4/13/16

HW 7 is due next Wed Apr 20.
Exam 2 is this Friday in class.

Today: Review for Exam 2.

The exam will cover the material from HW 5 & 6 and the corresponding Course Notes. First we need to remember the properties of matrix multiplication.

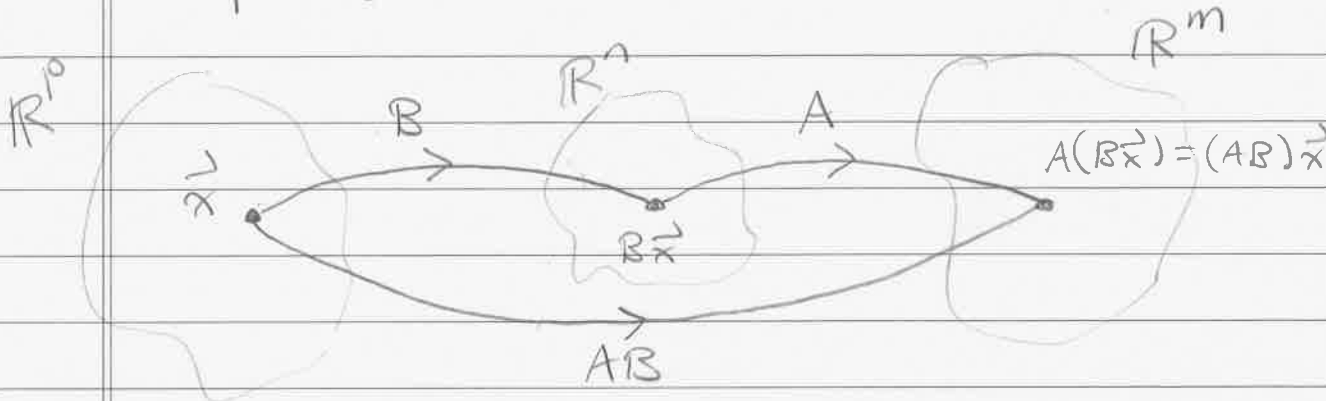
Let A be an $m \times n$ matrix and let B be an $n \times p$ matrix then the $m \times p$ product matrix AB satisfies

- $(i,j)^{th}$ entry of AB = (ith row A)(jth col B)
- (ith row of AB) = (ith row A) B
- (jth col of AB) = A (jth col B).

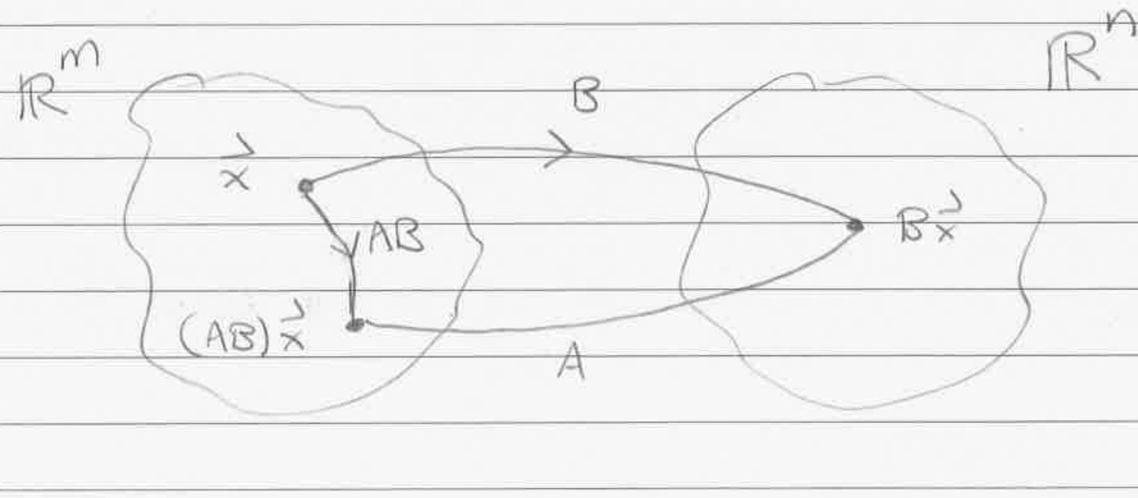
[Note that (ith row A) is a $1 \times n$ matrix and (jth col B) is an $n \times 1$ matrix. Their matrix product is the same as the good old dot product of vectors:

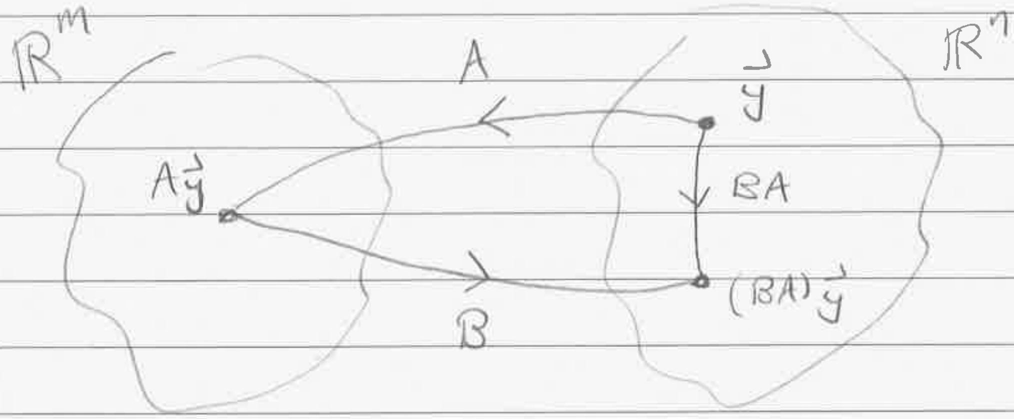
$$(i\text{th row } A)(j\text{th col } B) = (i\text{th row } A)^T \cdot (j\text{th col } B)$$
$$(1\times n \text{ matrix})(n\times 1 \text{ matrix}) = (\text{vector})\cdot(\text{vector})$$
$$1\times 1 \text{ matrix} = \text{number}$$

The matrix AB was originally <u>defined</u> by composing two functions:



In other words we have $(AB)\vec{x} = A(B\vec{x})$ for all $p\times 1$ vectors $\vec{x}$. Now if we have $p=m$ then we can also define the matrix BA and we obtain two pictures:





If both of AB & BA are "do nothing" functions, i.e., if

$$AB = I_m \quad \& \quad BA = I_n$$

then we will say that A & B are <u>inverses</u> of each other.

But I discussed in class that this is <u>impossible</u> when $m \neq n$. [Idea: If $m < n$ then RREF(A) has a non-pivot column which implies that A has a non-trivial row relation. But then if $BA = I_n$ then the formula

$$(j^{th} \text{ col } I_n) = B(j^{th} \text{ col } A)$$

implies that I_n has a non-trivial column relation, which is <u>impossible</u>.]

If $m=n$ then then the $n \times n$ matrix A <u>might</u> have an inverse. Suppose it does, i.e., suppose that there exists an $n \times n$ matrix B such that

$$AB = I_n \quad \& \quad BA = I_n.$$

To <u>compute</u> this matrix B we will solve the linear systems

$$A\,(j^{th} \text{ col } B) = (j^{th} \text{ col } I_n)$$

to get the columns of B. All of these systems can be solved simu<u>ltane</u>ously with a trick:

(*) $$\left(A \mid I_n \right) \xrightarrow{RREF} \left(I_n \mid B \right).$$

[Remark: The matrix B is unique so we call it "the" inverse of A and we write $B =$ "A^{-1}". Indeed, if we also have $AC = I_n$ & $CA = I_n$ then it follows that

$$C = C I_n = C(AB) = (CA)B = I_n B = B. \quad]$$

If we try to compute A^{-1} using (*) and it fails this means that A is <u>not</u> invertible. The reason (*) is because

$$\mathrm{RREF}(A) \neq I_n$$

and there are many things that can cause this, e.g.,

- A has a non-trivial column relation
- A has a non-trivial row relation
- $\det(A) = 0$.

You do <u>not</u> need to know the "Fundamental Theorem of Linear Algebra" for the exam. However, you <u>do</u> need to know the basic properties of matrix arithmetic, e.g.,

$$A(xB + yC) = xAB + yAC$$

$$(AB)^T = B^T A^T$$
$$(AB)^{-1} = B^{-1} A^{-1}$$
$$(A^T)^{-1} = (A^{-1})^T$$

etc.

=

We discussed "orthogonal projection" and its applications to "least squares regression".

Suppose we want to find the line $C + tD = b$ that is <u>closest</u> to the data points

$$\begin{pmatrix} t_1 \\ b_1 \end{pmatrix}, \begin{pmatrix} t_2 \\ b_2 \end{pmatrix}, \begin{pmatrix} t_3 \\ b_3 \end{pmatrix}.$$

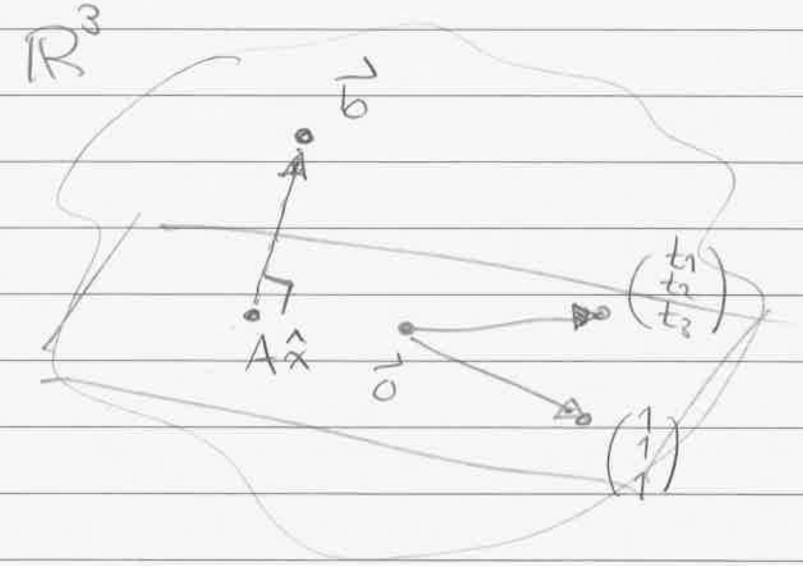
First we write down the silly equations

$$\begin{cases} C + t_1 D = b_1 \\ C + t_2 D = b_2 \\ C + t_3 D = b_3 \end{cases} \rightsquigarrow \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \end{pmatrix} \begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}.$$

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

This equation probably has no solution because the three points probably don't lie on a line. Alternatively, this means that the point $\vec{b}$ in $\mathbb{R}^3$ does not lie in the plane $A\vec{x} = C(1,1,1) + D(t_1, t_2, t_3)$ which is the "column space" of the matrix A:

$\downarrow$



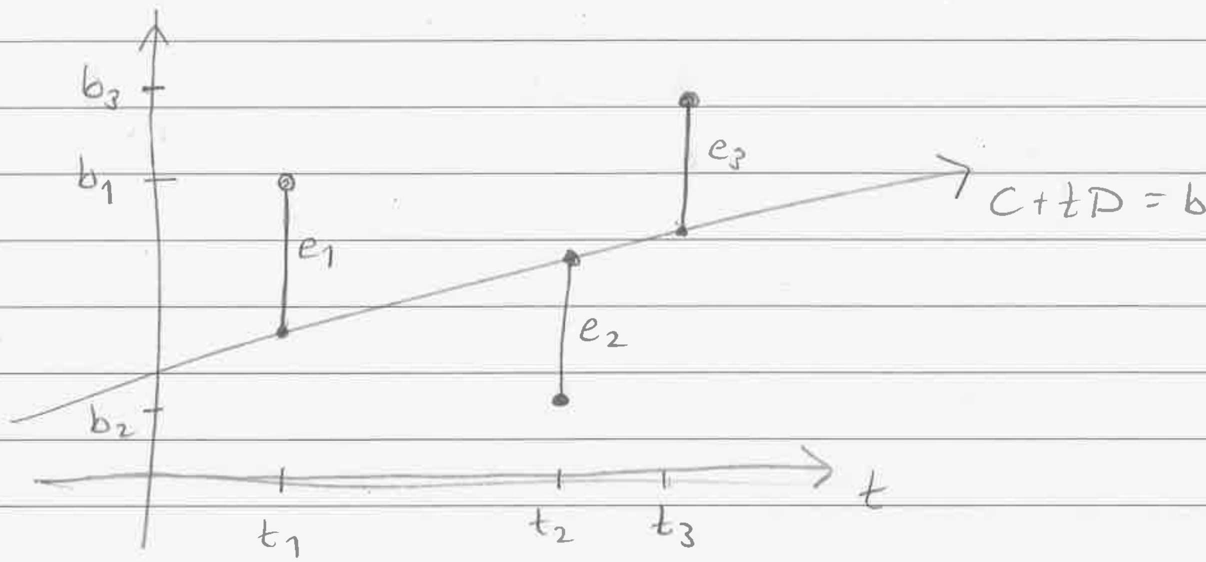
Gauss' idea is to replace $\vec{b}$ by the closest point $A\hat{x}$ in the column space. This will be accomplished when the error vector $\vec{e} := \vec{b} - A\hat{x}$ is perpendicular to all of the columns of A (the plane in the picture). We can express this condition with one matrix equation

$$(*) \qquad A^T \vec{e} = \vec{0}$$

After substituting $\vec{e} = \vec{b} - A\hat{x}$, this $(*)$ becomes the "normal equation"

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

Now we can solve this to find $\hat{x} = (C, D)$ and hence the best fit line:



The vertical errors are the entries of the error vector:

$$\vec{e} = \vec{b} - A\hat{x} = \begin{pmatrix} b_1 - (C + t_1 D) \\ b_2 - (C + t_2 D) \\ b_3 - (C + t_3 D) \end{pmatrix} = \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}$$

and the best fit line is "best" in the sense that

$$\|\vec{e}\|^2 = e_1^2 + e_2^2 + e_3^2$$

is as small as possible. ///

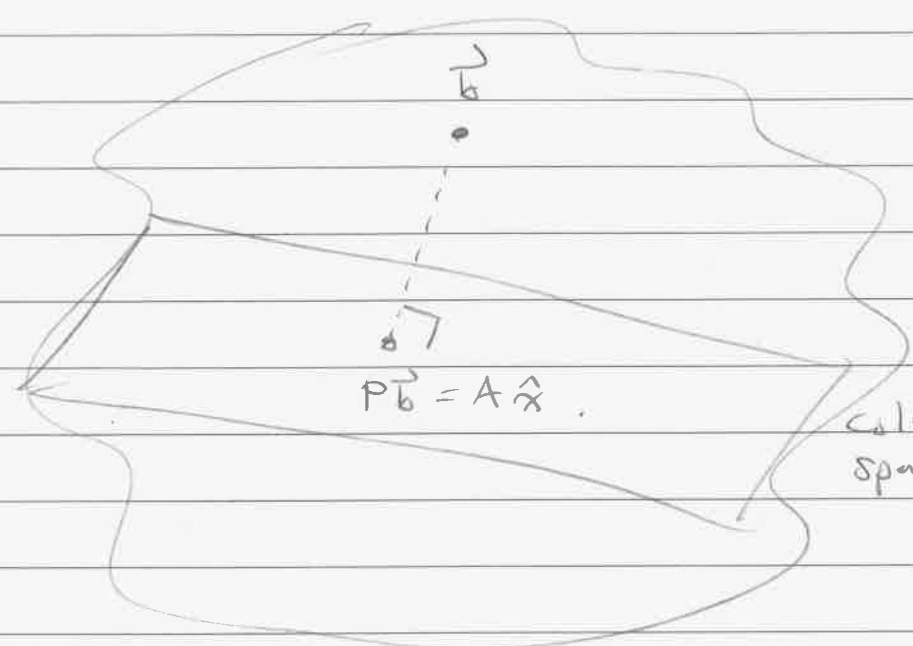
Least squares regression applies to a much broader range of problems. Given (almost) any linear system $A\vec{x} = \vec{b}$ with no solution, you can find the "best approximate solution" $\hat{x}$ by solving

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

The geometry behind this is to project $\vec{b}$ orthogonally onto the column space of A by using the projection matrix

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

Picture:



column space of A