18-660: Numerical Methods for Engineering Design and Optimization

Xin Li
Department of ECE
Carnegie Mellon University
Pittsburgh, PA 15213

Electrical & Computer ENGINEERING

Overview

- Over-determined Linear Equation Solver
  - Pseudo-inverse
  - QR decomposition

Over-Determined Linear Equation

- Fit an approximate function f(x) from sampling points

$$f(x) \approx \alpha_1 \cdot b_1(x) + \alpha_2 \cdot b_2(x) + \cdots$$

M samples

$$\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix} \cdot \alpha = \begin{bmatrix} \\ B \\ \\ \end{bmatrix}$$

N coefficients

(M > N)
Over-determined linear equation

Over-Determined Linear Equation

- Solve over-determined linear equation
  - No exact solution to satisfy all equations, but we can find the least-squares solution:

$$A \cdot \alpha = B$$

$$\Downarrow$$

$$\min_{\alpha} \quad \left\| A \cdot \alpha - B \right\|_2^2$$

(Least-squares solution)

Over-Determined Linear Equation

- There are two popular approaches to solve over-determined linear equations

M samples

$$\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix} \cdot \alpha = \begin{bmatrix} \\ B \\ \\ \end{bmatrix} \quad (M > N)$$

N coefficients

Over-Determined Linear Equation

- Solution 1

$$A \cdot \alpha = B$$

$$\Downarrow$$

$$\underbrace{\left[A^T \right]}_{N \times M} \cdot \underbrace{\left[A \right]}_{M \times N} \cdot \alpha = \underbrace{\left[A^T \right]}_{N \times M} \cdot \underbrace{\left[B \right]}_{M \times 1}$$

Over-Determined Linear Equation

- Solution 1

$$\left[\underset{N \times M}{A^T} \right] \cdot \left[\underset{M \times N}{A} \right] \cdot \alpha = \left[\underset{N \times M}{A^T} \right] \cdot \left[B \right]_{M \times 1}$$

$$\alpha = \underbrace{\left(A^T A\right)^{-1}}_{N \times N} \cdot \underbrace{\left(A^T B\right)}_{N \times 1}$$

We can prove that the solution yields least-squared error

Over-Determined Linear Equation

- Proof of optimality

$$\min_{\alpha} \quad \left\| A \cdot \alpha - B \right\|_2^2$$

$$F(\alpha) = \left\| A \cdot \alpha - B \right\|_2^2 = (A \cdot \alpha - B)^T \cdot (A \cdot \alpha - B)$$

$$F(\alpha) = \left\| A \cdot \alpha - B \right\|_2^2 = \left(\alpha^T \cdot A^T - B^T\right) \cdot (A \cdot \alpha - B)$$

$$F(\alpha) = \alpha^T A^T A \alpha - \alpha^T A^T B - B^T A \alpha + B^T B$$

Over-Determined Linear Equation

- Proof of optimality

$$F(\alpha) = \alpha^T A^T A\alpha - \alpha^T A^T B - B^T A\alpha + B^T B$$

$$\alpha^T A^T B = \left(\alpha^T A^T B\right)^T = B^T A\alpha$$

$$F(\alpha) = \alpha^T A^T A\alpha - 2B^T A\alpha + B^T B$$

$$\frac{\partial}{\partial\alpha} F(\alpha) = \frac{\partial}{\partial\alpha}\left(\alpha^T A^T A\alpha\right) - \frac{\partial}{\partial\alpha}\left(2B^T A\alpha\right) = 0$$

Over-Determined Linear Equation

- Proof of optimality

$$\frac{\partial}{\partial \alpha} F(\alpha) = \frac{\partial}{\partial \alpha}\left(\alpha^T \underbrace{A^T A}_{W} \alpha\right) - \frac{\partial}{\partial \alpha}\left(2 \underbrace{B^T A}_{P^T} \alpha\right) = 0$$

$$W = A^T A \qquad \alpha^T W \alpha = \sum_i \sum_j w_{ij} \alpha_i \alpha_j$$

$$\frac{\partial}{\partial \alpha_n}\left(\alpha^T W \alpha\right) = \frac{\partial}{\partial \alpha_n}\left(w_{nn}\alpha_n^2 + \sum_{i \neq n} w_{in}\alpha_i \alpha_n + \sum_{j \neq n} w_{nj}\alpha_n \alpha_j\right)$$

$$= 2w_{nn}\alpha_n + \sum_{i \neq n} w_{in}\alpha_i + \sum_{j \neq n} w_{nj}\alpha_j$$

$$= 2w_{nn}\alpha_n + \sum_{j \neq n} 2w_{nj}\alpha_j$$

$$= \sum_j 2w_{nj}\alpha_j$$

Over-Determined Linear Equation

- Proof of optimality

$$\frac{\partial}{\partial \alpha} F(\alpha) = \frac{\partial}{\partial \alpha}\left(\alpha^T \underbrace{A^T A}_{W} \alpha\right) - \frac{\partial}{\partial \alpha}\left(2 \underbrace{B^T A}_{P^T} \alpha\right) = 0$$

$$W = A^T A$$

$$\frac{\partial}{\partial \alpha_n}\left(\alpha^T W \alpha\right) = \sum_j 2 w_{nj} \alpha_j$$

$$\frac{\partial}{\partial \alpha}\left(\alpha^T W \alpha\right) = \begin{bmatrix} \frac{\partial}{\partial \alpha_1}\left(\alpha^T W \alpha\right) \\ \frac{\partial}{\partial \alpha_2}\left(\alpha^T W \alpha\right) \\ \vdots \end{bmatrix} = \begin{bmatrix} \sum_j 2 w_{1j} \alpha_j \\ \sum_j 2 w_{2j} \alpha_j \\ \vdots \end{bmatrix} = 2W\alpha = 2A^T A \alpha$$

