HW5 Notes

OK, I think we've now covered enough general theory for a first course in Linear Algebra. As promised, I will now show you two of the main applications of Linear Algebra:

① "least squares regression"
② "spectral analysis"

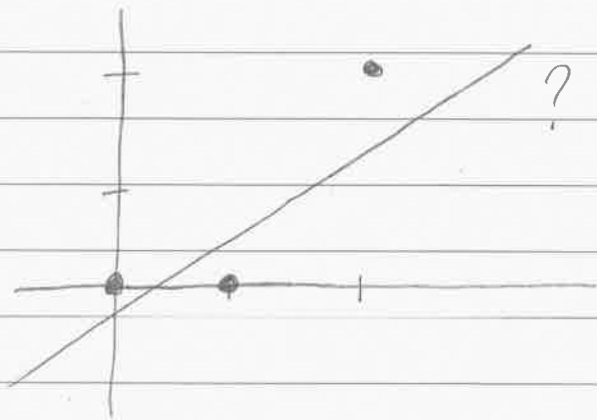
Application ① is the reason that Gauss invented "Gaussian elimination" in the first place. I'll introduce this topic with an example.

Motivating Example: Consider three points in the Cartesian plane

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

[Why did I use the funny letters t & b? Stay tuned.]

Do these three points lie on a line? Well, we could graph them and see that the answer is no:



? No.

Or we could use an algebraic approach. Suppose that the line $C + tD = b$ contains all three points, so that

$$\begin{cases} C + 0D = 0 \\ C + 1D = 0 \\ C + 2D = 2 \end{cases}$$

Now try to solve for C & D.

$$\left(\begin{array}{cc|c} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 2 \end{array}\right) \longrightarrow \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 2 \end{array}\right) \longrightarrow \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{array}\right),$$

After elimination we obtain the equation $0 \cdot C + 0 \cdot D = 2$, which tells us that there is NO SOLUTION. [This is not surprising based on the picture we drew.]

But what if we don't need an <u>exact</u> solution? Maybe we just want to find a good (or "best"?) approximate solution. In our example we might want to find the line $C + tD = b$ that is the "best fit" for the three given points.

This is precisely the problem that Gauss solved in 1795 (when he was 18 years old), and his solution has not been improved upon since. To describe Gauss' solution we will write our system in matrix form

$$A\vec{x} = \vec{b}$$

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix},$$

where
$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}, \quad \vec{x} = \begin{pmatrix} C \\ D \end{pmatrix}, \quad \vec{b} = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}.$$

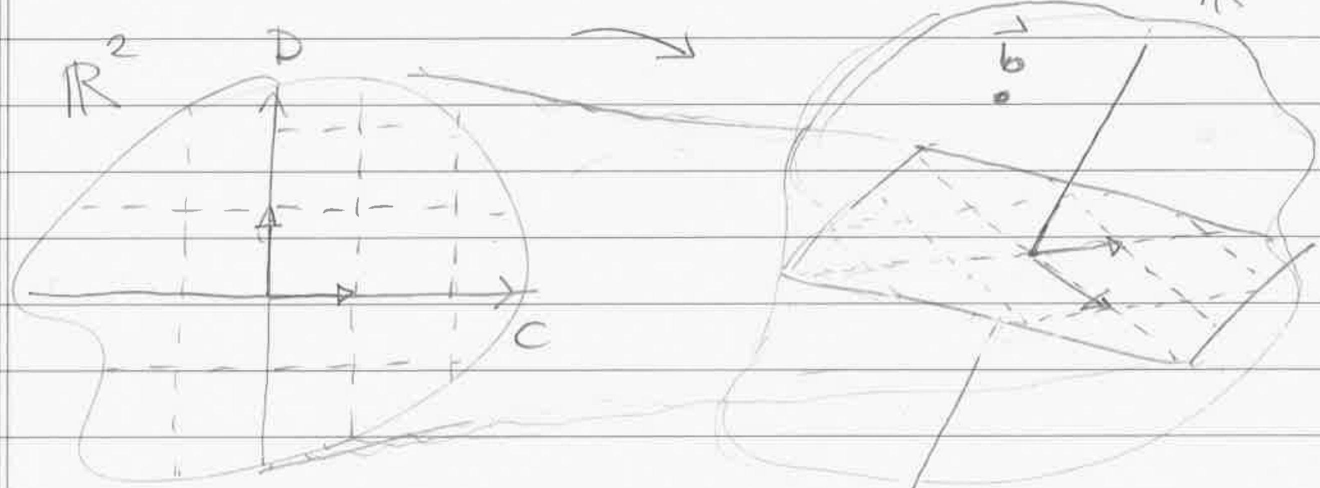
Expanding the equation in terms of columns gives us

$$A \begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} C \\ D \end{pmatrix} = C \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + D \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}$$

Thus we can think of A as a function from $\mathbb{R}^2$ to $\mathbb{R}^3$ that sends all of $\mathbb{R}^2$ onto the <u>plane</u>

$$C \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + D \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \quad \text{in } \mathbb{R}^3.$$

Picture:



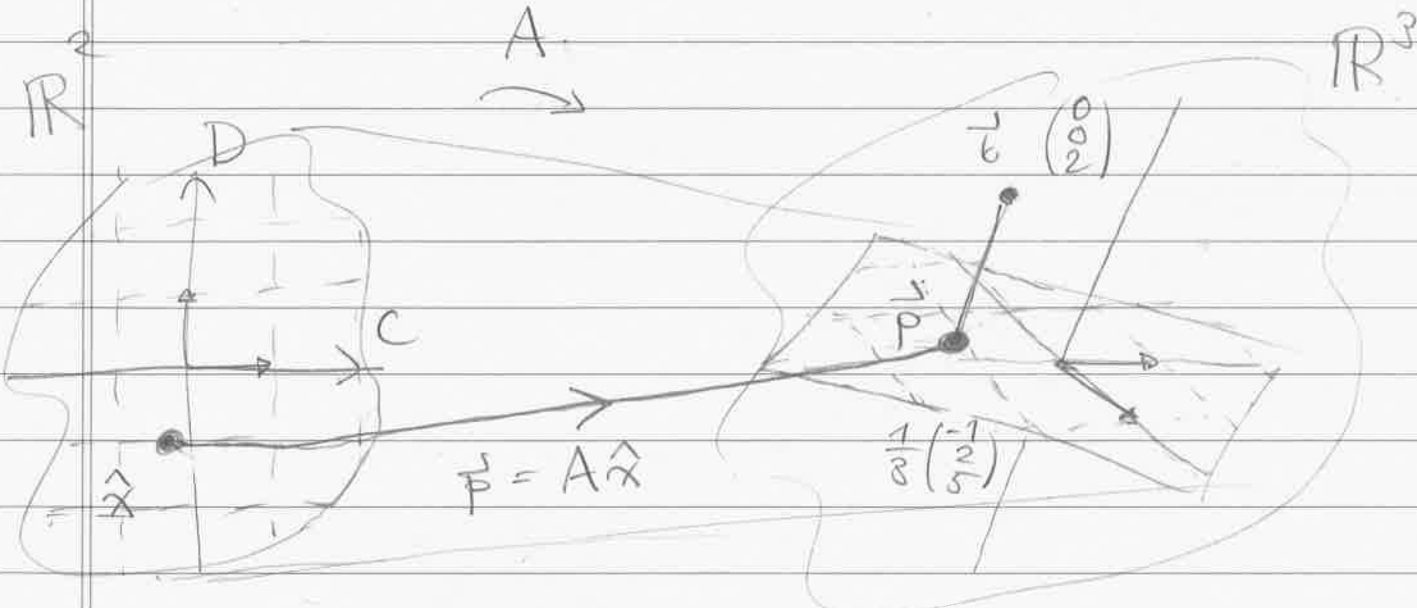
Jargon: We call this plane in $\mathbb{R}^3$ the image, or the column space, of the matrix A.

