◆□ > ◆□ > ◆目 > ◆目 > ○目 · のへで

Linear Systems 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 $\mathbf{A} \in \mathbb{R}^{m \times n}$, there is a solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$, only when $\mathbf{b} \in C(\mathbf{A})$.

$$\mathbf{b} = \sum_{i=1}^{n} \mathbf{v}_{i} a_{i} = \mathbf{V} \mathbf{a}; \ \mathbf{a} \in \mathbb{R}^{n}, \ \mathbf{V} = \begin{bmatrix} \mathbf{v}_{1} & \mathbf{v}_{2} & \dots & \mathbf{v}_{n} \end{bmatrix} \in \mathbb{R}^{n \times n}$$

 Can we have an approximate solution when ∄x such that Ax = b? Let us define "approximate" solution x̂ as the one that minimizes ||b - Ax̂||₂², ∀x ∈ ℝⁿ. This is the *least squares problem*.

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

minimize $\|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2$

- ▶ A and b come from the data.
- $\|\mathbf{b} \mathbf{A}\mathbf{x}\|^2$ is called the objective function.

Least Squares Problem

$$2x = 1$$

$$-1x = -2$$

$$\sqrt{2}x = 0$$

$$\rightarrow \mathbf{Ax} = \mathbf{b}, \ \mathbf{A} = \begin{bmatrix} 2\\ -1\\ \sqrt{2} \end{bmatrix}, \ \mathbf{x} \in \mathbb{R}, \ \mathbf{b} \in \begin{bmatrix} 1\\ -2\\ 0 \end{bmatrix}$$

$$\|\mathbf{b} - \mathbf{A}\mathbf{x}\|^2 = (1 - 2x)^2 + (-2 + x)^2 + (\sqrt{2}x)^2 = 7x^2 - 8x + 5 \ge 0$$



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

Least Squares Problem

$$2x_{1} - x_{2} = 2 \\ -x_{1} + x_{2} = 1 \\ 3x_{1} + 2x_{2} = -1 \end{cases} \longrightarrow \mathbf{Ax} = \mathbf{b}, \ \mathbf{A} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \\ 3 & 2 \end{bmatrix}, \ \mathbf{x} = \begin{bmatrix} x_{1} \\ x_{2} \end{bmatrix}, \ \mathbf{b} \in \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$$
$$\|\mathbf{b} - \mathbf{Ax}\|^{2} = 14x_{1}^{2} + 6x_{2}^{2} + 6x_{2} + 6x_{1}x_{2} + 6 \ge 0$$
$$J = 14x_{1}^{2} + 6x_{2}^{2} + 6x_{2} + 6x_{1}x_{2} + 6 \ge 0$$
$$J = 14x_{1}^{2} + 6x_{2}^{2} + 6x_{2} + 6x_{1}x_{2} + 6$$
$$\mathsf{The objective function assumes it minimum value at, \ \hat{x}_{1} = \frac{52}{75} \text{ and} \\ \hat{x}_{2} = \frac{3}{25}.$$

unction assumes it at, $\hat{x}_1 = rac{52}{75}$ and

Least Squares Methods

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

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

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

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

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

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

Least Squares Methods

- $\hat{\mathbf{x}}$ is the approximate least squares solution. $\hat{\mathbf{b}} = \mathbf{A}\hat{\mathbf{x}}$, which is in general not equal to \mathbf{b} . When is $\mathbf{b} = \hat{\mathbf{b}}$?
- \blacktriangleright We know two things about $\hat{\mathbf{b}},$

1.
$$\hat{\mathbf{b}} \in C(\mathbf{A})$$
: $\hat{\mathbf{b}}$ is the column space of \mathbf{A} .
2. $\left\| \mathbf{b} - \hat{\mathbf{b}} \right\|$ is minimum.



$$\begin{aligned} |\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}||_{2}^{2} \text{ is minimum } \implies (\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}) \perp \mathbf{A}\hat{\mathbf{x}} \\ (\mathbf{A}\hat{\mathbf{x}})^{T} (\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}) &= 0 \implies \hat{\mathbf{x}}^{T} \underbrace{\left(\mathbf{A}^{T}\mathbf{b} - \mathbf{A}^{T}\mathbf{A}\hat{\mathbf{x}}\right)}_{\text{Normal Equations}} = 0 \end{aligned}$$

