June 11, 2014

Least Squares Problems

- First posed and formulated by Gauss.

- Surveyors tried to identified boundaries lie between plots and land by measuring certain angles and distances from known landmarks.

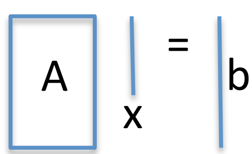
- To update the location of landmarks, new measurements of angles and distances between landmarks.

- Given a set of old locations $\{(x_i, y_i)\}$, find corrections $\{(\delta x_i, \delta y_i)\}$ such that $\{(x_i+\delta x_i, y_i+\delta y_i)\}$ better match new measurements.

- In general,

$$\text{constraints} \left\{ \begin{array}{c} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n = b_1 \\ \vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n = b_m \end{array} \right\} \text{observations}$$

i.e. $\quad A\,x = b$



In general, $r = b - A\,x \neq 0$.

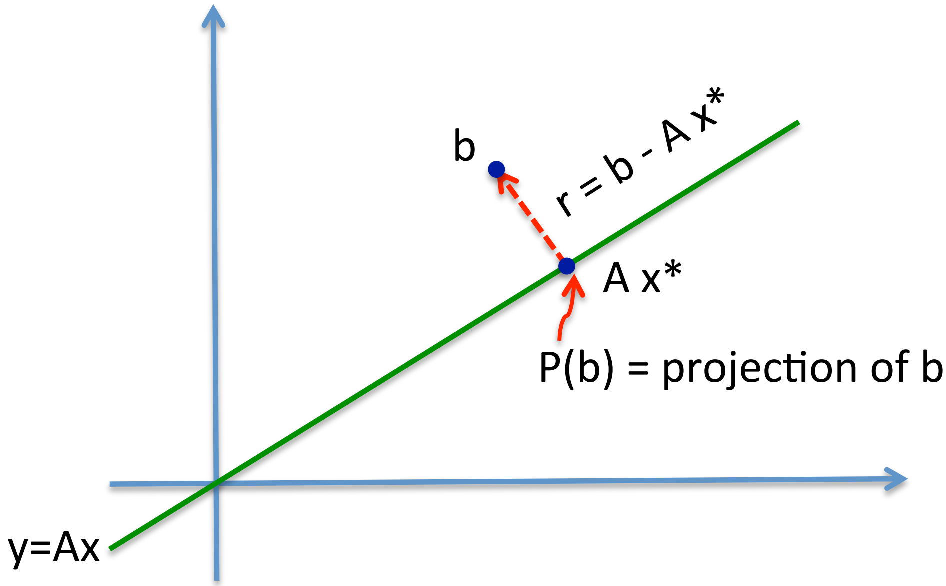
Idea: min the residual vector r

$$\min_{x \in R^n} \|b - Ax\|_2^2 \qquad A \in R^{m \times n},\ b \in R^m,\ m \geq n$$

- Least squares (LS) problems.

Solving LS problems

Geometric interpretation:



Theorem: Let $A \in R^{m x n}$, $b \in R^m$, $m \geq n$ and A has full rank. A vector $x \in R^n$ minimizes

$$|| r ||^2 = || b - A x ||^2$$

if and only if $r \perp \text{range}(A)$.

Hence $\quad r^T A = 0 \iff A^T r = 0 \iff A^T (b - A x) = 0$

$$\iff A^T A x = A^T b$$

Def: $A^+ = (A^T A)^{-1} A^T$ is called the pseudoinverse of A.

The least squares solution is given by:

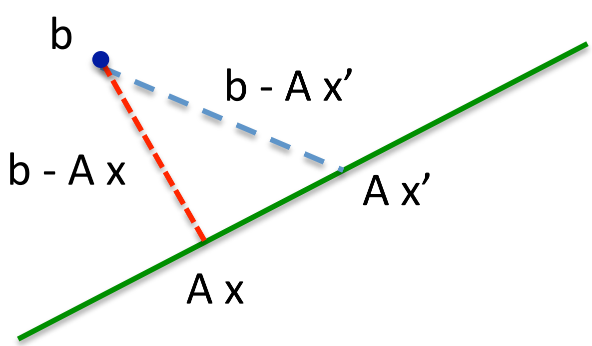
$$x = (A^T A)^{-1} A^T b = A^+ b$$

Why is x the minimizer of $||b-Ax||^2$?

Let $x' = x + e$ be another point.

$$
\begin{aligned}
||b - A x'||^2 &= (b - A x')^T (b - A x') \\
&= (b - A x - A e)^T (b - A x - A e) \\
&= (b - A x)^T (b - A x) + 2 (A e)^T (b - A x) + (Ae)^T (Ae) \\
&= ||b - A x||^2 + ||Ae||^2 + 2e^T \underbrace{(A^T b - A^T A x)}_{0} \\
&= ||b - A x||^2 + ||Ae||^2
\end{aligned}
$$

Hence $\quad ||b - A x'||^2 > ||b - A x||^2 \qquad$ if $e \neq 0$



Method 1: Normal equations

Solve $A^TA\,x = A^T b$

- Compute Cholesky factorization $A^TA = GG^T$, G = lower Δ
- Compute x by forward and backward solves

Complexity

flops(A^TA) ~ mn^2, flops(GG^T) ~ $1/3\,n^3$

Total flops ~ $mn^2 + 1/3\,n^3$ $\qquad$ ($m \geq n$)

