9. Equality constraints and tradeoffs

- More least squares
- Example: moving average model
- Minimum-norm least squares
- Equality-constrained least squares
- Optimal tradeoffs
- Example: hovercraft

Laurent Lessard (www.laurentlessard.com)

More least squares

Solving the least squares optimization problem:

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

Is equivalent to solving the normal equations:

$$A^\mathsf{T} A \hat{x} = A^\mathsf{T} b$$

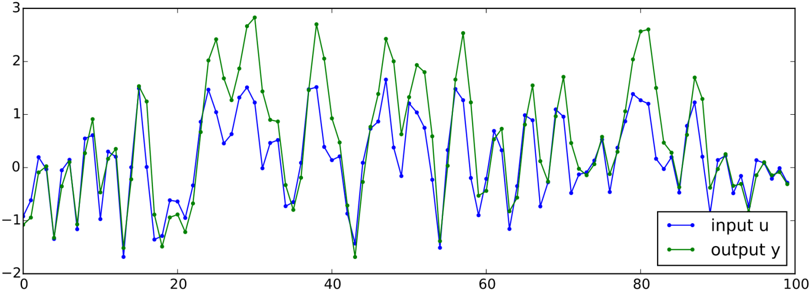
- If $A^\mathsf{T} A$ is invertible (A has linearly independent columns)

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

- $A^\dagger := (A^\mathsf{T} A)^{-1} A^\mathsf{T}$ is called the **pseudoinverse** of A.

Example: moving average model

- We are given a time series of input data $u_1, u_2, \ldots, u_T$ and output data $y_1, y_2, \ldots, y_T$. Example:



- A "moving average" model with window size k assumes each output is a weighted combination of k previous inputs:

$$y_t \approx w_1 u_t + w_2 u_{t-1} + \cdots + w_k u_{t-k+1} \qquad \text{for all } t$$

- find weights $w_1, \ldots, w_k$ that best agree with the data.

Example: moving average model

- Moving average model:

$$y_t \approx w_1 u_t + w_2 u_{t-1} + w_3 u_{t-2} \qquad \text{for all } t$$

- Writing all the equations (e.g. $k = 3$):

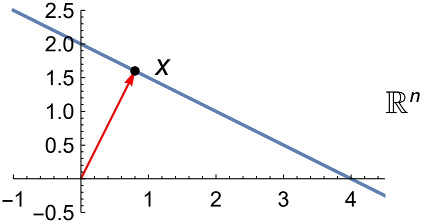
$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_T \end{bmatrix} \approx \begin{bmatrix} u_1 & 0 & 0 \\ u_2 & u_1 & 0 \\ u_3 & u_2 & u_1 \\ \vdots & \vdots & \vdots \\ u_T & u_{T-1} & u_{T-2} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$$

- Solve least squares problem! Moving Average.ipynb

Minimum-norm least squares

Underdetermined case: $A \in \mathbb{R}^{m \times n}$ is a wide matrix ($m \leq n$), so $Ax = b$ generally has infinitely many solutions.

- The set of solutions of $Ax = b$ forms an affine subspace. Recall: if $Ay = b$ and $Az = b$ then $A(\alpha y + (1 - \alpha)z) = b$.
- One possible choice: pick the x with smallest norm.



- **Insight:** The optimal $\hat{x}$ must satisfy $A\hat{x} = b$ and $\hat{x}^\mathsf{T}(\hat{x} - w) = 0$ for all w satisfying $Aw = b$.

Minimum-norm least squares

- We want: $\hat{x}^\mathsf{T}(\hat{x} - w) = 0$ for all w such that $Aw = b$.
- We also know that $A\hat{x} = b$. Therefore: $A(\hat{x} - w) = 0$.

In other words:

$$\hat{x} \perp (\hat{x} - w) \qquad \text{and} \qquad (\hat{x} - w) \perp (\text{all rows of } A)$$

Therefore, $\hat{x}$ is a linear combination of the rows of A.
Stated another way, $\hat{x} = A^\mathsf{T} z$ for some z.

