Least squares

CE282: Linear Algebra

Computer Engineering Department

Sharif University of Technology

Hamid R. Rabiee

Maryam Ramezani

Least squares problem

Theorem

- given $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$, find vector $x \in \mathbb{R}^n$ that minimizes

$$||Ax - b||^2 = \sum_{i=1}^{m}\left(\sum_{j=1}^{n} A_{ij}x_j - b_i\right)^2$$

- ''least squares'' because we minimize a sum of squares of affine functions:

$$||Ax - b||^2 = \sum_{i=1}^{m} r_i(x)^2, \qquad r_i(x) = \sum_{j=1}^{n} A_{ij}x_j - b_i$$

- the problem is also called the linear least squares problem

Least squares and linear equations

Important

$$\text{minimize } ||Ax - b||^2$$

solution of the least squares problem: any $\hat{x}$ that satisfies

- $$||A\hat{x} - b|| \leq ||Ax - b|| \quad \text{for all } x$$

Note

$\hat{r} = A\hat{x} - b$ is the residual vector

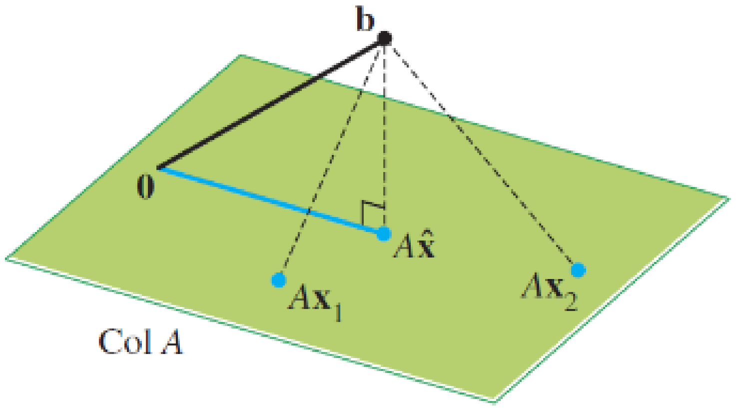
if $\hat{r} = 0$, then $\hat{x}$ solves the linear equation $Ax = b$

if $\hat{r} \neq 0$, then $\hat{x}$ is a least squares approximate solution of the equation

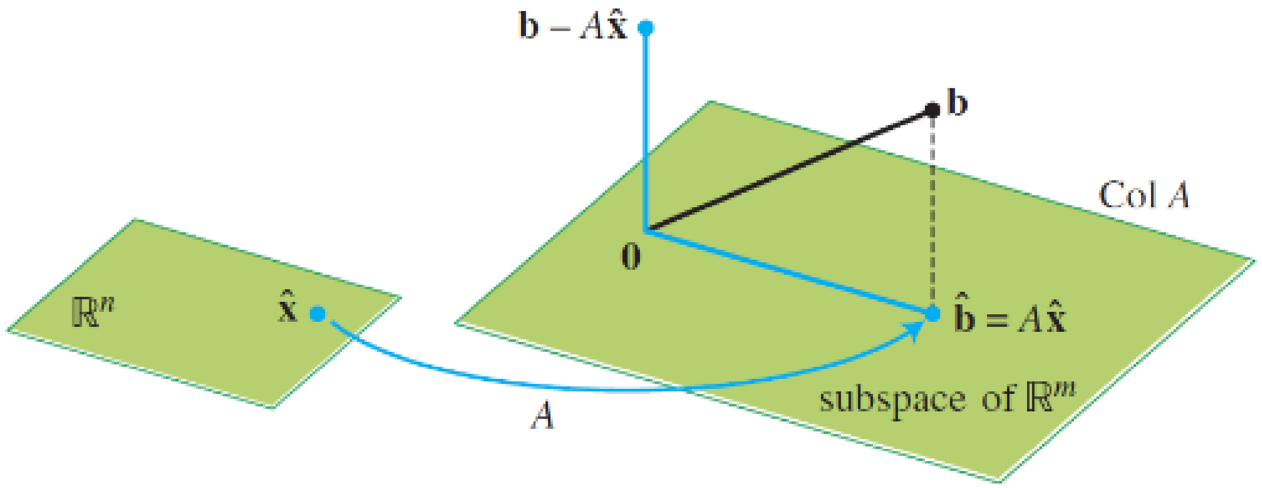
in most least squares applications, $m > n$ and $Ax = b$ has no solution

Note

The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the normal equations $A^T A\mathbf{x} = A^T\mathbf{b}$.



