-1)$. But this equation has no solution so instead we will solve the "normal equation"

$$A^T A \hat{x} = A^T \vec{b}$$

$$\begin{pmatrix} 1 & 2 & -1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 0 \\ -1 & 1 \end{pmatrix} \hat{x} = \begin{pmatrix} 1 & 2 & -1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 3 \\ -1 \\ -1 \end{pmatrix}$$

$$\begin{pmatrix} 6 & 0 \\ 0 & 2 \end{pmatrix} \hat{x} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}.$$

$$\hat{x} = \begin{pmatrix} 1/3 \\ 1 \end{pmatrix}.$$

Thus the point in the column space of A that is closest to $\vec{b}$ is given by

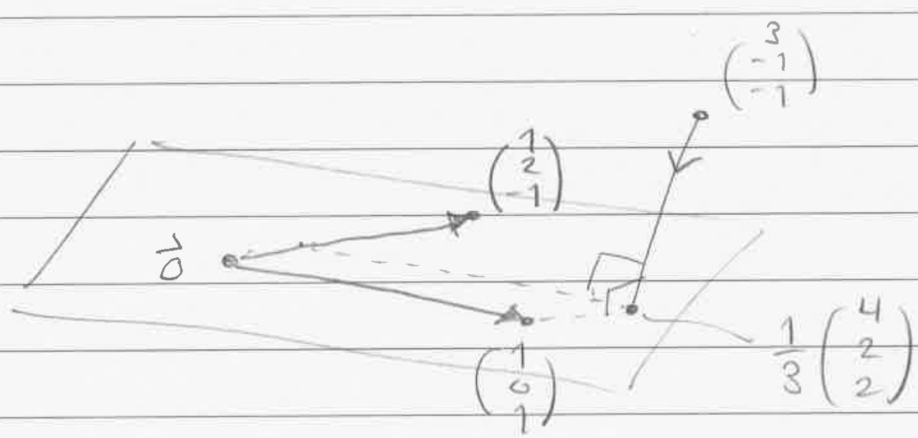
$$A\hat{x} = \begin{pmatrix} 1 & 1 \\ 2 & 0 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1/3 \\ 1 \end{pmatrix}$$

$$= \frac{1}{3}\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} + 1\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 4 \\ 2 \\ 2 \end{pmatrix}$$

[Alternatively, we could have solved this by first computing the projection matrix P onto the plane <u>or</u> the projection matrix Q onto the orthogonal line and then computing

$P\vec{b}$ <u>or</u> $(I-Q)\vec{b}$.]

Picture:

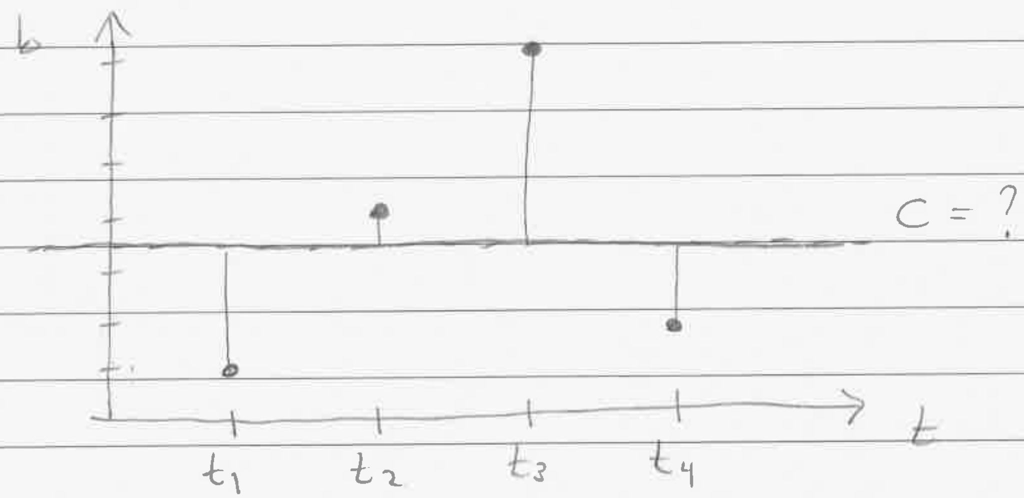


Problem 4: (a) Consider the data points.

$$\begin{pmatrix} t \\ b \end{pmatrix} = \begin{pmatrix} t_1 \\ 1 \end{pmatrix}, \begin{pmatrix} t_2 \\ 4 \end{pmatrix}, \begin{pmatrix} t_3 \\ 7 \end{pmatrix}, \begin{pmatrix} t_4 \\ 2 \end{pmatrix}.$$

We want to find the horizontal line $C + 0t = b$ that is the "best fit" for these points.

Picture:



"Best" will mean that we minimize the sum of the squares of the vertical errors in the picture. To this we first naïvely try to find a horizontal line that contains all four points:

$$\begin{cases} C = 1 \\ C = 4 \\ C = 7 \\ C = 2 \end{cases} \rightsquigarrow \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} (C) = \begin{pmatrix} 1 \\ 4 \\ 7 \\ 2 \end{pmatrix}.$$

This is very silly and clearly it has no solution. So now we pass to the "normal equation" by multiplying both sides on the left by $(1,1,1,1)^T$:

$$\begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} (c) = \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 4 \\ 7 \\ 2 \end{pmatrix}$$

$$(4)(c) = (14)$$

and this gives us $C = 14/4 = 7/2$. Note that this is just the <u>average</u> of the four values $1, 4, 7, 2$, as one would expect. This demonstrates that computing an average of numbers is a specific example of "least squares approximation".

