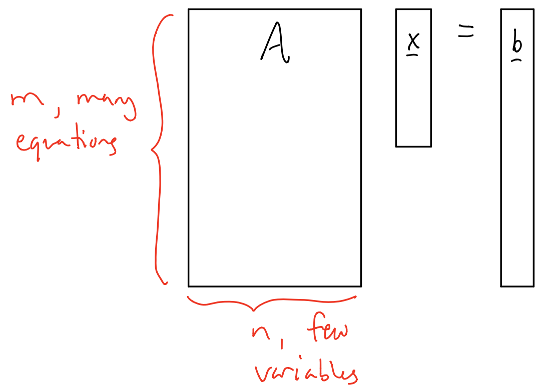
Sect 6.5 (Cont...)

Recall: We were interested in problems where there are more equations than unknowns. We called these problems over-determined



Often, there is no $\underline{x}$ that exact satisfies this system, so we try find an $\underline{x}$ such that

$$A\underline{x} \approx \underline{b}$$

So, we a close or approximate solution. We define "close" in terms of the norm;

$$\| \underline{b} - A\underline{x} \|$$

If this norm is small, the $\underline{x}$ is a better <u>least-squares solution</u>.

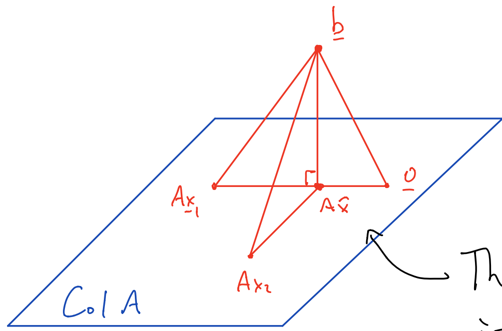
Def

If A is $m \times n$, and $\underline{b} \in \mathbb{R}^m$, a least-squares sol. of $A\underline{x} = \underline{b}$ is $\hat{\underline{x}} \in \mathbb{R}^n$ such that

$$\| \underline{b} - A\hat{x} \| \leq \| \underline{b} - A\underline{x} \|$$

for all $\underline{x} \in \mathbb{R}^n$.

Geometrically, this looks like



The least-squares solution $\hat{x}$ is the vector such that $A\hat{x} \in \mathrm{Col}(A)$ is closest to $\underline{b}$.

From Sect. 6.3, we know that closest point in $W = \mathrm{Col}\, A$ to a vector $\underline{b}$ is the orth. proj. of $\underline{b}$ onto W, given by

$$\hat{b} = \mathrm{proj}_{\mathrm{Col}\, A}\, \underline{b} = \mathrm{proj}_W\, \underline{b}$$

Since $\hat{b} \in \mathrm{Col}\, A$, the equation $A\underline{x} = \hat{\underline{b}}$ is consistent and there exists some $\hat{x}$ such that $A\hat{x} = \hat{\underline{b}}$

Note, that $\underline{b} - \hat{\underline{b}} = \underline{b} - A\hat{x}$ is orthogonal to every column of A, so we have

$$a_i^T (\underline{b} - A\hat{x}) = 0$$

Taking this over every column of A, we get

$$A^T (\underline{b} - A\hat{x}) = 0$$

$$A^T\underline{b} - A^T A \hat{x} = 0$$

$$A^T A \hat{x} = A^T \underline{b} \qquad (*)$$

Thus, any least-squares solution to $A\underline{x} = \underline{b}$ satisfies $(*)$, which we call the <u>normal equations</u>

<u>Thm</u>

The set of <u>least-squares solutions</u> to $A\underline{x} = \underline{b}$ is the non-empty set of solutions to the normal equations: $A^T A \underline{x} = A^T \underline{b}$

<u>Observations</u>

- $A^T A$ is $n \times n$ and $A^T \underline{b} \in \mathbb{R}^n$
- Normal equations are unstable on a computer (learn about this in higher level courses)
- The solution set for the Normal Equations is non-empty, i.e, always consistent!