Now we see that our system $A\vec{x} = \vec{b}$ has no solution because the target point

$$\vec{b} = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}$$

is not in the image of A!

Gauss' idea (called "least squares approximation") was to replace the bad target point $\vec{b}$ with a new target point $\vec{p}$ that is in the image of A:



The goal is to choose $\vec{p}$ in the image so that the distance

$$\| \vec{b} - \vec{p} \|$$

is as small as possible. Then the solution $\hat{x}$ of the equation $A\hat{x} = \vec{p}$ (which exists by the assumption that $\vec{p}$ is in the image of A) is called a least squares approximation to the system $A\vec{x} = \vec{b}$.

In our example, the point $\vec{p}$ is given by

$$\vec{p} = \frac{1}{3}\begin{pmatrix} -1 \\ 2 \\ 5 \end{pmatrix}.$$

[You'll just have to believe me for now...]

And our "least squares" solution is given by

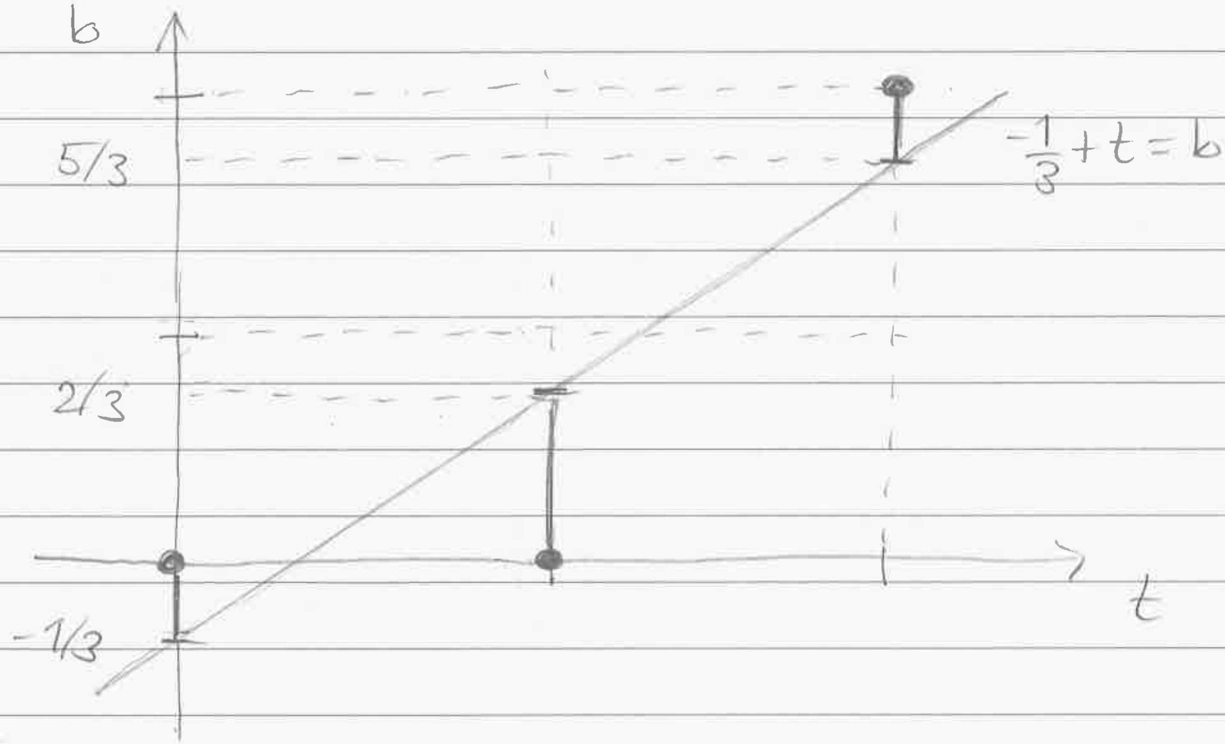
$$A\hat{x} = \vec{p}$$

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} -1/3 \\ 2/3 \\ 5/3 \end{pmatrix}.$$

Let's compute it:

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

We conclude that $\hat{x} = \begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} -1/3 \\ 1 \end{pmatrix}$,
so the "best fit" line to our three given points is $-1/3 + t = b$:



In what sense is this line "best"?

Well, we chose the point $\vec{p}$ so that the distance $\|\vec{b}-\vec{p}\|$ is minimized, which means that the distance squared is also minimized:

$$\|\vec{b}-\vec{p}\|^2 = \left\| \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix} - \begin{pmatrix} -1/3 \\ 2/3 \\ 5/3 \end{pmatrix} \right\|^2$$

$$= \left(0+\frac{1}{3}\right)^2 + \left(0-\frac{2}{3}\right)^2 + \left(2-\frac{5}{3}\right)^2$$

$$= \left(\frac{1}{3}\right)^2 + \left(-\frac{2}{3}\right)^2 + \left(-\frac{5}{3}\right)^2.$$

Note that this is the sum of the squares of the vertical errors in our picture.

Thus our line is "best" in the sense that sum of the squares of the vertical errors is minimized,

hence the name "Least squares".

Next time: How did I compute the point $\vec{p}$?

