Linear Control and Estimation

Least Squares Methods

Sivakumar Balasubramanian

Department of Bioengineering
Christian Medical College, Bagayam
Vellore 632002

References

- S Boyd, Introduction to Applied Linear Algebra: Chapters 12, 13, 15, 16 and 17.

Overdetermined System of linear equations

- For a tall, skinny matrix $A \in \mathbb{R}^{m \times n}$, there is a solution to $Ax = b$, only when $b \in C(A)$.

$$b = \sum_{i=1}^{n} v_i \alpha_i = V\alpha; \;\; \alpha \in \mathbb{R}^n, \;\; V = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix} \in \mathbb{R}^{n \times n}$$

- Can we have an approximate solution when $\nexists x$ such that $Ax = b$? Let us define "approximate" solution $\hat{x}$ as the one that minimizes $\|b - A\hat{x}\|_2^2, \;\; \forall x \in \mathbb{R}^n$. This is the *least squares problem*.

$$\text{Given } A \text{ and } b, \text{ choose } \hat{x} \text{ such that}$$

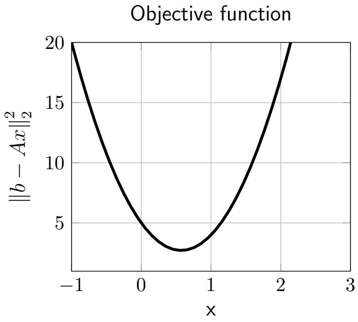
$$\text{minimize} \quad \|b - Ax\|_2^2$$

- A and b come from the data.
- $\|b - Ax\|^2$ is called the objective function.

Least Squares Problem

$$\left.\begin{aligned} 2x &= 1 \\ -1x &= -2 \\ \sqrt{2}x &= 0 \end{aligned}\right\} \longrightarrow Ax = b, \quad A = \begin{bmatrix} 2 \\ -1 \\ \sqrt{2} \end{bmatrix}, \; x \in \mathbb{R}, \; b \in \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}$$

$$\|b - Ax\|^2 = (1 - 2x)^2 + (-2 + x)^2 + \left(\sqrt{2}x\right)^2 = 7x^2 - 8x + 5 \geq 0$$



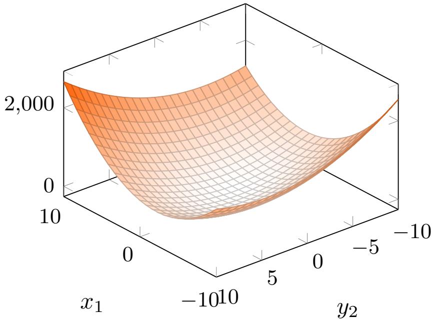
The objective function assumes its minimum value, at $\hat{x} = \frac{4}{7}$

Least Squares Problem

$$\left.\begin{aligned} 2x_1 - x_2 &= 2 \\ -x_1 + x_2 &= 1 \\ 3x_1 + 2x_2 &= -1 \end{aligned}\right\} \longrightarrow Ax = b, \quad A = \begin{bmatrix} 2 & -1 \\ -1 & 1 \\ 3 & 2 \end{bmatrix}, \; x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \; b \in \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$$

$$\|b - Ax\|^2 = 14x_1^2 + 6x_2^2 + 6x_2 + 6x_1x_2 + 6 \geq 0$$

$J = 14x_1^2 + 6x_2^2 + 6x_2 + 6x_1x_2 + 6$



The objective function assumes it minimum value at, $\hat{x}_1 = \frac{52}{75}$ and $\hat{x}_2 = \frac{3}{25}$.