Ex

Find a least-squares solution to $A\underset{\sim}{x} = \underline{b}$ for

$$A = \begin{vmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{vmatrix} \qquad \underline{b} = \begin{vmatrix} 2 \\ 0 \\ 11 \end{vmatrix}$$

Set up the normal equations

$$A^T A = \begin{vmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{vmatrix} \begin{vmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{vmatrix} = \begin{vmatrix} 17 & 1 \\ 1 & 5 \end{vmatrix}$$

$$A^T \underline{b} = \begin{vmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{vmatrix} \begin{vmatrix} 2 \\ 0 \\ 11 \end{vmatrix} = \begin{vmatrix} 19 \\ 11 \end{vmatrix}$$

So the normal equations are

$$A^T A \underset{\sim}{x} = A^T \underline{b}$$

$$\begin{vmatrix} 17 & 1 \\ 1 & 5 \end{vmatrix} \begin{vmatrix} x_1 \\ x_2 \end{vmatrix} = \begin{vmatrix} 19 \\ 11 \end{vmatrix}$$

The solution is

$$\underline{x} = (A^T A)^{-1} A^T \underline{b} = \frac{1}{84} \begin{vmatrix} 5 & -1 \\ -1 & 17 \end{vmatrix} \begin{vmatrix} 19 \\ 11 \end{vmatrix}$$

$$= \frac{1}{84} \begin{vmatrix} 84 \\ 168 \end{vmatrix} = \begin{vmatrix} 1 \\ 2 \end{vmatrix}$$

Note, here $A^T A$ is invertible. This may not always be the case. Use row reduction if necessary.

Thm

Let A be an $m \times n$ matrix, then the following are equivalent:

1.) $A\underline{x} = \underline{b}$ has unique least-squares solution for each $\underline{b}$ in $\mathbb{R}^m$

2.) The columns of A are linearly independent

3.) The matrix $A^T A$ is invertible.

When these are true, the least-squares solution is given by

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

Note

- The distance of $A\hat{x}$ to $\underline{b}$ is called the least-squares error of the solution

$$\| \underline{b} - A\hat{x} \|$$

This measures how well the solution $\hat{x}$ fits the system of equations $A\underline{x} = \underline{b}$

Connection to QR-factorization (Sect 6.4)

The normal equations are not the only way to find a least-squares solution. Another common way uses the QR factorization

Thm

Given an $m \times n$ matrix A with linearly independent columns, let $A = QR$ where Q is an $m \times n$ orthogonal matrix and R is an $n \times n$ upper triangular matrix and invertible, then for each $\underline{b} \in \mathbb{R}^m$, $A\underline{x} = \underline{b}$ has a unique least-squares solution given by

$$\hat{x} = R^{-1} Q^T \underline{b}$$

Pf. Start from $A\underline{x} = \underline{b}$. We know a least-squares satisfies the normal equations

$$A^T A \underline{x} = A^T \underline{b}$$

Substitute in $A = QR$, we get

$$(QR)^T (QR) \underline{x} = (QR)^T \underline{b}$$

$$R^T \underbrace{Q^T Q}_{= I} R \underline{x} = R^T Q^T \underline{b}$$

$$R^T R \underline{x} = R^T Q^T \underline{b}$$

$$R\underline{x} = \underbrace{(R^T)^{-1} R^T}_{= I} Q^T \underline{b}$$

$$R\underline{x} = Q^T \underline{b}$$

$$\hat{\underline{x}} = R^{-1} Q^T \underline{b}$$

Ex

Let $A = \begin{vmatrix} 1 & 3 & 5 \\ 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 3 & 3 \end{vmatrix}$ $\quad \underline{b} = \begin{vmatrix} 3 \\ 5 \\ 7 \\ -3 \end{vmatrix}$

The QR factorization is

$$A = QR = \begin{vmatrix} 1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & -1/2 \\ 1/2 & -1/2 & 1/2 \\ 1/2 & 1/2 & -1/2 \end{vmatrix} \begin{vmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{vmatrix}$$

Then $Q^T \underline{b} = \begin{vmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & -1/2 & 1/2 \\ 1/2 & -1/2 & 1/2 & -1/2 \end{vmatrix} \begin{vmatrix} 3 \\ 5 \\ 7 \\ -3 \end{vmatrix} = \begin{vmatrix} 6 \\ -6 \\ 4 \end{vmatrix}$

Lastly back substitute to solve $R\hat{x} = Q^T \underline{b}$ and get

$$\begin{vmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{vmatrix} \begin{vmatrix} x_1 \\ x_2 \\ x_3 \end{vmatrix} = \begin{vmatrix} 6 \\ -6 \\ 4 \end{vmatrix}$$

$$x_3 = 2$$

$$x_2 = \frac{-6 - 3(2)}{2} = -6 \qquad \Rightarrow \quad \underline{\hat{x}} = \begin{vmatrix} 10 \\ -6 \\ 2 \end{vmatrix}$$

$$x_1 = \frac{6 - 5(2) - 4(-6)}{2} = 10$$

Notes:

- QR factorization is usually prefered in computer implementations. It's more stable than the normal equations.