The vector $\mathbf{b}$ is closer to $A\hat{\mathbf{x}}$ than to $A\mathbf{x}$ for other $\mathbf{x}$.



The least-squares solution $\hat{\mathbf{x}}$ is in $\mathbb{R}^n$.

Column interpretation

Important

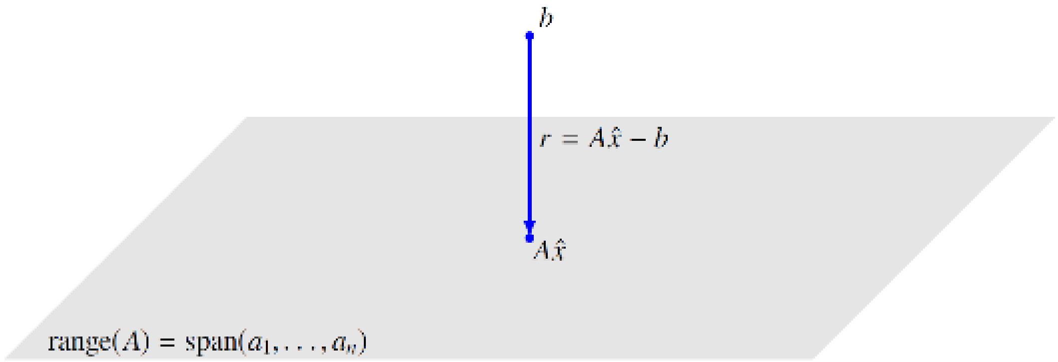
- least squares problem in terms of columns $a_1, a_2, \dots, a_n$ of A:

$$\text{minimize } \left|\left|Ax - b\right|\right|^2 = \left|\left|\sum_{j=1}^{n} a_j x_j - b\right|\right|^2$$

- The solution is closest to b among all linear combinations of columns of A

$$A\hat{x} = \hat{x}_1 a_1 + \cdots + \hat{x}_n a_n$$

- $A\hat{x}$ is the vector in range(A) = span($a_1, a_2, \dots, a_n$) closest to b
- geometric intuition suggests that $\hat{r} = A\hat{x} - b$ is orthogonal to range(A)



Row interpretation

Important

- suppose $\tilde{a}_1^T, \ldots, \tilde{a}_m^T$ are rows of A
- residual components are $r_i = \tilde{a}_i^T x - b_i$
- least squares objective is

$$||Ax - b||^2 = (\tilde{a}_1^T x - b_1)^2 + \cdots + (\tilde{a}_m^T x - b_m)^2$$

  the sum of squares of the residuals
- so least squares minimizes sum of squares of residuals
  - solving $Ax = b$ is making all residuals zero
  - least squares attempts to make them all small

Example

Example

$$A = \begin{bmatrix} 2 & 0 \\ -1 & 1 \\ 0 & 2 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

- $Ax = b$ has no solution
- least squares problem is to choose x to minimize

$$||Ax - b||^2 = (2x_1 - 1)^2 + (-x_1 + x_2)^2 + (2x_2 + 1)^2$$

- least squares approximate solution is $\hat{x} = (\frac{1}{3}, -\frac{1}{3})$ (say, via calculus)
- $||A\hat{x} - \mathrm{b}||^2 = \frac{2}{3}$ is smallest possible value of $||Ax - b||^2$
- $A\hat{x} = \left(\frac{2}{3}, -\frac{2}{3}, -\frac{2}{3}\right)$ is linear combination of columns of A closest to b

Solution of a least squares problem

Theorem

- A has linearly independent columns, then below vector is the unique solution of the least squares problem

$$\text{minimize } ||Ax - b||^2$$

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

$$= A^{\dagger} b$$

pseudo-inverse of a left-invertible matrix

- Proof?