Least Squares Methods

- The general solution to this least squares problem can be derived using calculus. Let $f(x) = \|b - Ax\|$

$$\nabla f(x) = 0 \longrightarrow \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = 0$$

Going through the algebra, we end up with the following expression for $\hat{x}$ that minimizes $f(x)$,

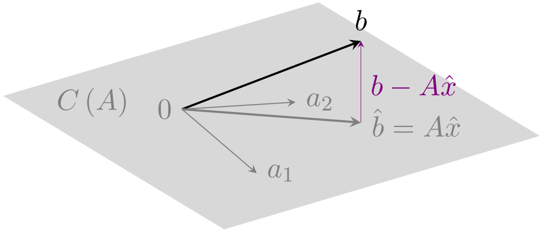
$$\underbrace{A^T A \hat{x} = A^T b}_{\text{Normal Equations}}$$

A is full rank, $\implies A^T A$ is invertible.

$$\implies \hat{x} = \underbrace{\left(A^T A\right)^{-1} A^T}_{\text{Pseudo-inverse}} b = A^{\dagger} b$$

Least Squares Methods

- $\hat{x}$ is the approximate least squares solution. $\hat{b} = A\hat{x}$, which is in general not equal to b. *When is $b = \hat{b}$?*
- We know two things about $\hat{b}$,
  1. $\hat{b} \in C(A)$: $\hat{b}$ is the column space of A.
  2. $\left\|b - \hat{b}\right\|$ is minimum.



$$\|b - A\hat{x}\|_2^2 \text{ is minimum } \implies (b - A\hat{x}) \perp A\hat{x}$$

$$(A\hat{x})^T (b - A\hat{x}) = 0 \implies \hat{x}^T \underbrace{\left(A^T b - A^T A\hat{x}\right)}_{\text{Normal Equations}} = 0$$

The least squares approximate solution of $Ax = b$ is the solution solution to the equation $Ax = \hat{b}$, where $\hat{b}$ is the projection of b onto the column space of A $(C(A))$

Multi-Objective Least Squares

- There are applications where there is more than one objective that must be optimized,

$$J_1 = \|A_1 x - b_1\|^2, \; J_2 = \|A_2 x - b_2\|^2, \; \ldots \; J_k = \|A_k x - b_k\|^2,$$

  and often these are conflicting objectives.

- We can define a single objective function J that is takes into account the different objective functions.

$$J = \sum_{i=1}^{k} \rho_i J_i, \;\; \rho_i > 0, \;\; \forall 1 \leq i \leq k$$

- The ρ_is indicate the relative weightage given to the individual objectives.

$$J = J_1 + \sum_{i=2}^{k} \rho_i J_i$$

Multi-Objective Least Squares

$$J = \rho_1 \left\|A_1 x - b_1\right\|^2 + \ldots + \rho_k \left\|A_k x - b_k\right\|^2 = \left\|\sqrt{\rho_1} A_1 x - \sqrt{\rho_1} b_1\right\|^2 + \ldots + \rho_k \left\|\sqrt{\rho_k} A_k x - \sqrt{\rho_k} b_k\right\|^2$$

$$J = \left\| \begin{bmatrix} \sqrt{\rho_1} A_1 \\ \sqrt{\rho_2} A_1 \\ \vdots \\ \sqrt{\rho_k} A_k \end{bmatrix} x - \begin{bmatrix} \sqrt{\rho_1} b_1 \\ \sqrt{\rho_2} b_1 \\ \vdots \\ \sqrt{\rho_k} b_k \end{bmatrix} \right\|^2 = \left\|\tilde{A} x - \tilde{b}\right\|^2 \implies \hat{x} = \left(\tilde{A}^T \tilde{A}\right)^{-1} \tilde{A}^T \tilde{b}$$

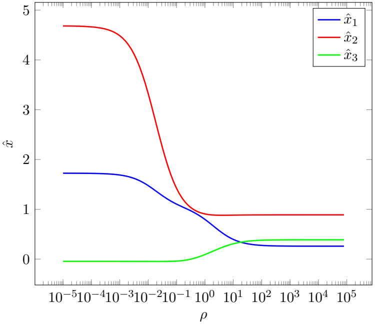
The columns of $\tilde{A}$ are must be independent, which happens if the columns of at least one of the A_is is independent.

