CS475 / CS675
Lecture 10: June 2, 2016

Least Squares Problems

Reading: [TB] Chapt 11

Least Squares Problems

- First posed and formulated by Gauss.

- Surveyors tried to identify boundaries by measuring certain angles and distances from known landmarks.

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

Surveying Example

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

Surveying Example

- Non-linear constraint:

$$\cos^2 \theta_i = measurement$$
$$= \frac{\left[\left(z_j - z_i\right)^T (z_k - z_i)\right]^2}{\left(z_j - z_i\right)^T \left(z_j - z_i\right)(z_k - z_i)^T (z_k - z_i)}$$

- Suppose $(\delta x_i, \delta y_i) \ll (x_i, y_i)$
  - Multiply through the denominator
  - Multiply out all the terms to get a quartic polynomial in all δ-variables
  - Throw away all terms containing $\delta^2, \delta^3, \delta^4$

  $\Longrightarrow$ linear constraint

Surveying example

- Collect all linear constraints for all angles and distance measurements

  $\Longrightarrow$ overdetermined linear system

- In general:

$$\text{constraints}\left\{\begin{array}{c} a_n x_1 + a_{12} x_2 + \cdots + a_{1n} x_n = b_1 \\ \vdots \\ a_{m1} x_1 + a_{m2} x_2 + \cdots + a_{mn} x_n = b_m \end{array}\right\}\text{observations}$$

Picture: $Ax = b$

Overdetermined system

- In general: $r = b - Ax \neq 0$
- Idea: minimize the residual vector r
  - Optimization problem:
- Least squares (LS) problems

Solving LS Problems

- Geometric interpretation:

Solving LS Problems

- <u>Theorem:</u> Let $A \in \Re^{m \times n}, b \in \Re^m, m \geq n.$ A vector $x \in \Re^n$ minimizes
$$||r||_2^2 = ||b - Ax||_2^2$$
if and only if $r \perp range(A)$

- $\therefore r^T A = 0 \Leftrightarrow A^T r = 0$
$$\Leftrightarrow A^T(b - Ax) = 0$$
$$\Leftrightarrow A^T A x = A^T b$$

Pseudoinverse

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

- The least squares solution is given by

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

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

Pseudoinverse

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

$$\left|\left|b - Ax'\right|\right|_2^2 = (b - Ax')^T(b - Ax')$$
$$= (b - Ax - Ae)^T(b - Ax - Ae)$$
$$= (b - Ax)^T(b - Ax) + 2(Ae)^T(b - Ax) + (Ae)^T Ae$$
$$= \left|\left|b - Ax\right|\right|^2 + \left|\left|Ae\right|\right|^2 + 2e^T(A^T b - A^T Ax)$$
$$= \left|\left|b - Ax\right|\right|^2 + \left|\left|Ae\right|\right|^2$$

$$\therefore \left|\left|b - Ax'\right|\right|^2 > \left|\left|b - Ax\right|\right|^2 \quad \text{if } e \neq 0$$

Pseudoinverse

- Picture:

Method 1: Normal Equations

- Solve $A^T A x = A^T b$
- Compute Cholesky factorization
  - i.e., $A^T A = G G^T$, $G =$ lower Δ

$$G G^T x = A^T b$$

- Compute x by forward and backward solves

- Complexity:
  - $flops(A^T A) \sim mn^2$, $flops(G G^T) \sim n^3/3$

    $\therefore total\ flops \sim mn^2 + n^3/3 \qquad (m \geq n)$

Method 2: QR Factorization

- Def: Q is orthogonal if $Q^{-1} = Q^T$
  - i.e., $Q^T Q = Q Q^T = I$
- <u>Theorem:</u> $||Qx||_2 = ||x||_2$
- <u>Proof:</u>

Orthogonal Q

- Note: multiplication by Q

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

- Picture:

QR Factorization (reduced version)

- Let $A = [a_1\ a_2\ \ \dots a_n]$. Want to find orthonormal vectors $\{q_i\}$ such that
  $$span\{q_1, \dots, q_j\} = span\{a_1, \dots, a_j\} \quad j = 1,2, \dots, n$$
- This amounts to:

- Matrix form: $A = \hat{Q}\hat{R}$
  - $\hat{Q}$ has orthonormal columns, $\hat{R}$ is upper Δ

QR Factorization (full version)

- Append additional $m - n$ orthogonal cols to $\hat{Q}$
  - i.e., $[q_{n+1}\ q_{n+2}\ \ldots q_m] \equiv \hat{Q}_{m-n}$
- Then

- Usually for theoretical purpose

QR Factorization

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

- Note: Cols of $\hat{Q}$ are orthogonal to each other and their norm $= 1$

QR Factorization

- Consider the LS problem:

$$\min_x \left|\left|Ax - b\right|\right|^2$$

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

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

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

QR Factorization

- Note: $\hat{Q}(\hat{R}x - \hat{Q}^T b) \perp \left(I - \hat{Q}\hat{Q}^T\right)b$
- Proof: $\left[\hat{Q}\left(\hat{R}x - \hat{Q}b\right)\right]^T \left(I - \hat{Q}\hat{Q}^T\right)b$

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

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

$= 0$ (since $\hat{Q}^T\hat{Q} = I$)

- Picture:

QR Factorization

- Pythagoras theorem:

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

- The RHS is minimized if the first term is 0
  - i.e., $\hat{R}x = \hat{Q}^T b \Longrightarrow x = \hat{R}^{-1}\hat{Q}^T b$
- Notes
  1. $A^\dagger = \hat{R}^{-1}\hat{Q}^T$
  2. $A^T A x = A^T b \iff \left(\hat{R}^T\hat{Q}^T\right)\left(\hat{Q}\hat{R}\right)x = \left(\hat{R}^T\hat{Q}^T\right)b$

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