Derivation from calculus

Important

- $f(x) = ||Ax - b||^2 = \sum_{i=1}^{m}(\sum_{j=1}^{n} A_{ij}x_j - b_i)^2$
- partial derivative of f with respect to x_k

$$\frac{\partial f}{\partial x_k}(x) = 2\sum_{i=1}^{m} A_{ik}\left(\sum_{j=1}^{n} A_{ij}x_j - b_i\right) = 2\left(A^T(Ax - b)\right)_k$$

- gradient of f is

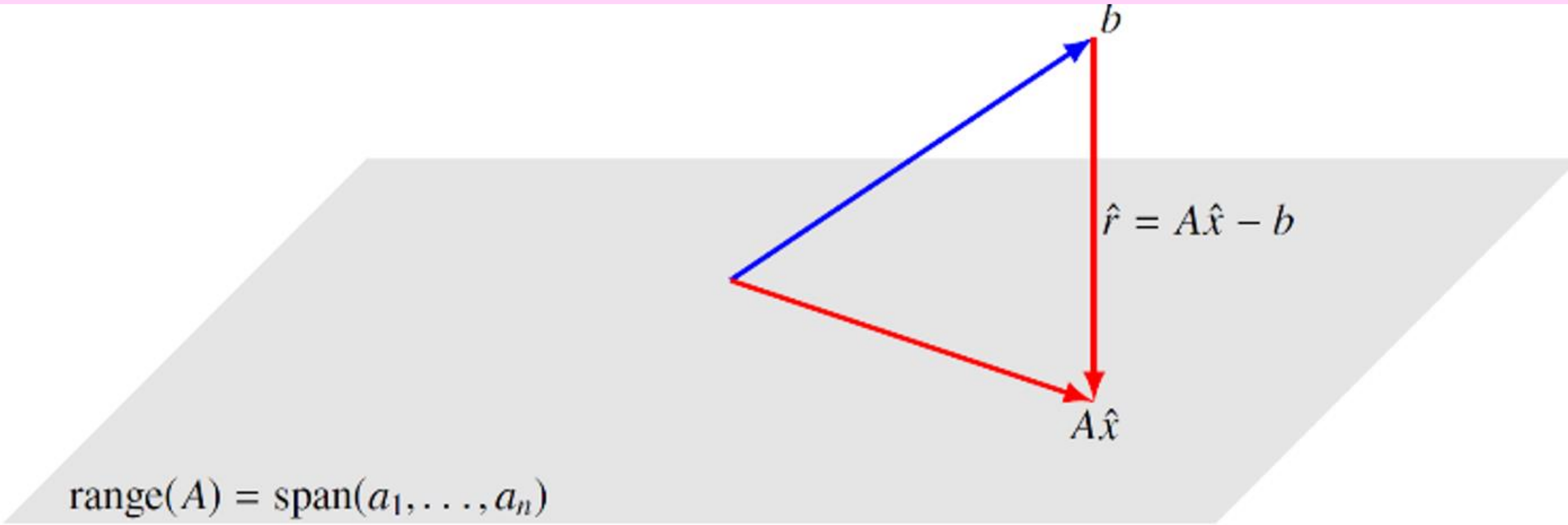
$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), \dots, \frac{\partial f}{\partial x_n}(x)\right) = 2A^T(Ax - b)$$

- minimizer $\hat{x}$ of $f(x)$ satisfies $\nabla f(\hat{x}) = 2A^T(A\hat{x} - b) = 0 \rightarrow \hat{x} = (A^T A)^{-1} A^T b$

Geometric interpretation

Important

- residual vector $\hat{r} = A\hat{x} - b$ satisfies $A^T\hat{r} = \mathrm{A}^{\mathrm{T}}(\mathrm{A}\hat{x} - b) = 0$



residual vector $\hat{r}$ is orthogonal to every column of A; hence, to range(A) projection on range(A) is a matrix-vector multiplication with the matrix

$$A(A^TA)^{-1}A^T = AA^\dagger$$

Conclusion

Important

Let A be an $m \times n$ matrix. The following statements are logically equivalent:

a. The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for each $\mathbf{b}$ in $\mathbb{R}^m$

b. The columns of A are linearly independent.

c. The matrix $A^T A$ is invertible.