Consider a two objective case, $J = J_1 + \rho J_2$.

$$\hat{x} = \begin{cases} \text{argmin}_x \left\|A_1 x - b_1\right\|^2 & \rho = 0 \\ \text{argmin}_x \left\|A_2 x - b_2\right\|^2 & \rho \to \infty \end{cases}$$

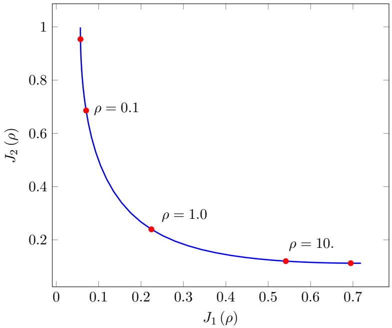
Multi-Objective Least Squares

Least squares approximate solution $\hat{x}(\rho)$



Any solution that lies on this curve is called the *Pareto optimal* solution. There exists no solution $\tilde{x}$, such that $J_1(\tilde{x}) \leq J_1(\hat{x})$ and $J_2(\tilde{x}) \leq J_2(\hat{x})$ where, both inqualities hold strictly.

Optimal trade-off curve



Multi-Objective Least Squares

- Multi-objective least squares methods play an important role in both control and estimation problems.
- Appropriate choice of the objective functions can also help deal with conditions where the columns of A_i are not independent. Consider the following example,

$$J_1 = \|A_1 x - b_1\|^2 \quad \text{and} \quad J_2 = \|A_2 x - b - 2\|^2$$

  where, $A_1 \in \mathbb{R}^{m_1 \times n}$ and $A_2 \in \mathbb{R}^{m_2 \times n}$, such that $m_1, m_2 < n$. Thus, the columns of A_1 and A_2 are not independent! However, if $m_1 + m_2 \geq n$, then it is possible that the columns of $\tilde{A}$ are independent.
- This is called **regularized least squares**.
- **Tichonov regularization**: $J = \|Ax - y\|^2 + \rho \|x\|^2$, where $\rho > 0$.

$$\tilde{A} = \begin{bmatrix} A \\ \sqrt{\rho} I \end{bmatrix} \implies \hat{x} = \left(A^T A + \rho I\right)^{-1} A^T b$$

Multi-Objective Least Squares

- **Tichonov regularization**: $J = \|Ax - y\|^2 + \rho \|x\|^2$, where $\rho > 0$.

$$\tilde{A} = \begin{bmatrix} A \\ \sqrt{\rho} I \end{bmatrix} \implies \hat{x} = \left(A^T A + \rho I\right)^{-1} A^T b$$

- $\hat{x}$ gives a unique solution in minimizing J, even when A is not full rank.
- Even when A is full rank, the regularization term can be used to improve the condition number of the matrix.

Multi-Objective Least Squares

- **Tichonov regularization**: $J = \|Ax - y\|^2 + \rho\|x\|^2$, where $\rho > 0$.

$$\tilde{A} = \begin{bmatrix} A \\ \sqrt{\rho} I \end{bmatrix} \implies \hat{x} = \left(A^T A + \rho I\right)^{-1} A^T b$$

- $\hat{x}$ gives a unique solution in minimizing J, even when A is not full rank.
- Even when A is full rank, the regularization term can be used to improve the condition number of the matrix.

Constrained Least Squares

- **Problem**:

$$\begin{aligned} \text{minimize} \quad & \|Ax - b\|^2 \\ \text{subject to} \quad & Cx = d \end{aligned}$$

  where, $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $C \in \mathbb{R}^{p \times n}$ and $d \in \mathbb{R}^p$.

- This can be solved using the *method of Lagrange multipliers*. When we do this, we finally arrive the following set of equations, called the *Karush-Kuhn-Tucker* (KKT) equation,

$$2\left(A^T A\right)\hat{x} - 2A^T b + C^T \hat{z} = 0$$

$$\begin{bmatrix} 2\left(A^T A\right) & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{z} \end{bmatrix} = \begin{bmatrix} 2A^T b \\ d \end{bmatrix}$$

- The coefficient matrix on the LHS of the KKT equation a square matrix of dimensions $(n+p) \times (n+p)$ is invertible, if and only if, $\begin{bmatrix} A \\ C \end{bmatrix}$ is full rank.