(b) More generally, consider any vector
$\vec{b} = (b_1, b_2, \ldots, b_m)$ and let
$\vec{a} = (1, 1, \ldots, 1)$. Then performing the
same calculation as in part (a) gives

$$\vec{a}^T \vec{a} \hat{x} = \vec{a}^T \vec{b}$$

$$(1\ 1 \cdots 1) \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \hat{x} = (1\ 1 \cdots 1) \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

$$(m)\, \hat{x} = \left(\sum_{i=1}^{m} b_i \right)$$

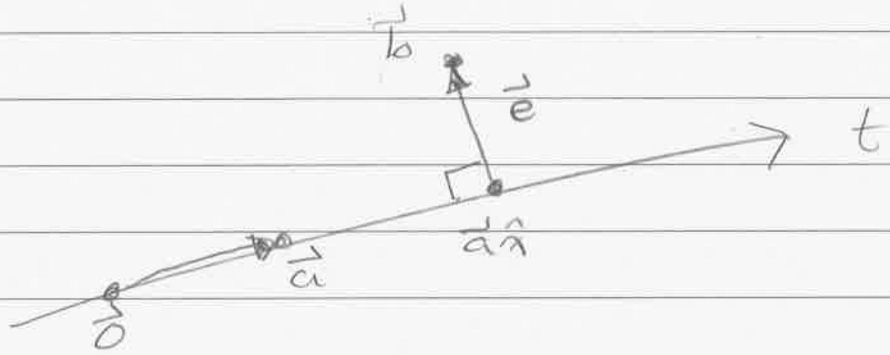
and we conclude that

$$\hat{x} = \frac{\sum_{i=1}^{m} b_i}{m},$$

which is just the average of $b_1, \ldots, b_m$.

(c) What was minimized here?
Geometrically, we have projected the
point $\vec{b}$ orthogonally onto the line
$t\vec{a}$:

↓



This means that we have minimized the <u>length</u> of the vector

$$\vec{e} = \vec{b} - \vec{a}\hat{x}$$

$$= \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} - \begin{pmatrix} \hat{x} \\ \hat{x} \\ \vdots \\ \hat{x} \end{pmatrix} = \begin{pmatrix} b_1 - \hat{x} \\ b_2 - \hat{x} \\ \vdots \\ b_m - \hat{x} \end{pmatrix},$$

which is defined by

$$\|\vec{e}\|^2 = \vec{e}^T\vec{e}$$

$$= (b_1 - \hat{x})^2 + (b_2 - \hat{x})^2 + \cdots + (b_m - \hat{x})^2.$$

If you've taken a statistics couse (like MTH 224) then you will recognize the quantity $\|\vec{e}\|^2/m$ as the <u>variance</u> of the numbers $b_1, \ldots, b_m$.

Problem 5: Find the equation $C + tD = b$ of the best fit line for the data points

$$\begin{pmatrix} t \\ b \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

(a) First we write down the naive system

$$\begin{cases} C - D = 3 \\ C + 0 = 2 \\ C + D = 2 \\ C + 2D = 1 \end{cases} \rightsquigarrow \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ 2 \\ 1 \end{pmatrix}$$

and express it as $A\vec{x} = \vec{b}$. This system has no solution because the four points don't actually fit on a single line.

(b) So now we pass to the normal equation

$$A^T A \hat{x} = A^T \vec{b}$$

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \hat{x} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \\ 2 \\ 1 \end{pmatrix}.$$

$$\begin{pmatrix} 4 & 2 \\ 2 & 6 \end{pmatrix} \hat{x} = \begin{pmatrix} 8 \\ 1 \end{pmatrix}$$

and then we solve it to get

$$\left(\begin{array}{cc|c} 4 & 2 & 8 \\ 2 & 6 & 1 \end{array}\right) \longrightarrow \left(\begin{array}{cc|c} 2 & 1 & 4 \\ 2 & 6 & 1 \end{array}\right)$$

$$\longrightarrow \left(\begin{array}{cc|c} 2 & 1 & 4 \\ 0 & 5 & -3 \end{array}\right) \longrightarrow \left(\begin{array}{cc|c} 2 & 1 & 4 \\ 0 & 1 & -3/5 \end{array}\right)$$

$$\longrightarrow \left(\begin{array}{cc|c} 2 & 0 & 23/5 \\ 0 & 1 & -3/5 \end{array}\right) \longrightarrow \left(\begin{array}{cc|c} 1 & 0 & 23/10 \\ 0 & 1 & -3/5 \end{array}\right),$$

hence $\hat{x} = \begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} 23/10 \\ -3/5 \end{pmatrix}$.

We conclude that the best fit line is

$$\frac{23}{10} - \frac{3}{5} t = b.$$

(c) The error vector is

$$\vec{e} = \vec{b} - A\hat{x}.$$

$$= \begin{pmatrix} 3 \\ 2 \\ 2 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 23/10 \\ -3/5 \end{pmatrix}$$

$$= \begin{pmatrix} 3 \\ 2 \\ 2 \\ 1 \end{pmatrix} - \frac{23}{10} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \frac{3}{5} \begin{pmatrix} -1 \\ 0 \\ 1 \\ 2 \end{pmatrix}$$

$$= \begin{pmatrix} 1/10 \\ -3/10 \\ 3/10 \\ -1/10 \end{pmatrix}$$

Picture: