Most important ideas:

- The least squares solution $\hat{\vec{x}} = (A^T A)^{-1} A^T \vec{b}$ to the problem $A\vec{x} = \vec{b}$.
- Projecting the vector $\vec{b}$ onto the column space of A *non-orthogonal* columns.

Reminder: Where $A = [\,\vec{a}_1\ \ \vec{a}_2 \cdots\ \ \vec{a}_n\,]$, $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, we can write

$$A\vec{x} = \vec{b} \quad \text{as} \quad [\,\vec{a}_1\ \ \vec{a}_2 \cdots\ \ \vec{a}_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \vec{b}, \quad \text{i.e.,} \quad x_1\vec{a}_1 + x_2\vec{a}_2 + \cdots x_n\vec{a}_n = \vec{b},$$

that is, $\vec{b}$ is a linear combination of the columns of A, that is, $\vec{b}$ is in the column space of A.

Recall: $\vec{b}$ is in the column space of A means there is some $\vec{x}$ so that $\vec{b} = A\vec{x}$.

Example 1: Is $\begin{bmatrix} 5 \\ 4 \\ 3 \end{bmatrix}$ in $Col\ A$ where $A = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$? Yes, since $\begin{bmatrix} 5 \\ 4 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} -3 \\ 2 \end{bmatrix}$.

Question: What if there is no solution to $A\vec{x} = \vec{b}$, that is, what if $\vec{b}$ is not in the column space of A?

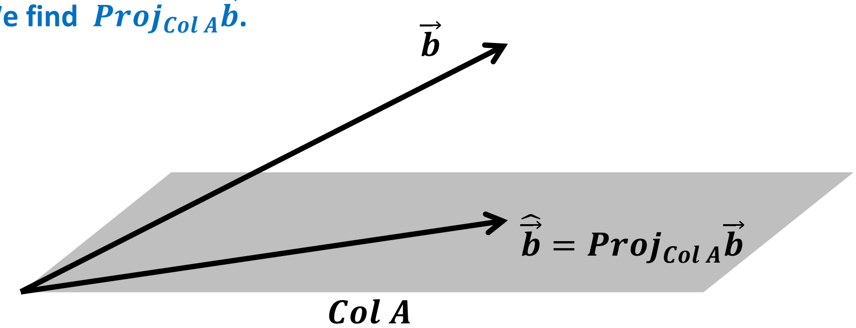
Answer: **We do the best we can.**

Question: What does that mean?

Answer: **We find the vector $\hat{\vec{b}}$ in $Col\ A$ that is closest to $\vec{b}$.**

Question: How do we do that?

Answer: **We find $Proj_{Col\ A}\vec{b}$.**



Question: So you mean that where $m \times n$ $A = [\,\vec{a}_1\ \ \vec{a}_2 \cdots\ \ \vec{a}_n]$ we have

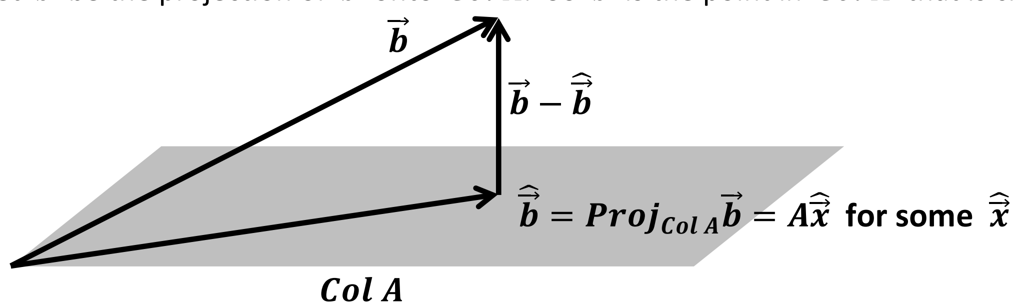
$$Proj_{Col\ A}\vec{b} = \frac{\vec{b} \cdot \vec{a}_1}{\vec{a}_1 \cdot \vec{a}_1}\vec{a}_1 + \frac{\vec{b} \cdot \vec{a}_2}{\vec{a}_2 \cdot \vec{a}_2}\vec{a}_2 + \cdots + \frac{\vec{b} \cdot \vec{a}_n}{\vec{a}_n \cdot \vec{a}_n}\vec{a}_n\,?$$

Answer: **No. This would be true only if the columns of A were orthogonal. (See the example at the bottom of Handout 6.2 page 3.)**

Question: So how do we find $Proj_{Col\,A}\vec{b}$ <u>if the columns of A are not orthogonal</u>?

Answer: **Great question. Glad you asked.**

Notation: Let $\hat{\vec{b}}$ be the projection of $\vec{b}$ onto $Col\,A$. So $\hat{\vec{b}}$ is the point in $Col\,A$ that is closest to $\vec{b}$.



Note: if $\hat{\vec{b}}$ is in the column space of A, then there must be some $\hat{\vec{x}}$ such that $\hat{\vec{b}} = A\hat{\vec{x}}$.

Question: So how do you find $\hat{\vec{b}}$?

Answer: **You actually find the $\hat{\vec{x}}$ so that $A\hat{\vec{x}} = \hat{\vec{b}}$.**

Question: So how do you find $\hat{\vec{x}}$?

Answer:

As shown in the diagram above, find $\hat{\vec{b}}$ so that $\vec{b} - \hat{\vec{b}} \perp Col\,A$.

That is, find $\hat{\vec{x}}$ so that $\vec{b} - A\hat{\vec{x}} \perp Col\,A$.

That is, find $\hat{\vec{x}}$ so that $\vec{b} - A\hat{\vec{x}} \perp$ each column of A.

So where $A = [\ \vec{a}_1\ \vec{a}_2\ \cdots\ \vec{a}_n\ ]$, find $\hat{\vec{x}}$ so that

$$\begin{aligned} \vec{a}_1^T(\vec{b} - A\hat{\vec{x}}) &= 0 \\ \vec{a}_2^T(\vec{b} - A\hat{\vec{x}}) &= 0 \\ &\vdots \\ \vec{a}_n^T(\vec{b} - A\hat{\vec{x}}) &= 0 \end{aligned}$$

which can be written as a matrix $\times$ a vector $\begin{bmatrix} \vec{a}_1^T \\ \vec{a}_2^T \\ \vdots \\ \vec{a}_n^T \end{bmatrix}(\vec{b} - A\hat{\vec{x}}) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$,

that is, $A^T(\vec{b} - A\hat{\vec{x}}) = \vec{0}$, that is, $A^T\vec{b} - A^TA\hat{\vec{x}} = \vec{0}$, that is, $A^TA\hat{\vec{x}} = A^T\vec{b}$.

So we have discovered that $A\vec{x} = \vec{b}$ can be modified (left multiply both sides by A^T) to be

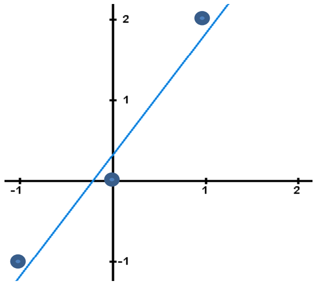
$$A^TA\hat{\vec{x}} = A^T\vec{b}$$

which leads to

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

assuming that A^TA has an inverse, which it usually (but not always) does.

Example: Find the line $y = mx + b$ that best fits the points $(-1,-1), (0,\ 0), (1,\ 2)$.



That is, find m and b so that

$$\begin{aligned} -1 \cdot m + b &= -1 \\ 0 \cdot m + b &= \ \ 0 \\ 1 \cdot m + b &= \ \ 2 \end{aligned}$$

That is

$$\begin{bmatrix} -1 & 1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}$$

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

$$\boldsymbol{A^T A} = \begin{bmatrix} -1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}, \quad \boldsymbol{A^T \vec{b}} = \begin{bmatrix} -1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix},$$

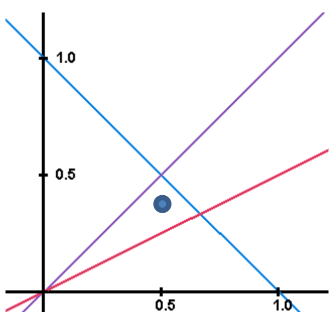
$$\Rightarrow \quad \begin{bmatrix} \boldsymbol{m} \\ \boldsymbol{b} \end{bmatrix} = \widehat{\vec{\boldsymbol{x}}} = (\boldsymbol{A^T A})^{-1} \boldsymbol{A^T} \widehat{\vec{\boldsymbol{b}}} = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/3 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 3/2 \\ 1/3 \end{bmatrix},$$

so the best fit line is $\boldsymbol{y = \frac{3}{2} x + \frac{1}{3}}$, as seen in the diagram above.

Why did we suspect that there is not a solution (that is, not an exact solution) to this problem?
There are more equations (3) than unknowns (2). It is overdetermined.

Another way to think about this: we get to choose two parameters, m and b, but there are three restrictions that must be met: the three points the line must fit. So the system is overdetermined (it is the restrictions that *determine* the values of the parameters).

Example: Find the point (x, y) that is on (or at least closest to) the lines



$$\begin{aligned} x + \ \ y &= 1 \\ x - \ \ y &= 0 \\ x - 2y &= 0 \end{aligned}, \quad \text{that is,} \quad \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

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

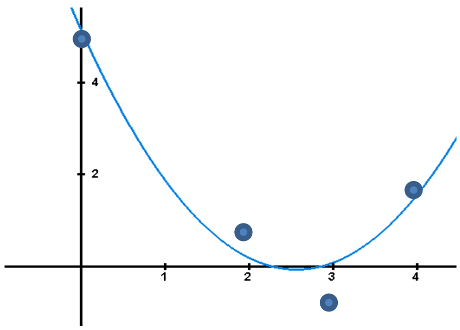
$$\boldsymbol{A^T A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & -2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & -2 \end{bmatrix} = \begin{bmatrix} 3 & -2 \\ -2 & 6 \end{bmatrix}, \quad \boldsymbol{A^T \vec{b}} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

$$\Rightarrow \quad \widehat{\vec{\boldsymbol{x}}} = (\boldsymbol{A^T A})^{-1} \boldsymbol{A^T} \widehat{\vec{\boldsymbol{b}}} = \frac{1}{(3)(6)-(-2)(-2)} \begin{bmatrix} 6 & 2 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 8/14 \\ 5/14 \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix},$$

So the point that is closest to the three lines is (8/14, 5/14), as seen in the diagram above.

Example: find the quadratic $y = a_0 + a_1 x + a_2 x^2$ that best fits that data

$$(4,2), (2,\ 1), (0,\ 5), (3, -1).$$



That is, find a_0 , a_1, a_2 so that

$$\begin{aligned} a_0 + a_1(4) + a_2(4^2) &= 2 \\ a_0 + a_1(2) + a_2(2^2) &= 1 \\ a_0 + a_1(0) + a_2(0^2) &= 5 \\ a_0 + a_1(3) + a_2(3^2) &= -1 \end{aligned}$$

That is

$$\begin{bmatrix} 1 & 4 & 16 \\ 1 & 2 & 4 \\ 1 & 0 & 0 \\ 1 & 3 & 9 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 5 \\ -1 \end{bmatrix}$$

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

$$\boldsymbol{A^T A} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 4 & 2 & 0 & 3 \\ 16 & 4 & 0 & 9 \end{bmatrix} \begin{bmatrix} 1 & 4 & 16 \\ 1 & 2 & 4 \\ 1 & 0 & 0 \\ 1 & 3 & 9 \end{bmatrix} = \begin{bmatrix} 4 & 9 & 29 \\ 9 & 29 & 99 \\ 29 & 99 & 353 \end{bmatrix}, \quad \boldsymbol{A^T \vec{b}} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 4 & 2 & 0 & 3 \\ 16 & 4 & 0 & 9 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 5 \\ -1 \end{bmatrix} = \begin{bmatrix} 7 \\ 7 \\ 27 \end{bmatrix}$$

so (using technology)

$$\widehat{\vec{\boldsymbol{x}}} = (\boldsymbol{A^T A})^{-1} \boldsymbol{A^T} \widehat{\vec{\boldsymbol{b}}} = \begin{bmatrix} 4 & 9 & 29 \\ 9 & 29 & 99 \\ 29 & 99 & 353 \end{bmatrix}^{-1} \begin{bmatrix} 7 \\ 7 \\ 27 \end{bmatrix} \approx \begin{bmatrix} 5.136 \\ -4.068 \\ 0.795 \end{bmatrix} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}.$$

So the quadratic polynomial that best fits the data is $\boldsymbol{5.136 - 4.068x + .795x^2}$.
It's not a great fit, but given the data, it's the best we can do.

You can see how this would generalize to finding a higher degree polynomial to fit several data points.

Question: To what types of problems does this least squares process apply?

Answer: **Overdetermined systems, that is, systems with more equations than unknowns.**

Question: Well what about underdetermined systems in which # equations < # unknowns?

Answer: **Remember that when there are fewer equations than unknowns there is an infinite number of solutions—assuming there is a solution at all.**

And of course: **If the number of equations is equal to the number of unknowns, there is typically exactly one solution, but as we've seen there can be exceptions.**

Example: Find the function $y = e^{a+bx} = e^a e^{bx} = Ce^{bx}$ that best fits the data $(1,1.9), (2,4), (3,8)$. Note that there are two parameters a and b that we get to choose, but three restrictions (three points to fit), so this system is <u>overdetermined</u>. Let's work this in class.

Question: What if A is square and A (and thus also A^T) has an inverse?

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

Let's find a general formula for the least squares <u>line</u> $y = mx + b$ that best fits given data

$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.

Find m and b so that

$$\begin{aligned} x_1 \cdot m + b &= y_1 \\ x_2 \cdot m + b &= y_2 \\ &\vdots \\ x_n \cdot m + b &= y_n \end{aligned}$$

That is

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Then $A^T A = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} = \begin{bmatrix} \sum x^2 & \sum x \\ \sum x & \sum 1 \end{bmatrix} = \begin{bmatrix} \sum x^2 & \sum x \\ \sum x & n \end{bmatrix}$

and $A^T \vec{b} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum xy \\ \sum y \end{bmatrix}$

Then $\begin{bmatrix} m \\ b \end{bmatrix} = (A^T A)^{-1} A^T \vec{b} = \begin{bmatrix} \sum x^2 & \sum x \\ \sum x & n \end{bmatrix}^{-1} \begin{bmatrix} \sum xy \\ \sum y \end{bmatrix} = \frac{1}{(\sum x^2)\, n - (\sum x)(\sum x)} \begin{bmatrix} n & -\sum x \\ -\sum x & \sum x^2 \end{bmatrix} \begin{bmatrix} \sum xy \\ \sum y \end{bmatrix}$

So the slope $m = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$

and y–intercept $b = \frac{-(\sum x)(\sum xy) + (\sum x^2)(\sum y)}{n(\sum x^2) - (\sum x)^2}$

Another view of finding the least squares solution. For a given matrix A and a given right hand side $\vec{b}$, the best (i.e. the least squares solution) to $A\vec{x} = \vec{b}$ is the $\vec{x}$ which minimizes $\|\vec{b} - A\vec{x}\|$. That is, if we can't find a vector $\vec{x}$ so that $A\vec{x} = \vec{b}$ exactly, find the $\hat{\vec{x}}$ so that $A\hat{\vec{x}}$ is as close as possible to $\vec{b}$. The $\vec{x}$ which minimizes $\|\vec{b} - A\vec{x}\|$ is also the the $\vec{x}$ which minimizes

$$f(\vec{x}) = \|\vec{b} - A\vec{x}\|^2$$

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

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

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

$$= \vec{b}^T\vec{b} - 2\vec{x}^T A^T\vec{b} - \vec{x}^T A^T A\vec{x} \text{ since } \vec{b}^T A\vec{x} = \vec{x}^T A^T\vec{b} \text{ (both are simply numbers)}$$

Then $f'(\vec{x}) = 2A^T\vec{b} - 2A^T A\vec{x}$ so that $f'(\vec{x}) = \vec{0} \Rightarrow 2A^T\vec{b} - 2A^T A\vec{x} = \vec{0}$, i.e. $A^T A\vec{x} = A^T\vec{b}$, which of course is the same equation we found earlier using Linear Algebra (projecting $\vec{b}$ onto the column space of A).

Question: It seemed in the examples we've done that A^TA is always symmetric? Is that the case?

Answer: **Yes.** $(A^TA)^T = A^T(A^T)^T = A^TA.$

Question: Does A^TA always have an inverse?

Short Answer: Only if the columns of A are linearly independent, which is typically the case.

A bit more detail. First, if A is $m \times n$, then A^T is $n \times m$, and A^TA is $n \times n$. So for A^TA to be invertible, it must have full rank of n.

Since $rank\ AB \leq rank\ A$ (or B) and since $rank\ A^T = rank\ A$, then $\boldsymbol{rank\ A^TA \leq rank\ A}$. Matrix A is $m \times n$, where $m > n$, so we have $rank\ A^TA \leq rank\ A \leq \min(m, n) = n$. So for A^TA to have an inverse, it must be that $rank\ A = n$ (recall that A is $m \times n$). It turns out that if $rank\ A = n$ (if all n columns of A are linearly independent), then $rank\ A^TA = n$.

Example: Find the line $y = mx + b$ that best fits the points

$$(4,1),(4,2),(4,3).$$

Let's work this in class.

Also: See HW 6.5.25.

Finally: Suppose we have the QR factorization of A where Q is orthogonal and R is (square) upper triangular and invertible.

Then $A^TA = (QR)^T(QR) = R^TQ^TQR = R^TIR = R^TR$ and $A^T\vec{b} = (QR)^T\vec{b} = R^TQ^T\vec{b}$.

So $A^TA\hat{\vec{x}} = A^T\vec{b}$ becomes $R^TR\hat{\vec{x}} = R^TQ^T\vec{b}$

which (multiplying both sides by $(R^T)^{-1}$) leads to $R\hat{\vec{x}} = Q^T\vec{b}$

which is fairly easy to solve where R is upper triangular.

One final note: There is also multilinear regression if you are trying to fit functions of more than one variable to data from R^3 or higher. So instead of fitting something like

$$y = a_0 + a_1x + a_2x^2 \quad \text{to} \quad (x_1,y_1),(x_2,y_2),\dots,(x_n,y_n)$$

we would fit something like

$$z = a_0 + a_1x + a_2y + a_3x^2 + a_4xy + a_5y^2 \quad \text{to} \quad (x_1,y_1,z_1),(x_2,y_2,z_2),\dots,(x_n,y_n,z_n).$$

Remember you can use technology (e.g. Excel) as appropriate to do the computation.