When these statements are true, the least-squares solution $\hat{\mathbf{x}}$ is given by

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

When a least-squares solution $\hat{\mathbf{x}}$ is used to produce $A\hat{\mathbf{x}}$ as an approximation to $\mathbf{b}$, the distance from $\mathbf{b}$ to $A\hat{\mathbf{x}}$ is called the **least-squares error** of this approximation.

When $\hat{\mathbf{x}} = A^{-1}\mathbf{b}$?

Solving least squares problems (Method 1)

Example

- Normal equations of the least squares problem $A^T Ax = A^T b$
  - Coefficient matrix $A^T A$ is the ......
  - Equivalent to $\nabla f(x) = 0$ where $f(x) =$
  - All solutions of the least squares problem satisfy the normal equations

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

Solving least squares problems (Method 2): QR factorization

Example

- Rewrite least squares solution using QR factorization $A = QR$

- Complexity: $2mn^2$

Algorithm: Least squares via QR factorization
Input: A : $m \times n$ left-invertible **Input:** b : $m \times 1$ **output:** x_{LS} : $n \times 1$ Find QR factorization $A = QR$ Compute $Q^T b$ Solve $Rx_{LS} = Q^T b$ using back substitution

- Identical to algorithm for solving $Ax = b$ for square invertible A, but when A is tall, gives least squares approximate solution

Solving least squares problems

Example

a 3×2 matrix with "almost linearly dependent" columns

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

round intermediate results to 8 significant decimal digits

- Solve using both methods
    - Which one is more stable? Why?

Note

- we choose the model $\hat{f}(x)$ from a family models

$$\hat{f}(x) = \theta_1 f_1(x) + \theta_2 f_2(x) + \cdots + \theta_p f_p(x)$$

model parameters

scalar valued basis functions (chosen by us)

Solution of weighted least squares

Example

weighted least squares is equivalent to a standard least squares problem

- $$\text{minimize} \quad \left\| \begin{bmatrix} \sqrt{\lambda_1} A_1 \\ \sqrt{\lambda_2} A_2 \\ \vdots \\ \sqrt{\lambda_k} A_k \end{bmatrix} x - \begin{bmatrix} \sqrt{\lambda_1} b_1 \\ \sqrt{\lambda_2} b_2 \\ \vdots \\ \sqrt{\lambda_k} b_k \end{bmatrix} \right\|^2$$

- Solution is unique if the *stacked matrix* has linearly independent columns
- Each matrix A_i may have linearly dependent columns (or be a wide matrix)
- if the stacked matrix has linearly independent columns, the solution is

$$\hat{x} = \left(\lambda_1 A_1^T A_1 + \cdots + \lambda_k A_k^T A_k\right)^{-1} (\lambda_1 A_1^T b_1 + \cdots + \lambda_k A_k^T b_k)$$

Lagrange multiplier

Example

$$f(x) = \min(x_1 x_2)$$

$$g(x) = 1 - \mathrm{x}_1 - \mathrm{x}_2$$

$$g(x) = 0$$

$$L(x, \lambda) = f(x) + \lambda g(x)$$

$$\nabla f(x)$$

Constrained Least Square

Example

- $\begin{cases} \min\limits_x ||Ax - b||^2 & A: m \times n \\ s.t. \quad Cx = d & C: p \times n \end{cases}$

$$L(x, \lambda) = ||Ax - b||^2 + \lambda^T (Cx - d)$$

$$\begin{cases} \nabla_x L = 2A^T Ax - 2A^T b + C^T \lambda = 0 \\ \nabla_\lambda L = Cx - d = 0 \end{cases} \rightarrow \begin{bmatrix} 2A^T A & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \lambda^* \end{bmatrix} = \begin{bmatrix} 2A^T b \\ d \end{bmatrix}$$

Note

- #equations: $n + p$ #Unkowns: $n + p$
- KKT equations
- Least Square problem is a KKT problem with $A = I, b = 0$

Class Activity

Class Activity

Does the least squared error method give more weight to points with larger residuals when calculating the sum of squared residuals?