Last time we decided that the line

$$-\frac{1}{3} + t = b$$

is the "best fit" for the three data points

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

How did we do it? First we assumed (naïvely) that the three points line on the same line $C + tD = b$, which leads to the unsolvable system of equations

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

which we will write as

$$A\vec{x} = \vec{b}.$$

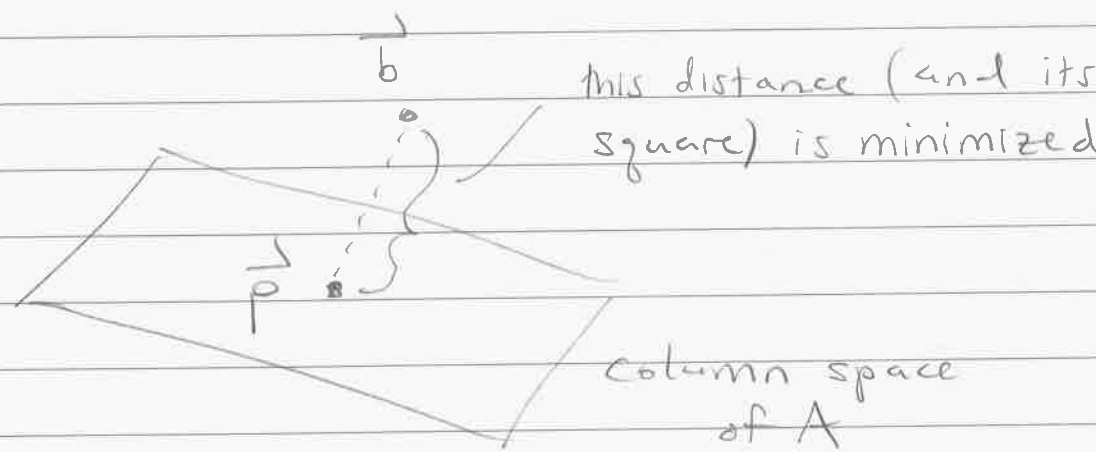
As $\vec{x}$ ranges over all points (C, D) in $\mathbb{R}^2$ the expression $A\vec{x}$ ranges over the points

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} C \\ D \end{pmatrix} = C \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + D \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix},$$

which form a plane in $\mathbb{R}^3$. [Jargon: we call this plane the "column space" of the matrix A, for obvious reasons.]

Then the reason $A\vec{x} = \vec{b}$ has no solution is because the target point $\vec{b} = (0, 0, 2)$ is not in the column space of A.

Gauss' solution is to let $\vec{p}$ be the closest point to $\vec{b}$ in the column space of A:



Then the resulting system $A\hat{x} = \vec{p}$ (of "normal equations") does have a solution $\hat{x}$, which is called a least squares approximation to the system $A\vec{x} = \vec{b}$.

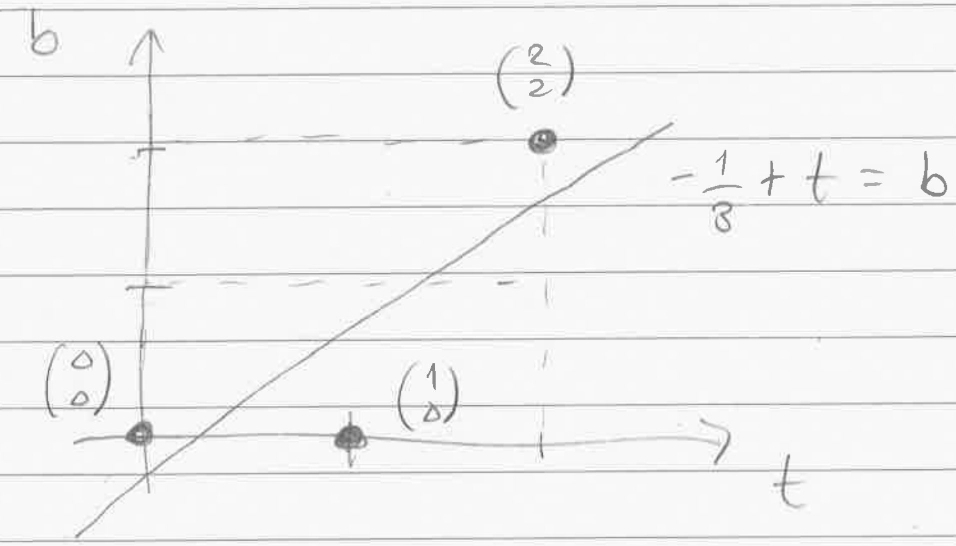
In our case, I told you that

$$\vec{p} = \frac{1}{3}\begin{pmatrix} -1 \\ 2 \\ 5 \end{pmatrix}$$

and then we solved the normal equation to get

$$\hat{x} = \begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} -1/3 \\ 1 \end{pmatrix},$$

which leads to the "best fit" line:

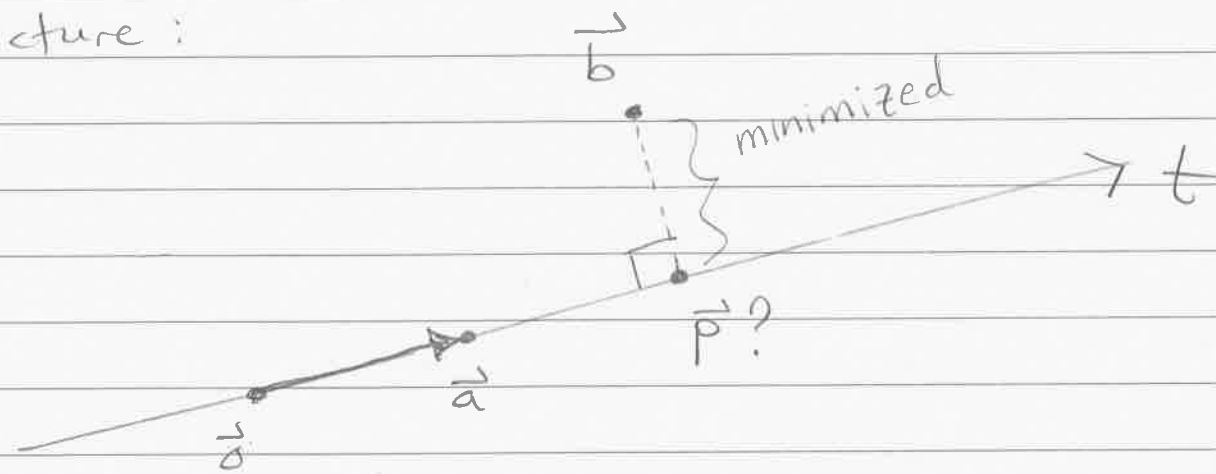


But how did I find the point $\vec{p}$?