Over-Determined Linear Equation

- Proof of optimality

$$\frac{\partial}{\partial \alpha} F(\alpha) = \frac{\partial}{\partial \alpha}\left(\alpha^T \underbrace{A^T A}_{W} \alpha\right) - \frac{\partial}{\partial \alpha}\left(2 \underbrace{B^T A}_{P^T} \alpha\right) = 0$$

$$P = \left(B^T A\right)^T = A^T B \qquad 2P^T \alpha = \sum_i 2 p_i \alpha_i$$

$$\frac{\partial}{\partial \alpha_n}\left(2P^T \alpha\right) = \frac{\partial}{\partial \alpha_n}\left(\sum_i 2 p_i \alpha_i\right) = 2 p_i$$

$$\frac{\partial}{\partial \alpha}\left(2P^T \alpha\right) = \begin{bmatrix} \frac{\partial}{\partial \alpha_1}\left(2P^T \alpha\right) \\ \frac{\partial}{\partial \alpha_2}\left(2P^T \alpha\right) \\ \vdots \end{bmatrix} = \begin{bmatrix} 2p_1 \\ 2p_2 \\ \vdots \end{bmatrix} = 2P = 2A^T B$$

Over-Determined Linear Equation

- Proof of optimality

$$\frac{\partial}{\partial \alpha} F(\alpha) = \frac{\partial}{\partial \alpha}\left(\alpha^T \underbrace{A^T A}_{W} \alpha\right) - \frac{\partial}{\partial \alpha}\left(2 \underbrace{B^T A}_{P^T} \alpha\right) = 0$$

$$\frac{\partial}{\partial \alpha}\left(\alpha^T W \alpha\right) = 2A^T A\alpha \qquad \frac{\partial}{\partial \alpha}\left(2P^T \alpha\right) = 2A^T B$$

$$2A^T A\alpha - 2A^T B = 0$$

$$\alpha = \left(A^T A\right)^{-1} \cdot \left(A^T B\right)$$

Over-Determined Linear Equation

- Solution 2

$$A \cdot \alpha = B$$

QR decomposition

$$A = \begin{bmatrix} \\ Q \\ \text{M x N} \\ \\ \end{bmatrix} \cdot \begin{bmatrix} \times & \times & \times \\ & R & \times \\ & \text{N x N} & \times \end{bmatrix}$$

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

Upper triangular

QR Decomposition

$$\begin{bmatrix} A_1 & A_2 & \cdots \end{bmatrix} = \begin{bmatrix} Q_1 & Q_2 & \cdots \end{bmatrix} \cdot \begin{bmatrix} r_{11} & r_{12} & \cdots \\ & r_{22} & \cdots \\ & & \ddots \end{bmatrix}$$

$$A_1 = Q_1 \cdot r_{11} \qquad Q_1^T Q_1 = 1$$

$$\|A_1\|_2 = r_{11} \cdot \|Q_1\|_2$$

$$r_{11} = \|A_1\|_2$$

$$Q_1 = \frac{A_1}{r_{11}}$$

QR Decomposition

$$\begin{bmatrix} A_1 & A_2 & \cdots \end{bmatrix} = \begin{bmatrix} Q_1 & Q_2 & \cdots \end{bmatrix} \cdot \begin{bmatrix} r_{11} & r_{12} & \cdots \\ & r_{22} & \cdots \\ & & \ddots \end{bmatrix}$$

$$A_2 = Q_1 \cdot r_{12} + Q_2 \cdot r_{22} \quad Q_2^T Q_2 = 1 \quad Q_1^T Q_2 = 0$$

$$Q_1^T A_2 = Q_1^T Q_1 \cdot r_{12} + Q_1^T Q_2 \cdot r_{22}$$

$$r_{12} = Q_1^T A_2$$

QR Decomposition

$$\begin{bmatrix} A_1 & A_2 & \cdots \end{bmatrix} = \begin{bmatrix} Q_1 & Q_2 & \cdots \end{bmatrix} \cdot \begin{bmatrix} r_{11} & r_{12} & \cdots \\ & r_{22} & \cdots \\ & & \ddots \end{bmatrix}$$

$$A_2 = Q_1 \cdot r_{12} + Q_2 \cdot r_{22} \qquad Q_2^T Q_2 = 1 \qquad Q_1^T Q_2 = 0$$

