Least Squares Problems

Motivation

In many real-world situations, the linear system

$$Ac = y$$

is **overdetermined** and has **no exact solution**.

Motivation

In many real-world situations, the linear system

$$Ac = y$$

is **overdetermined** and has **no exact solution**.

$$\boxed{\nexists\, c \text{ such that } Ac = y \quad \Longleftrightarrow \quad y \notin \operatorname{Im}(A)}$$

Motivation

In many real-world situations, the linear system

$$Ac = y$$

is **overdetermined** and has **no exact solution**.

$$\boxed{\nexists\, c \text{ such that } Ac = y \quad \Longleftrightarrow \quad y \notin \mathrm{Im}(A)}$$

Goal of Least Squares

Find an *approximate solution* that best fits the data.

Least Squares: Problem Statement

Given:

- $A \in \mathbb{R}^{m \times n}$ with $m > n$,
- $y \in \mathbb{R}^m$,

Least Squares: Problem Statement

Given:

- $A \in \mathbb{R}^{m \times n}$ with $m > n$,
- $y \in \mathbb{R}^m$,

We seek $c \in \mathbb{R}^n$ minimizing:

$$\boxed{\|y - Ac\|_2^2}$$

Least Squares: Problem Statement

Given:

- $A \in \mathbb{R}^{m \times n}$ with $m > n$,
- $y \in \mathbb{R}^m$,

We seek $c \in \mathbb{R}^n$ minimizing:

$$\boxed{\|y - Ac\|_2^2}$$

Define the **residual vector**

$$r = y - Ac$$

and solve:

$$\boxed{\min_c \|r\|_2^2}$$

Interpretation

- Ac lies in the column space $\mathrm{Im}(A)$,
- y generally does not,
- least squares finds the closest point in $\mathrm{Im}(A)$ to y.

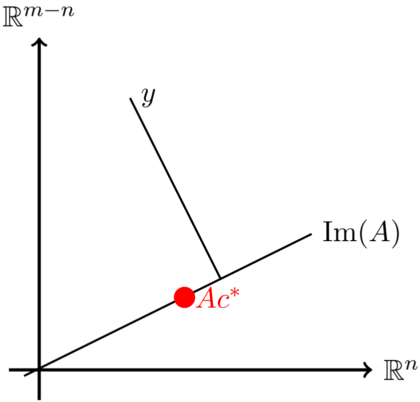
Interpretation

- Ac lies in the column space $\mathrm{Im}(A)$,
- y generally does not,
- least squares finds the closest point in $\mathrm{Im}(A)$ to y.

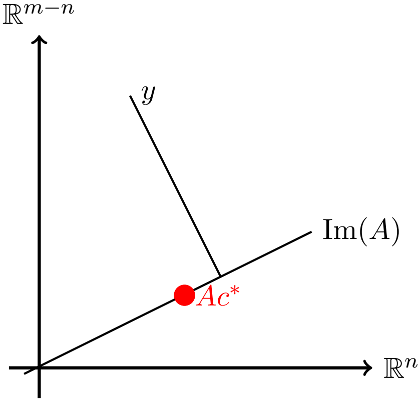
Key idea

Minimizing the residual norm is equivalent to minimizing the Euclidean distance between y and its approximation.

Geometric Intuition: Projection



Geometric Intuition: Projection



$$\boxed{r = y - Ac^* \perp \mathrm{Im}(A)}$$

Orthogonality Condition (Theorem)

c^* minimizes $\|y - Ac\|_2^2$ if and only if:

$$\boxed{y - Ac^* \;\perp\; \mathrm{Im}(A)}$$

Orthogonality Condition (Theorem)

c^* minimizes $\|y - Ac\|_2^2$ if and only if:

$$\boxed{y - Ac^* \;\perp\; \mathrm{Im}(A)}$$

Equivalently:

$$\forall z \in \mathbb{R}^n, \quad (Az)^\top r = 0$$

Orthogonality Condition (Theorem)

c^* minimizes $\|y - Ac\|_2^2$ if and only if:

$$\boxed{y - Ac^* \perp \mathrm{Im}(A)}$$

Equivalently:

$$\forall z \in \mathbb{R}^n, \quad (Az)^\top r = 0$$

This implies:

$$\boxed{A^\top r = 0}$$

Normal Equations

Since $r = y - Ac^*$:

$$A^\top (y - Ac^*) = 0$$

Normal Equations

Since $r = y - Ac^*$:

$$A^\top(y - Ac^*) = 0$$

$$\boxed{A^\top Ac^* = A^\top y}$$

Normal Equations

Since $r = y - Ac^*$:

$$A^\top (y - Ac^*) = 0$$

$$\boxed{A^\top A c^* = A^\top y}$$

Normal equations

They characterize the least-squares solution.