In general, let A be any $m \times n$ matrix and let $\vec{b}$ be any $m \times 1$ vector. We want to find the point $\vec{p}$ in the column space of A that is closest to $\vec{b}$.

Let's start with the case when $n = 1$, so that $A = \vec{a}$ is an $m \times 1$ vector and the "column space" of A is just the line $t\vec{a}$.

Picture:



We want to find the point $\vec{p} = t\vec{a}$ so that the distance $\|\vec{b} - \vec{p}\|$ is minimized.

For geometric reasons, we see that this will be accomplished when the vector $\vec{b} - \vec{p}$ is perpendicular to the line, i.e. when

$$\vec{b} - \vec{p} \perp \vec{a} .$$

Then the dot product immediately gives us the solution. We have $\vec{p} = t\vec{a}$ and $\vec{a} \cdot (\vec{b} - \vec{p}) = 0$, hence

$$\begin{aligned} \vec{a} \cdot (\vec{b} - t\vec{a}) &= 0 \\ \vec{a}^T (\vec{b} - t\vec{a}) &= 0 \\ \vec{a}^T\vec{b} - t\,\vec{a}^T\vec{a} &= 0 \\ \vec{a}^T\vec{b} &= t\,\vec{a}^T\vec{a} \end{aligned}$$

$$\Longrightarrow \quad t = \frac{\vec{a}^T\vec{b}}{\vec{a}^T\vec{a}} ,$$

and we conclude that

$$\boxed{\vec{p} = \left(\frac{\vec{a}^T\vec{b}}{\vec{a}^T\vec{a}}\right)\vec{a}}$$

Jargon: We call this the <u>orthogonal projection</u> of $\vec{b}$ onto the line $t\vec{a}$.

[so $\vec{p}$ is for "projection" ☺].

Examples:

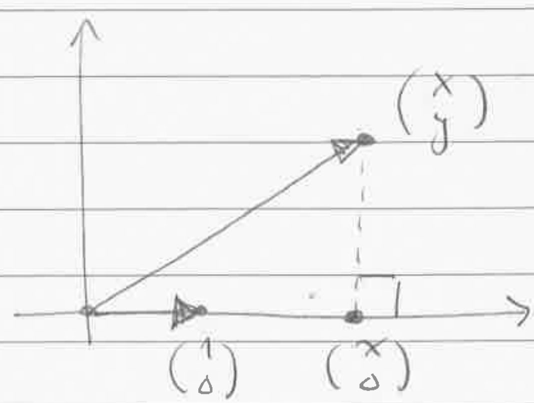
- Project the general point (x, y) onto the line $t(1, 0)$.

We have $\vec{b} = \begin{pmatrix} x \\ y \end{pmatrix}$ & $\vec{a} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, so the projection is

$$\vec{p} = \left(\frac{\vec{a}^T \vec{b}}{\vec{a}^T \vec{a}} \right) \vec{a}$$

$$= \left(\frac{(1 \; 0)\begin{pmatrix} x \\ y \end{pmatrix}}{(1 \; 0)\begin{pmatrix} 1 \\ 0 \end{pmatrix}} \right) \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

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



no surprise here.

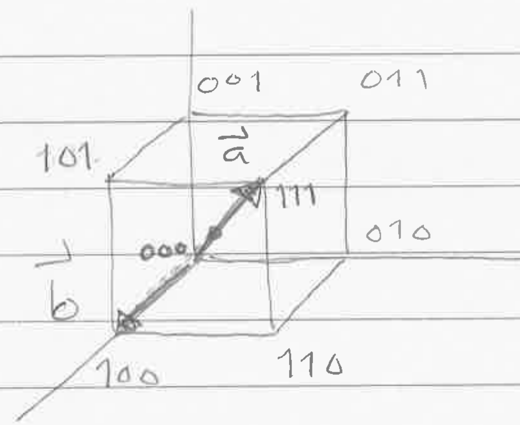
- Project the point $(1,0,0)$ onto the line $t(1,1,1)$.

We have $\vec{b} = \begin{pmatrix}1\\0\\0\end{pmatrix}$ & $\vec{a} = \begin{pmatrix}1\\1\\1\end{pmatrix}$ so that

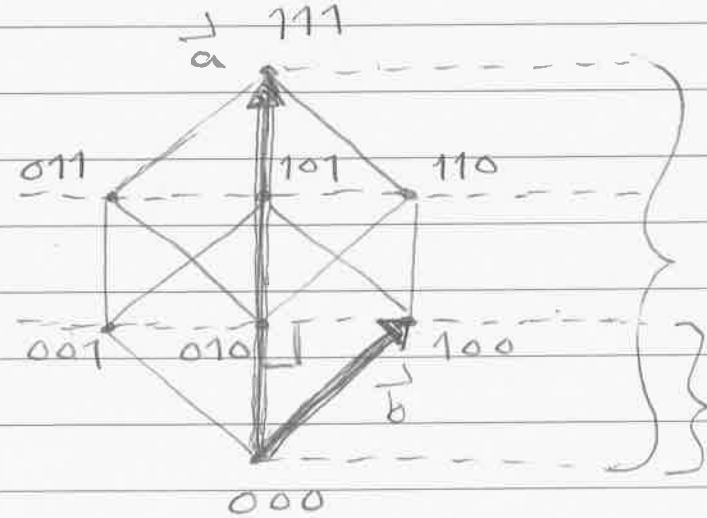
$$\vec{p} = \left(\frac{(1\ 1\ 1)\begin{pmatrix}1\\0\\0\end{pmatrix}}{(1\ 1\ 1)\begin{pmatrix}1\\1\\1\end{pmatrix}}\right)\begin{pmatrix}1\\1\\1\end{pmatrix} = \frac{1}{3}\begin{pmatrix}1\\1\\1\end{pmatrix}.$$

Is this a surprise?

Bad Picture:



Better Picture:

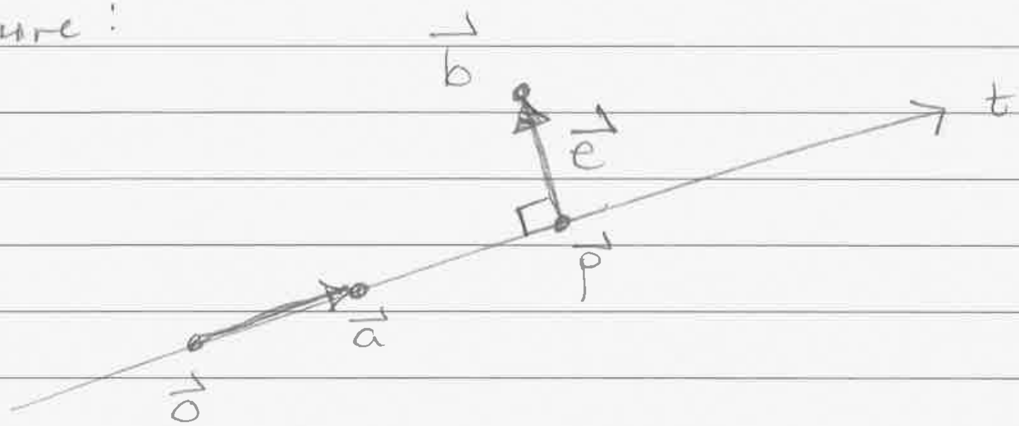


The projection is 1/3 of the way to $\begin{pmatrix}1\\1\\1\end{pmatrix}$.