Method 2: QR factorization

Def: Q is orthogonal if $Q^{-1} = Q^T$

i.e. $Q^T Q = Q Q^T = I$

Theorem: $||Q x||_2 = ||x||_2$

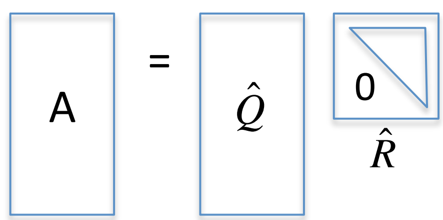
Pf: $||Q x||^2 = (Q x)^T (Q x) = x^T Q^T Q x = x^T x = ||x||^2$

Note:

$$\text{multiplication by } Q = \begin{cases} \text{rotation} & \text{if } \det(Q) = 1 \\ \text{reflection} & \text{if } \det(Q) = -1 \end{cases}$$

Theorem: Suppose $A \in R^{m \times n}$ has full rank. Then there exist a unique orthogonal matrix $\hat{Q} \in R^{m \times n}$ $(Q^T Q = I)$ and a unique upper Δ matrix $\hat{R} \in R^{n \times n}$ with positive diagonals $(r_{ii} > 0)$ such that

$$A = \hat{Q}\hat{R}$$ i.e.



Note:

The column of $\hat{Q}$ are orthogonal to each other and their norm = 1.

Consider the LS problem:

$$\min \ || A x - b ||^2$$

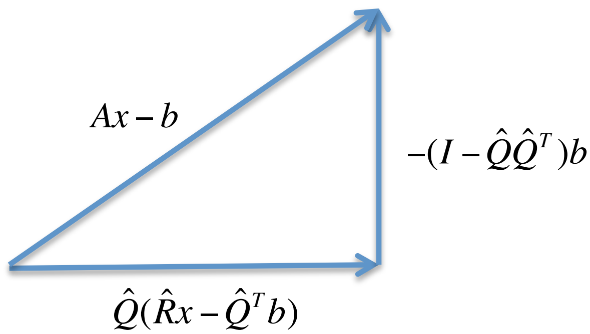
- $Ax - b = \hat{Q}\hat{R}x - b$

$$= \hat{Q}\hat{R}x - (\hat{Q}\hat{Q}^T + I - \hat{Q}\hat{Q}^T)b$$

$$= \hat{Q}(\hat{R}x - \hat{Q}^T b) - (I - \hat{Q}\hat{Q}^T)b$$

Note: $\hat{Q}(\hat{R}x-\hat{Q}^T b) \quad \perp \quad (I-\hat{Q}\hat{Q}^T)b$

Pf: $$[\hat{Q}(\hat{R}x-\hat{Q}^T b)]^T (I-\hat{Q}\hat{Q}^T)b$$
$$=(\hat{R}x-\hat{Q}^T b)\hat{Q}^T (I-\hat{Q}\hat{Q}^T)b$$
$$=(\hat{R}x-\hat{Q}^T b)(\hat{Q}^T-\hat{Q}^T\hat{Q}\hat{Q}^T)b \qquad (\because \hat{Q}^T\hat{Q}=I)$$



Pythagoras thm: $$\|Ax-b\|^2=\left\|\hat{Q}(\hat{R}x-\hat{Q}^T b)\right\|^2+\left\|(I-\hat{Q}\hat{Q}^T)b\right\|^2$$
$$=\left\|\hat{R}x-\hat{Q}^T b\right\|^2+\left\|(I-\hat{Q}\hat{Q}^T)b\right\|^2$$

The RHS is min if the first term is 0.

i.e. $$\hat{R}x=\hat{Q}^T b \quad\Rightarrow\quad x=\hat{R}^{-1}\hat{Q}^T b$$

<u>Notes</u>

1) $A^+=\hat{R}^{-1}\hat{Q}^T$

2) $A^T Ax=A^T b \quad\Leftrightarrow\quad (\hat{R}^T\hat{Q}^T)(\hat{Q}\hat{R})x=(\hat{R}^T\hat{Q}^T)b$
$$\hat{R}^T\hat{R}x=\hat{R}^T\hat{Q}^T b$$
$$\hat{R}x=\hat{Q}^T b$$
$$x=\hat{R}^{-1}\hat{Q}^T b$$

QR factorization (reduced version)

Let $A = [a_1 \ a_2 \ \ldots \ a_n]$. Want to find orthogonal vectors $\{ q_i \}$ such that

$$\text{span}\{ q_1, \ldots, q_j \} = \text{span}\{ a_1, \ldots, a_j \} \qquad j = 1, 2, \ldots, n.$$

This amounts to:

$$\begin{bmatrix} \\ a_1 & a_2 & \cdots & a_n \\ \\ \end{bmatrix} = \begin{bmatrix} \\ q_1 & q_2 & \cdots & q_n \\ \\ \end{bmatrix} \begin{bmatrix} r_{11} & \cdots & r_{1n} \\ & \ddots & \vdots \\ & & r_{nn} \end{bmatrix}$$

i.e.

$$a_1 = r_{11} q_1$$

$$a_2 = r_{12} q_1 + r_{22} q_2$$

$$\vdots$$

Matrix form: $A = \hat{Q}\hat{R}$

- $\hat{Q}$ has orthonormal columns, $\hat{R}$ upper Δ.

QR factorization (full version)

- Append additional m-n orthonormal columns to

  i.e. $[q_{n+1} \ \ q_{n+2} \ \ \cdots \ \ q_m] \equiv \hat{Q}_{m-n}$

  Then

$$\begin{bmatrix} A \end{bmatrix}_{m\times n} = \begin{bmatrix} \hat{Q} & \hat{Q}_{m-n} \end{bmatrix}_{m\times m} \begin{bmatrix} \hat{R} \\ 0 \end{bmatrix}_{m\times n} \Big\} \text{m-n zero rows}$$

- Usually for theoretical purpose.