Derivation via Objective Expansion

$$\|Ac - y\|_2^2 = (Ac - y)^\top (Ac - y)$$

Derivation via Objective Expansion

$$\|Ac - y\|_2^2 = (Ac - y)^\top (Ac - y)$$

$$= c^\top A^\top Ac - c^\top A^\top y - y^\top Ac + y^\top y$$

Derivation via Objective Expansion

$$\|Ac - y\|_2^2 = (Ac - y)^\top (Ac - y)$$

$$= c^\top A^\top Ac - c^\top A^\top y - y^\top Ac + y^\top y$$

Taking the gradient:

$$\nabla_c = 2A^\top Ac - 2A^\top y$$

Derivation via Objective Expansion

$$\|Ac - y\|_2^2 = (Ac - y)^\top (Ac - y)$$

$$= c^\top A^\top Ac - c^\top A^\top y - y^\top Ac + y^\top y$$

Taking the gradient:

$$\nabla_c = 2A^\top Ac - 2A^\top y$$

Setting it to zero yields:

$$\boxed{A^\top Ac^* = A^\top y}$$

Explicit Solution and Pseudo-Inverse

If $\mathrm{rank}(A) = n$, then $A^\top A$ is invertible and:

$$\boxed{c^* = (A^\top A)^{-1} A^\top y = A^+ y}$$

Explicit Solution and Pseudo-Inverse

If $\mathrm{rank}(A) = n$, then $A^\top A$ is invertible and:

$$\boxed{c^* = (A^\top A)^{-1} A^\top y = A^+ y}$$

Moore–Penrose pseudo-inverse

Maps $y \in \mathbb{R}^m$ to the least-squares optimal $c^* \in \mathbb{R}^n$.

Numerical Considerations

Directly computing:

$$(A^\top A)^{-1} A^\top$$

is **numerically dangerous**.

Numerical Considerations

Directly computing:

$$(A^\top A)^{-1} A^\top$$

is **numerically dangerous**.

- squares the condition number,
- unstable for ill-conditioned or nearly rank-deficient A.

Numerical Considerations

Directly computing:

$$(A^\top A)^{-1} A^\top$$

is **numerically dangerous**.

- squares the condition number,
- unstable for ill-conditioned or nearly rank-deficient A.

In practice

Use QR or SVD decompositions.

Least Squares via QR Decomposition

If $A = QR$ with:

- $Q^\top Q = I$,
- R upper triangular,

Least Squares via QR Decomposition

If $A = QR$ with:

- $Q^\top Q = I$,
- R upper triangular,

Solve:

$$Rc = Q^\top y$$

Least Squares via QR Decomposition

If $A = QR$ with:

- $Q^\top Q = I$,
- R upper triangular,

Solve:

$$Rc = Q^\top y$$

$$\boxed{A^+ = R^{-1}Q^\top}$$

Least Squares via QR Decomposition

If $A = QR$ with:

- $Q^\top Q = I$,
- R upper triangular,

Solve:

$$Rc = Q^\top y$$

$$\boxed{A^+ = R^{-1}Q^\top}$$

Advantages

Stable and efficient.

Least Squares via SVD

For general matrices:

$$A = U\Sigma V^\top$$

Least Squares via SVD

For general matrices:

$$A = U\Sigma V^\top$$

$$\boxed{A^+ = V\Sigma^{-1}U^\top}$$

Least Squares via SVD

For general matrices:

$$A = U\Sigma V^\top$$

$$\boxed{A^+ = V\Sigma^{-1}U^\top}$$

- handles rank deficiency,
- reveals effective dimensionality,
- yields minimum-norm solution.

Summary

- Least squares handles overdetermined systems
- Solution $=$ orthogonal projection onto $\mathrm{Im}(A)$
- Leads to normal equations
- Computed stably using QR or SVD

Summary

- Least squares handles overdetermined systems
- Solution $=$ orthogonal projection onto $\mathrm{Im}(A)$
- Leads to normal equations
- Computed stably using QR or SVD

Key message

Least squares is both a geometric projection and a numerical optimization problem.