Now: Least-Squares Approximation.

Last time we found the formula for <u>projecting</u> the point $\vec{b}$ onto the line $t\vec{a}$.

Picture:



Let $\vec{p} = t\vec{a}$ be a point on the line. We want to compute t. To do this we define the "error vector"

$$\vec{e} = \vec{b} - \vec{p}.$$

For geometric reasons, the length $\|\vec{e}\|$ will be <u>minimized</u> when $\vec{e} \perp \vec{a}$, i.e. when $\vec{a}^T \vec{e} = 0$.

Then by plugging in $\vec{e} = \vec{b} - \vec{p} = \vec{b} - t\vec{a}$ we can solve for t:

$$\vec{a}^T \vec{e} = 0$$
$$\vec{a}^T (\vec{b} - t\vec{a}) = 0$$
$$\vec{a}^T \vec{b} - t\, \vec{a}^T \vec{a} = 0$$
$$\vec{a}^T \vec{b} = t\, \vec{a}^T \vec{a}$$

$$\Rightarrow \quad t = \frac{\vec{a}^T \vec{b}}{\vec{a}^T \vec{a}} .$$

[I know this looks fancy, but keep in mind that it is just a number.]

We conclude that the orthogonal projection of the point $\vec{b}$ onto the line $t\vec{a}$ is the point

$$\vec{p} = \underbrace{\left(\frac{\vec{a}^T \vec{b}}{\vec{a}^T \vec{a}} \right)}_{\text{number}} \underset{\text{vector}}{\underset{\uparrow}{\vec{a}}} .$$

More generally, we want to minimize the distance between a point $\vec{b}$ in $\mathbb{R}^n$ and some arbitary "subspace" of $\mathbb{R}^n$.

Example : Find the linear combination of the vectors $(1,1,1)$ & $(0,1,2)$ that is closest to the point $(0,0,2)$.

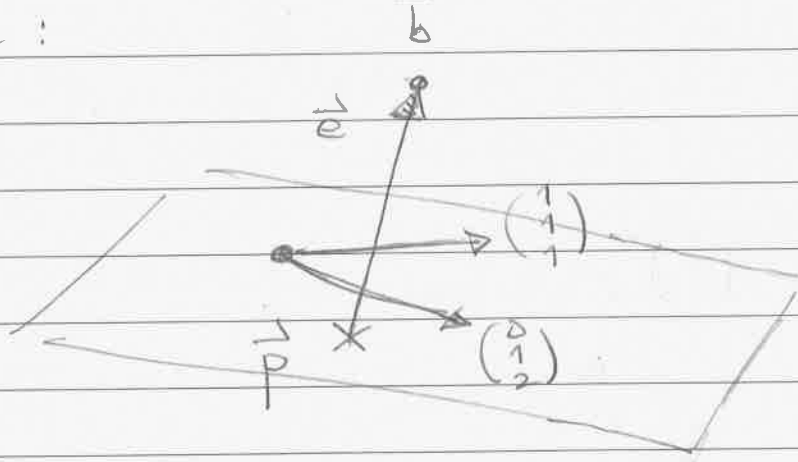
I'll solve this by using the general method. We define a vector and a matrix.

$$\vec{b} = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix} \quad \& \quad A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}$$

so that linear combinations of $(1,1,1)$ $(0,1,2)$ have the form

$$A\vec{x} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} C \\ D \end{pmatrix} = C\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + D\begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix},$$

Picture :



Now we let $\vec{p} = A\hat{x}$ be some arbitrary point in the plane and we define the "error vector"

$$\vec{e} = \vec{b} - \vec{p}.$$

For geometric reasons the distance will be minimized when $\vec{e}$ is perpendicular to the plane, but how can we turn this idea into an equation?

Note that $\vec{e}$ is $\perp$ to the plane when $\vec{e}$ is $\perp$ to both of $(1,1,1)$ & $(0,1,2)$. In other words we must have

$$(1\ 1\ 1)\,\vec{e} = 0 \quad \& \quad (0\ 1\ 2)\,\vec{e} = 0$$

Is there a way to express these two vector equations as one matrix equation? Certainly: we can write them as

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix} \vec{e} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

$$A^T \vec{e} = \vec{0}.$$

[In general, a vector $\vec{e}$ is perpendicular to all of the columns of a matrix A when

$$A^T \vec{e} = \vec{0}.]$$

Now we can substitute $\vec{e} = \vec{b} - \vec{p}$ and solve for the point $\vec{p}$

$$\begin{aligned} A^T \vec{e} &= \vec{0} \\ A^T(\vec{b} - \vec{p}) &= \vec{0} \\ A^T\vec{b} - A^T\vec{p} &= \vec{0} \\ A^T\vec{b} &= A^T\vec{p}. \end{aligned}$$

$$\boxed{A^T\vec{p} = A^T\vec{b}}$$

Or we could go further by substituting $\vec{p} = A\hat{x}$ and then solving for $\hat{x}$

$$\begin{aligned} A^T\vec{p} &= A^T\vec{b} \\ A^T(A\hat{x}) &= A^T\vec{b} \\ (A^TA)\hat{x} &= A^T\vec{b}. \end{aligned}$$

The matrix A may not be square, but the matrix A^TA always is square.

If A^TA is invertible then we can go further by multiplying both sides on the left by the inverse $(A^TA)^{-1}$:

$$\cancel{(A^TA)^{-1}(A^TA)}\hat{x} = (A^TA)^{-1}A^T\vec{b}$$

$$I\hat{x} = (A^TA)^{-1}A^T\vec{b}$$

$$\boxed{\hat{x} = (A^TA)^{-1}A^T\vec{b}}$$

Finally, we can get back to $\vec{p}$ if we want by multiplying on the left by A.

$$A\hat{x} = A(A^TA)^{-1}A^T\vec{b}$$

$$\boxed{\vec{p} = A(A^TA)^{-1}A^T\vec{b}}$$

This is the formula for the orthogonal projection of the point $\vec{b}$ onto the column space of the matrix A.