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

▲□▶▲圖▶▲≣▶▲≣▶ ≣ のへで

8/13

Multi-Objective Least Squares

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

$$J_1 = \|\mathbf{A}_1\mathbf{x} - \mathbf{b}_1\|^2, \ J_2 = \|\mathbf{A}_2\mathbf{x} - \mathbf{b}_2\|^2, \ \dots \ J_k = \|\mathbf{A}_k\mathbf{x} - \mathbf{b}_k\|^2,$$

and often these are conflicting objectives.

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

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

• The ρ_i s indicate the relative weightage given to the individual objectives.

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

Multi-Objective Least Squares

$$J = \rho_1 \|\mathbf{A}_1 \mathbf{x} - \mathbf{b}_1\|^2 + \dots + \rho_k \|\mathbf{A}_k \mathbf{x} - \mathbf{b}_k\|^2$$
$$= \|\sqrt{\rho_1} \mathbf{A}_1 \mathbf{x} - \sqrt{\rho_1} \mathbf{b}_1\|^2 + \dots + \|\sqrt{\rho_k} \mathbf{A}_k \mathbf{x} - \sqrt{\rho_k} \mathbf{b}_k\|^2$$
$$J = \left\| \begin{bmatrix} \sqrt{\rho_1} \mathbf{A}_1 \\ \sqrt{\rho_2} \mathbf{A}_2 \\ \vdots \\ \sqrt{\rho_k} \mathbf{A}_k \end{bmatrix} \mathbf{x} - \begin{bmatrix} \sqrt{\rho_1} \mathbf{b}_1 \\ \sqrt{\rho_2} \mathbf{b}_1 \\ \vdots \\ \sqrt{\rho_k} \mathbf{b}_k \end{bmatrix} \right\|^2 = \left\| \mathbf{\tilde{A}} \mathbf{x} - \mathbf{\tilde{b}} \right\|^2 \implies \hat{x} = \left(\mathbf{\tilde{A}}^T \mathbf{\tilde{A}} \right)^{-1} \mathbf{\tilde{A}}^T \mathbf{\tilde{b}}$$

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

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

$$\hat{\mathbf{x}} = \begin{cases} \operatorname{argmin}_{x} \|\mathbf{A}_{1}\mathbf{x} - \mathbf{b}_{1}\|^{2} & \rho = 0\\ \operatorname{argmin}_{x} \|\mathbf{A}_{2}\mathbf{x} - \mathbf{b}_{2}\|^{2} & \rho \to \infty \end{cases}$$

・ロト < 団ト < 三ト < 三ト 三 のへぐ

Multi-Objective Least Squares



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



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,

$$J_1 = \|\mathbf{A}_1 \mathbf{x} - \mathbf{b}_1\|^2$$
 and $J_2 = \|\mathbf{A}_2 \mathbf{x} - \mathbf{b}_2\|^2$

where, $\mathbf{A}_1 \in \mathbb{R}^{m_1 \times n}$ and $\mathbf{A}_2 \in \mathbb{R}^{m_2 \times n}$, such that $m_1, m_2 < n$. Thus, the columns of A_1 and A_2 are not independent! However, if $m_1 + m_2 \ge n$, then it is possible that the columns of \tilde{A} are independent.

This is called regularized least squares.

• Tichonov regularization: $J = \|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 + \rho \|\mathbf{x}\|^2$, where $\rho > 0$.

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

(ロ)、(型)、(E)、(E)、 E) のQ(()

12/13

Multi-Objective Least Squares

• Tichonov regularization:
$$J = \|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 + \rho \|\mathbf{x}\|^2$$
, where $\rho > 0$.

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

 $\mathbf{\hat{x}}$ gives a unique solution in minimizing J, even when \mathbf{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:

minimize $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$ subject to $\mathbf{C}\mathbf{x} = \mathbf{d}$

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

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,

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

▶ 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, $\begin{bmatrix} \mathbf{A} \\ \mathbf{C} \end{bmatrix}$ is full rank.