$$A_2 - r_{12} Q_1 = Q_2 r_{22}$$

$$\left\| A_2 - r_{12} Q_1 \right\|_2 = r_{22} \cdot \left\| Q_2 \right\|_2$$

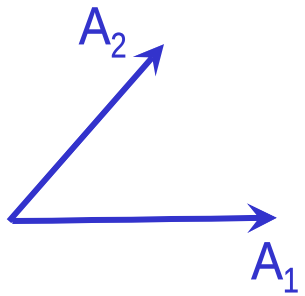
$$r_{22} = \left\| A_2 - r_{12} Q_1 \right\|_2$$

$$Q_2 = \frac{A_2 - r_{12} Q_1}{r_{22}}$$

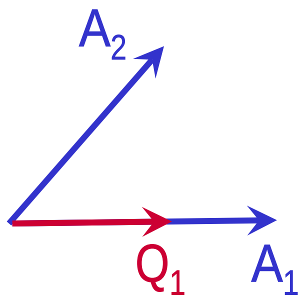
Continue iteration until Q and R are found

QR Decomposition

- Geometrical interpretation



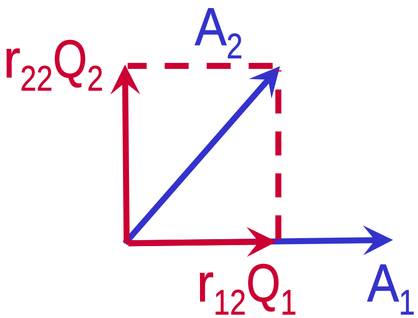
Start from two vectors A_1 and A_2



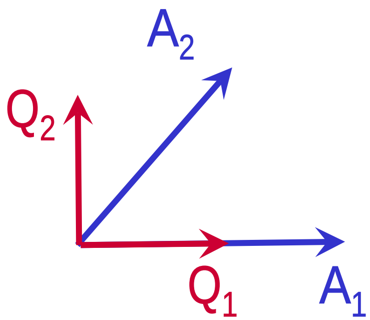
Normalized A_1 to determine Q_1

QR Decomposition

- Geometrical interpretation



Decompose A_2 into $r_{12}Q_1$ and $r_{22}Q_2$



Normalized $r_{22}Q_2$ to determine Q_2

QR Decomposition

- It is referred to as classical Gram-Schmidt algorithm
  - Q may not be orthogonal due to numerical errors
- Modified Gram-Schmidt algorithm was proposed to further improve numerical stability
- More details can be found at

  Numerical Recipes: The Art of Scientific Computing, 2007

Over-Determined Linear Equation

- Solution 2

$$A \cdot \alpha = B \quad A = Q \cdot R$$

$$Q \cdot R \cdot \alpha = B$$

$$Q^T \cdot Q \cdot R \cdot \alpha = Q^T \cdot B$$

$$Q^T Q = I$$

$$R \cdot \alpha = Q^T \cdot B$$

$$\alpha = R^{-1} \cdot \left(Q^T B\right)$$

$$\left[\; R \;\right]$$

N x N

$$\left[\; Q^T \;\right] \cdot \left[\; B \;\right] = \left[\; Q^T B \;\right]$$

N x M

M x 1

N x 1

Over-Determined Linear Equation

- In theory, these two approaches yield identical results

$$A \cdot \alpha = B \quad A = Q \cdot R$$

$$A^T A = R^T Q^T Q R = R^T R$$

$$\left(A^T A\right)^{-1} = \left(R^T R\right)^{-1} = R^{-1}\left(R^T\right)^{-1}$$

$$A^T B = R^T Q^T B$$

$$\alpha = \left(A^T A\right)^{-1} \cdot \left(A^T B\right) = R^{-1}\left(R^T\right)^{-1} \cdot R^T Q^T B = R^{-1} Q^T B$$

$(A^TA)^{-1}A^T = R^{-1}Q^T$ is often referred to as pseudo-inverse of A

Over-Determined Linear Equation

- In practice, Solution 2 is preferred since it is more robust to numerical noise
  - MATLAB uses a modified version of solution 2 to solve over-determined linear equations
  - You can use a simple command “α = A\B” to solve over-determined linear equations in MATLAB

- To understand the numerical difference, we check the condition number of both approaches

Over-Determined Linear Equation

- A simple example based on L_1 matrix norm

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 10^{-5} \\ 10^{-10} & 10^{-10} \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 1 & 10^{-20} \\ 10^{-20} & 10^{-10} \end{bmatrix}$$

$$R = \begin{bmatrix} 1 & 10^{-20} \\ 0 & 10^{-5} \end{bmatrix}$$

$$k\left(A^T A\right) = 10^{10}$$

Solution 1

$$k\left(R\right) = 10^{5}$$

Solution 2

Over-Determined Linear Equation

- Comparison on two solutions

Solution 1 $$\alpha = \left(A^T A\right)^{-1} \cdot \left(A^T B\right)$$

Solution 2 $$\alpha = R^{-1} \cdot \left(Q^T B\right)$$

$$\Downarrow$$

$$k\left(A^T A\right) >> k(R)$$

Solution 2 is more numerically robust

Summary

- Over-determined linear equation solver
  - Pseudo-inverse
  - QR decomposition