Let's compute the projection in our example

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \quad \& \quad \vec{b} = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}.$$

First we have

$$A^T A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}$$

$$= \begin{pmatrix} 1+1+1 & 0+1+2 \\ 0+1+2 & 0+1+4 \end{pmatrix}$$

$$= \begin{pmatrix} 3 & 3 \\ 3 & 5 \end{pmatrix}.$$

Note that this is a square, invertible matrix with

$$(A^T A)^{-1} = \begin{pmatrix} 3 & 3 \\ 3 & \end{pmatrix}^{-1}$$

$$= \frac{1}{3 \cdot 5 - 3 \cdot 3} \begin{pmatrix} 5 & -3 \\ -3 & 3 \end{pmatrix}$$

$$= \frac{1}{6} \begin{pmatrix} 5 & -3 \\ -3 & 3 \end{pmatrix}.$$

Then we compute

$$A(A^TA)^{-1}A^T = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \frac{1}{6}\begin{pmatrix} 5 & -3 \\ -3 & 3 \end{pmatrix}\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix}$$

$$= \frac{1}{6}\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 5 & 2 & -1 \\ -3 & 0 & 3 \end{pmatrix}$$

$$= \frac{1}{6}\begin{pmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{pmatrix}.$$

This is the matrix that projects <u>any</u> vector in $\mathbb{R}^3$ onto the column space of A, i.e., onto the plane spanned by $(1,1,1)$ & $(0,1,2)$. Finally, we apply this projection to the point $\vec{b} = (0,0,2)$ to get

$$\vec{p} = A(A^TA)^{-1}A^T\vec{b}$$

$$= \frac{1}{6}\begin{pmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{pmatrix}\begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}$$

$$= \frac{1}{6}\begin{pmatrix} -2 \\ 4 \\ 10 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} -1 \\ 2 \\ 5 \end{pmatrix},$$

which verifies what I told you in Monday's class. ‼ ◡

Whew, isn't there a faster way to do that? Sure. If you don't need to know the projection matrix you can just solve the system $A^TA\hat{x} = A^T\vec{b}$:

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

$$\begin{pmatrix} 3 & 3 \\ 3 & 5 \end{pmatrix}\hat{x} = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$$

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

$$\longrightarrow \left(\begin{array}{cc|c} 3 & 0 & -1 \\ 0 & 1 & 1 \end{array}\right) \longrightarrow \left(\begin{array}{cc|c} 1 & 0 & -1/3 \\ 0 & 1 & 1 \end{array}\right) \longrightarrow \hat{x} = \begin{pmatrix} -1/3 \\ 1 \end{pmatrix}$$

and then the projection is

$$\vec{p} = A\hat{x} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} -1/3 \\ 1 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} -1 \\ 2 \\ 5 \end{pmatrix} \quad \checkmark$$

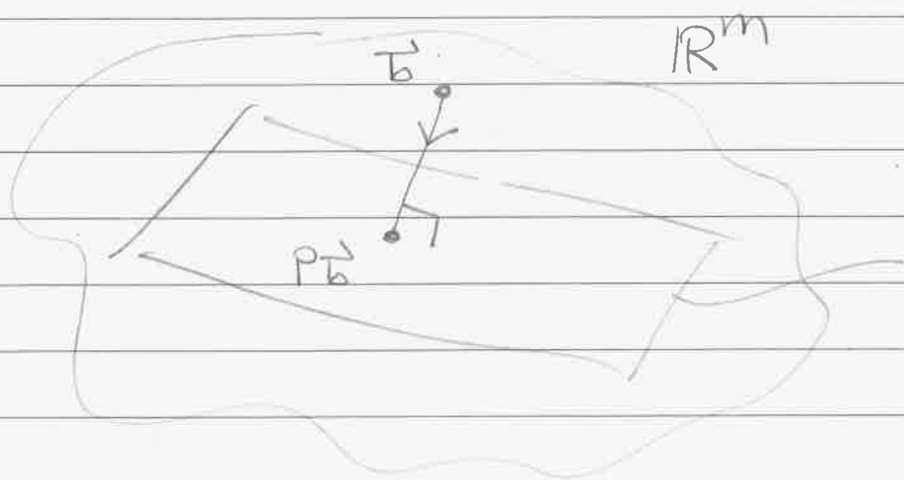
=

Recall from last time:

Let A be an $m \times n$ matrix such that the inverse $(A^TA)^{-1}$ exists [which happens precisely when A has no column relations]. Then the column space of A is an n dimensional "subspace" of $\mathbb{R}^m$, and the $m \times m$ matrix

$$P := \underbrace{\underset{m\times n}{A}\,\underset{n\times n}{(A^TA)^{-1}}\,\underset{n\times m}{A^T}}$$

corresponds to the function $\mathbb{R}^m \to \mathbb{R}^m$ that "orthogonally projects" any point $\vec{b}$ in $\mathbb{R}^m$ onto the column space:



n-dimensional column space of A.

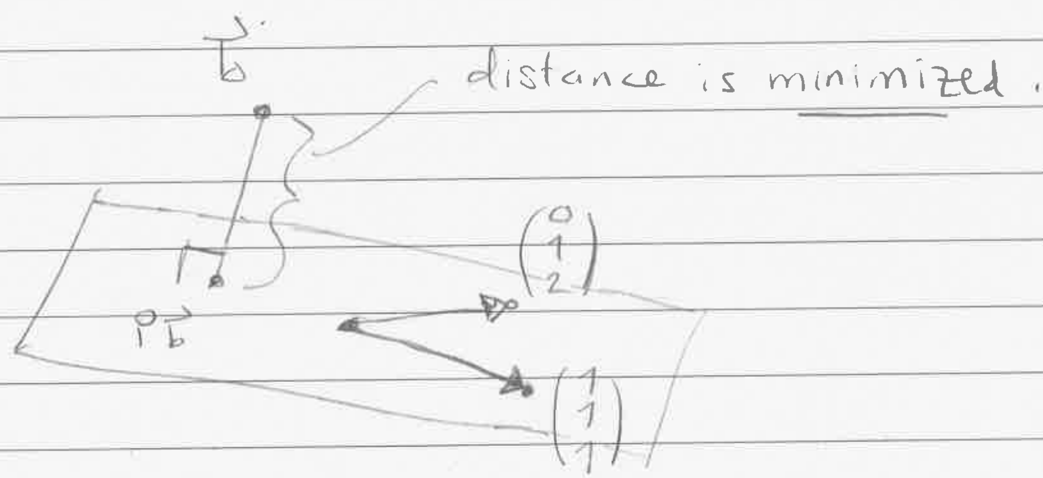
[On HW 6.1 you will examine the "abstract algebraic" properties of the matrix P. In particular, you will show that $P^2 = P$, which makes sense from the geometric point of view (projecting twice is the same as projecting once).]

