

Applied Linear Algebra for Data

Linear Least Squares

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 $\mathbf{Ax} = \mathbf{b}$, only when $\mathbf{b} \in C(\mathbf{A})$.

$$\mathbf{b} = \sum_{i=1}^n \mathbf{v}_i a_i = \mathbf{Va}; \quad \mathbf{a} \in \mathbb{R}^n, \quad \mathbf{V} = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \quad \mathbf{v}_n] \in \mathbb{R}^{n \times n}$$

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

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

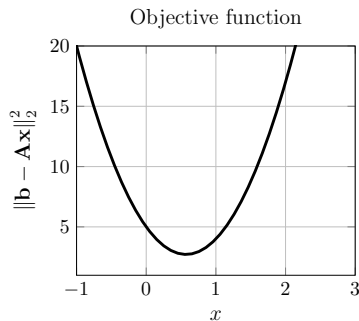
$$\text{minimize} \quad \|\mathbf{b} - \mathbf{Ax}\|_2^2$$

- ▶ \mathbf{A} and \mathbf{b} come from the data.
- ▶ $\|\mathbf{b} - \mathbf{Ax}\|_2^2$ is called the objective function.

Least Squares Problem

$$\left. \begin{array}{rcl} 2x & = & 1 \\ -1x & = & -2 \\ \sqrt{2}x & = & 0 \end{array} \right\} \rightarrow \mathbf{Ax} = \mathbf{b}, \quad \mathbf{A} = \begin{bmatrix} 2 \\ -1 \\ \sqrt{2} \end{bmatrix}, \quad \mathbf{x} \in \mathbb{R}, \quad \mathbf{b} \in \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}$$

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



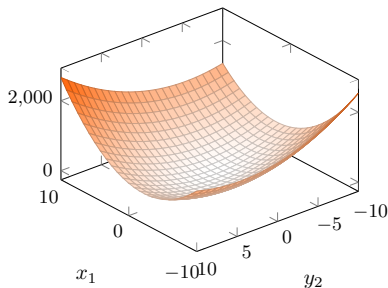
The objective function assumes its minimum value, at $\hat{\mathbf{x}} = \frac{4}{7}$

Least Squares Problem

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

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

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



The objective function assumes its 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 $f(\mathbf{x}) = \|\mathbf{b} - \mathbf{A}\mathbf{x}\|$

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

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

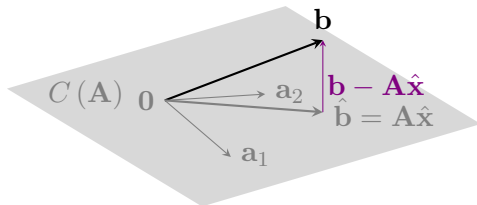
$$\underbrace{\mathbf{A}^T \mathbf{A} \hat{\mathbf{x}} = \mathbf{A}^T \mathbf{b}}_{\text{Normal Equations}}$$

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

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

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 $\mathbf{b} = \hat{\mathbf{b}}$?
- ▶ 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. $\|\mathbf{b} - \hat{\mathbf{b}}\|$ is minimum.



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

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

The least squares approximate solution of $\mathbf{A}\mathbf{x} = \mathbf{b}$ is the 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} ($C(\mathbf{A})$).

Multi-Objective Least Squares

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

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

and often these are conflicting objectives.

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

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

- ▶ 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

$$J = \rho_1 \|\mathbf{A}_1 \mathbf{x} - \mathbf{b}_1\|^2 + \dots + \rho_k \|\mathbf{A}_k \mathbf{x} - \mathbf{b}_k\|^2$$

$$= \|\sqrt{\rho_1} \mathbf{A}_1 \mathbf{x} - \sqrt{\rho_1} \mathbf{b}_1\|^2 + \dots + \|\sqrt{\rho_k} \mathbf{A}_k \mathbf{x} - \sqrt{\rho_k} \mathbf{b}_k\|^2$$

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

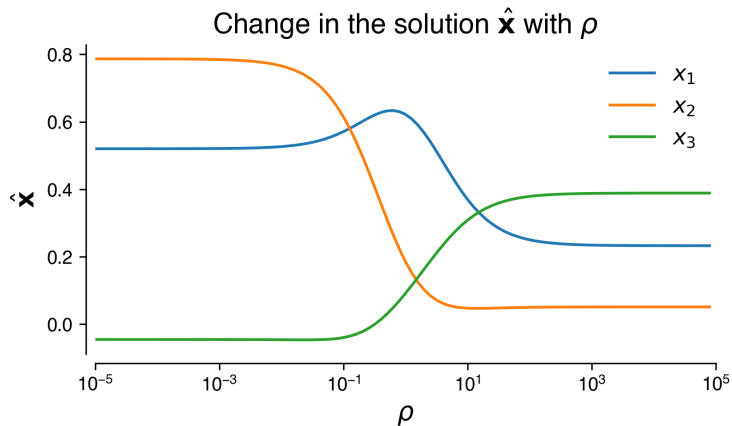
The columns of $\tilde{\mathbf{A}}$ are must be independent, which happens if the columns of at least one of the \mathbf{A}_i s is independent.

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

$$\hat{\mathbf{x}} = \begin{cases} \operatorname{argmin}_x \|\mathbf{A}_1 \mathbf{x} - \mathbf{b}_1\|^2 & \rho = 0 \\ \operatorname{argmin}_x \|\mathbf{A}_2 \mathbf{x} - \mathbf{b}_2\|^2 & \rho \rightarrow \infty \end{cases}$$

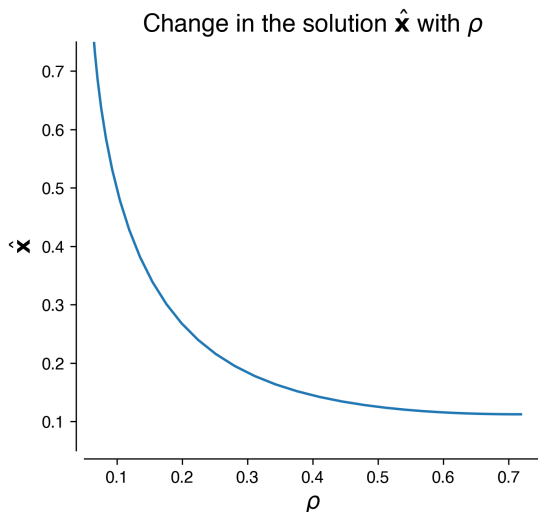
Multi-Objective Least Squares

Consider a problem with the objective function, $J(\mathbf{x}) = J_1(\mathbf{x}) + \rho J_2(\mathbf{x})$, where $\mathbf{x} \in \mathbb{R}^3$.



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 inequalities 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,

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

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 $m_1 + m_2 \geq n$, 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 $\rho > 0$.

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

Multi-Objective Least Squares

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

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

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

Constrained Least Squares

► **Problem:**

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

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,

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

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

- 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.