For example, when

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

we found that $P = \dfrac{1}{6}\begin{pmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{pmatrix}$.

Picture:



[The key property from our point of view is that $P\vec{b}$ is the point in the column space of A that minimizes the distance

$$\| \vec{b} - P\vec{b} \|.$$
]

Now let's consider the line that is orthogonal to the plane

$$s\begin{pmatrix}1\\1\\1\end{pmatrix} + t\begin{pmatrix}0\\1\\2\end{pmatrix},$$

What vector spans this line? We want to find a vector $\vec{x}$ that is simultaneously perpendicular to $(1,1,1)$ & $(0,1,2)$. In other words we want to solve the two vector equations

$$(1\ 1\ 1)\vec{x} = 0 \quad \& \quad (0\ 1\ 2)\vec{x} = 0,$$

which can be expressed together as one matrix equation:

$$\begin{pmatrix}1 & 1 & 1\\0 & 1 & 2\end{pmatrix}\vec{x} = \begin{pmatrix}0\\0\end{pmatrix}.$$

Now solve :

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 1 & 2 & 0 \end{array}\right) \longrightarrow \begin{array}{c} \begin{array}{ccc} x & y & z \end{array} \\ \left(\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 \end{array}\right) \end{array} \text{ RREF } \checkmark$$

If $\vec{x} = (x, y, z)$ then x & y are the pivot variables and z is free. We conclude that

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

[Remark: If you have taken physics or multivariable calculus then you will know that there is a shortcut to this answer called the "cross product"

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

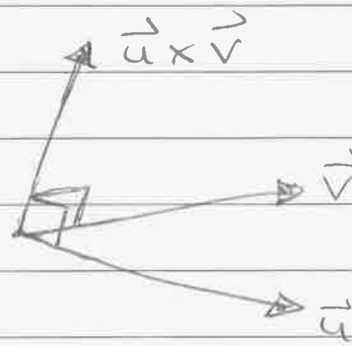
In general, given two vectors

$$\vec{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \quad \& \quad \vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$$

in $\mathbb{R}^3$, their cross product is defined by

$$\vec{u} \times \vec{v} = \begin{pmatrix} u_2 v_3 - u_3 v_2 \\ u_3 v_1 - u_1 v_3 \\ u_1 v_2 - u_2 v_1 \end{pmatrix}$$

The important property of the vector $\vec{u} \times \vec{v}$ is that it is perpendicular to both $\vec{u}$ & $\vec{v}$:



]

So let's define the vector $\vec{a} = (1, -2, 1)$ and try to compute the matrix that projects orthogonally onto the line $t\vec{a}$. To do this we observe that

$$\vec{a} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$$

itself is a matrix whose column space is the desired line. So the projection matrix is given by the formula

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

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

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

$$= \frac{1}{6}\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}\begin{pmatrix} 1 & -2 & 1 \end{pmatrix}$$

$$= \frac{1}{6}\begin{pmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{pmatrix}.$$

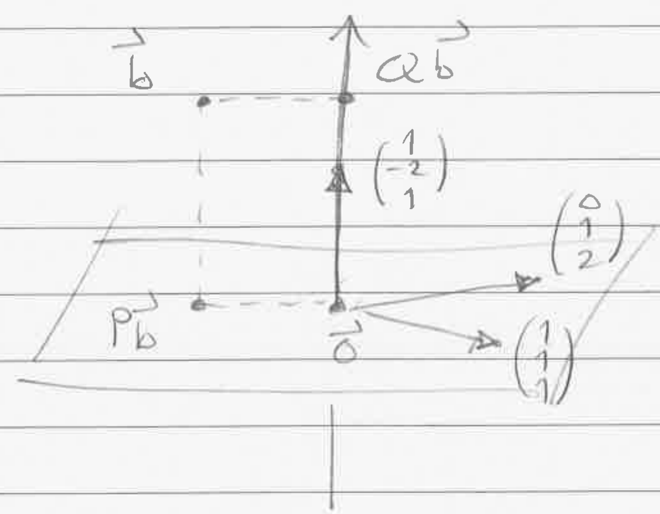
Next we observe that

$$P + Q = \frac{1}{6}\begin{pmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{pmatrix} + \frac{1}{6}\begin{pmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{pmatrix}$$

$$= \frac{1}{6}\begin{pmatrix} 5+1 & 2-2 & -1+1 \\ 2-2 & 2+4 & 2-2 \\ -1+1 & 2-2 & 5+1 \end{pmatrix}$$

$$= \frac{1}{6}\begin{pmatrix} 6 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = I_3.$$

Is that a surprise? NO. Let $\vec{b}$ be any point in $\mathbb{R}^3$ and consider the projections $P\vec{b}$ & $Q\vec{b}$ onto the plane and its orthogonal line:



Note that the points $\vec{0}$, $P\vec{b}$, $Q\vec{b}$, $\vec{b}$ are the four vertices of a 2D rectangle living in $\mathbb{R}^3$. Thus the "parallelogram Law" of vector addition tells us that

$$\vec{b} = P\vec{b} + Q\vec{b} = (P+Q)\vec{b}$$

Then since this is true for all points $\vec{b}$ it follows that

$$P + Q = I. \quad ///$$

This is a general phenomenon: the projection matrices onto a pair of "orthogonal subspaces" will always sum to the identity matrix. This is a good trick because it means we only have to compute one of the matrices; then we get the other for free.

Example: Let $\vec{a}$ be any nonzero vector in $\mathbb{R}^n$. The projection matrix onto the line $t\vec{a}$ is given by

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

Note that $\vec{a}^T\vec{a}$ is just a 1×1 number so we can move it to the front,

$$P = \underbrace{\left(\frac{1}{\vec{a}^T\vec{a}}\right)}_{\text{number}} \underbrace{\vec{a}\,\vec{a}^T}_{n\times n \text{ matrix}}$$

Or to be fancy we could write it as

$$P = \frac{\vec{a}\,\vec{a}^T}{\vec{a}^T\vec{a}} \quad \begin{matrix} n\times n \text{ matrix} \\ 1\times 1 \text{ number.} \end{matrix}$$

Then the projection matrix onto the "hyperplane orthogonal to $\vec{a}$" is given by

$$Q = I_n - P = I_n - \frac{\vec{a}\vec{a}^T}{\vec{a}^T\vec{a}} .$$

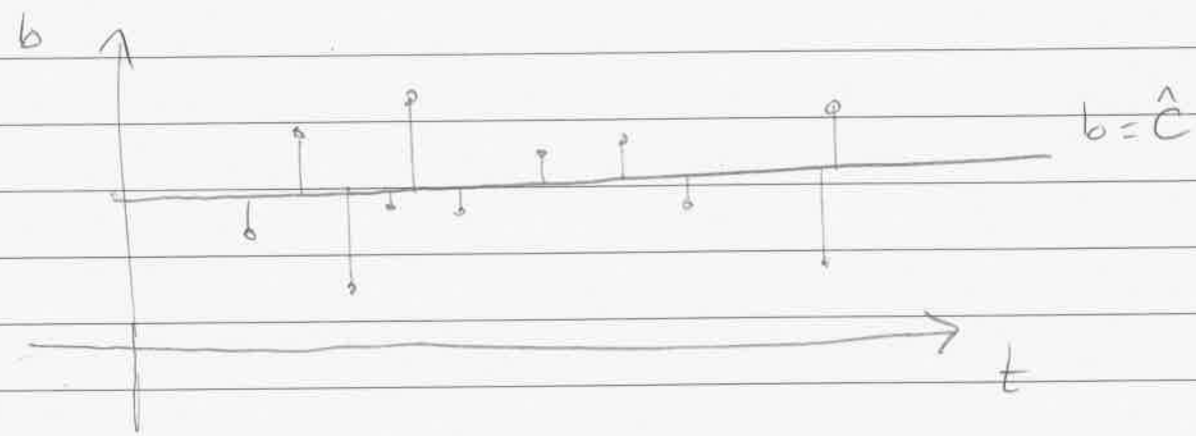
Note that we got this for free, i.e., without having to find a matrix A whose column space is the hyperplane. That's pretty good!

Old Homework Discussion

In Problem 4 we considered the problem of fitting a horizontal line to a bunch of data points

$$\begin{pmatrix} t_1 \\ b_1 \end{pmatrix}, \begin{pmatrix} t_2 \\ b_2 \end{pmatrix}, \ldots, \begin{pmatrix} t_n \\ b_n \end{pmatrix}.$$

Picture:



For example, these could be observations of a certain made at times $t_1, t_2, \ldots, t_n$. If we assume that the quantity never changes (say it's the acceleration due to gravity at the earth's surface) then the times of the observations don't matter.

The "silly equation" is

$$\begin{cases} C = b_1 \\ C = b_2 \\ \vdots \\ C = b_n \end{cases} \longrightarrow \underbrace{\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}}_{A} \underbrace{C}_{\vec{x}} = \underbrace{\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}}_{\vec{b}}$$

which has no solution (that's why it's so silly). Instead we solve the "normal equation"

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

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

$$n \hat{C} = \sum_{i=1}^{n} b_i .$$

and hence $\hat{C} = \dfrac{\sum_{i=1}^{n} b_i}{n}$.

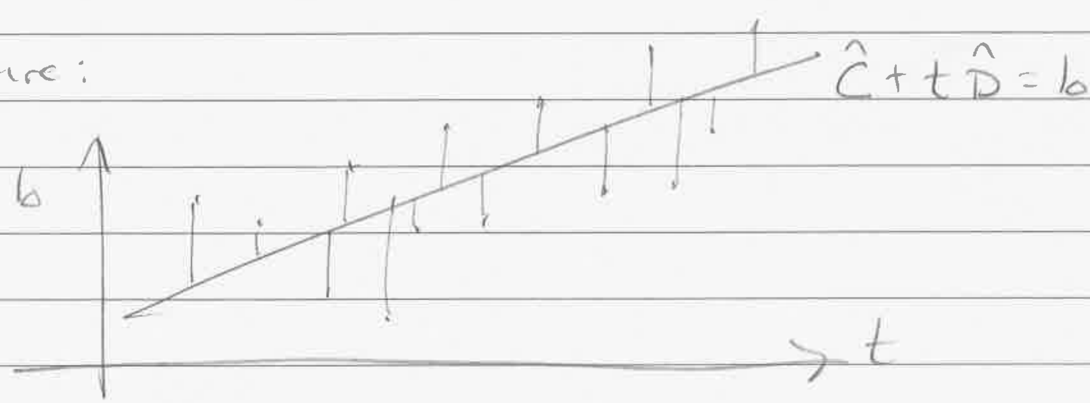
Note that this is just the average value of our observations. The error vector

$$\vec{e} = \vec{b} - \vec{a}\hat{C} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \hat{C} = \begin{pmatrix} b_1 - \hat{C} \\ b_2 - \hat{C} \\ \vdots \\ b_n - \hat{C} \end{pmatrix}$$

encodes the deviation of each observation from the average. [These are the vertical bars in the picture.]

Now suppose we're observing a quantity that does change with time (or maybe we want to test if the acceleration due to gravity changes with time). In this case we want to find the general line $\hat{C} + t\hat{D} = b$ that is the best fit for our data.

Picture:



(Again, we want to minimize the sum of the squares of the vertical errors). This time the "silly equation" is

$$\begin{cases} C + t_1 D = b_1 \\ C + t_2 D = b_2 \\ \vdots \\ C + t_n D = b_n \end{cases} \longrightarrow \underbrace{\begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} C \\ D \end{pmatrix}}_{\vec{x}} = \underbrace{\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}}_{\vec{b}}$$

which still has no solution. So instead we solve the "normal equation"

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

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_n \end{pmatrix} \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix} \begin{pmatrix} \hat{C} \\ \hat{D} \end{pmatrix} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_n \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}.$$

$$\begin{pmatrix} n & \sum_{i=1}^{n} t_i \\ \sum_{i=1}^{n} t_i & \sum_{i=1}^{n} (t_i)^2 \end{pmatrix} \begin{pmatrix} \hat{C} \\ \hat{D} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} b_i \\ \sum_{i=1}^{n} t_i b_i \end{pmatrix}.$$

$\downarrow$

This translates to the following system of two linear equations in two unknowns:

$$\begin{cases} n\hat{c} + \left(\sum_{i=1}^{n} t_i\right)\hat{D} = \sum_{i=1}^{n} b_i \\ \left(\sum_{i=1}^{n} t_i\right)\hat{c} + \left(\sum_{i=1}^{n} (t_i)^2\right)\hat{D} = \sum_{i=1}^{n} t_i b_i \end{cases}$$

In statistics these are called the "normal equations" and you are usually asked to memorize them.

Now you see that this memorization is unnecessary. All you will ever need to know is the following recipe:

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

This includes the "best fit line" as a special case, and does a lot more besides. It is probably the most useful ($) thing you will learn in this class.