- Therefore, we must find z and $\hat{x}$ such that:

$$A\hat{x} = b \quad \text{and} \quad A^\mathsf{T} z = \hat{x}$$

(this also follows from $\mathcal{R}(A)^\perp = \mathcal{N}(A^\mathsf{T})$)

Minimum-norm least squares

Theorem: If there exists $\hat{x}$ and z that satisfy $A\hat{x} = b$ and $A^\mathsf{T} z = \hat{x}$, then $\hat{x}$ is a solution to the minimum-norm problem

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \|x\|^2 \\ \text{subject to:} \quad & Ax = b \end{aligned}$$

Proof: Suppose $A\hat{x} = b$ and $A^\mathsf{T} z = \hat{x}$. For any x that satisfies $Ax = b$, we have:

$$\begin{aligned} \|x\|^2 &= \|x - \hat{x} + \hat{x}\|^2 \\ &= \|x - \hat{x}\|^2 + \|\hat{x}\|^2 + 2\hat{x}^\mathsf{T}(x - \hat{x}) \\ &= \|x - \hat{x}\|^2 + \|\hat{x}\|^2 + 2z^\mathsf{T} A(x - \hat{x}) \\ &= \|x - \hat{x}\|^2 + \|\hat{x}\|^2 \\ &\geq \|\hat{x}\|^2 \end{aligned}$$

Minimum-norm least squares

Solving the minimum-norm least squares problem:

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \|x\|^2 \\ \text{subject to:} \quad & Ax = b \end{aligned}$$

Is equivalent to solving the linear equations:

$$A\hat{x} = b \quad \text{and} \quad A^\mathsf{T} z = \hat{x} \qquad \Longrightarrow \qquad AA^\mathsf{T} z = b$$

- If AA^T is invertible (A has linearly independent rows)

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

- $A^\dagger := A^\mathsf{T}(AA^\mathsf{T})^{-1}$ is **also** called the **pseudoinverse** of A.

Equality-constrained least squares

A more general optimization problem:

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

(Equality-constrained least squares)

- If $C = 0$, $d = 0$, we recover ordinary least squares
- If $A = I$, $b = 0$, we recover minimum-norm least squares

Equality-constrained least squares

Solving the equality-constrained least squares problem:

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

Is equivalent to solving the linear equations:

$$A^\mathsf{T} A\hat{x} + C^\mathsf{T} z = A^\mathsf{T} b \quad \text{and} \quad C\hat{x} = d$$

Equality-constrained least squares

Proof: Suppose $\hat{x}$ and z satisfy $A^T A\hat{x} + C^T z = A^T b$ and $C\hat{x} = d$. Let x be any other point satisfying $Cx = d$. Then,

$$
\begin{aligned}
\|Ax - b\|^2 &= \|A(x - \hat{x}) + (A\hat{x} - b)\|^2 \\
&= \|A(x - \hat{x})\|^2 + \|A\hat{x} - b\|^2 + 2(x - \hat{x})^T A^T (A\hat{x} - b) \\
&= \|A(x - \hat{x})\|^2 + \|A\hat{x} - b\|^2 - 2(x - \hat{x})^T C^T z \\
&= \|A(x - \hat{x})\|^2 + \|A\hat{x} - b\|^2 - 2(Cx - C\hat{x})^T z \\
&= \|A(x - \hat{x})\|^2 + \|A\hat{x} - b\|^2 \\
&\geq \|A\hat{x} - b\|^2
\end{aligned}
$$

Therefore $\hat{x}$ is an optimal choice.

Recap so far

Several different variants of least squares problems are **easy** to solve in the sense that they are equivalent to solving systems of linear equations.

Least squares

$$\min_x \; \|Ax - b\|^2$$

Minimum-norm

$$\begin{aligned} \min_x \quad & \|x\|^2 \\ \text{s.t.} \quad & Ax = b \end{aligned}$$

Equality constrained

$$\begin{aligned} \min_x \quad & \|Ax - b\|^2 \\ \text{s.t.} \quad & Cx = d \end{aligned}$$

Optimal tradeoffs

We often want to optimize several different objectives simultaneously, but these objectives are **conflicting**.

- risk vs expected return (finance)
- power vs fuel economy (automobiles)
- quality vs memory (audio compression)
- space vs time (computer programs)
- mittens vs gloves (winter)

Optimal tradeoffs

- Suppose $J_1 = \|Ax - b\|^2$ and $J_2 = \|Cx - d\|^2$.
- We would like to make **both** J_1 and J_2 small.
- A sensible approach: solve the optimization problem:

$$\underset{x}{\text{minimize}} \quad J_1 + \lambda J_2$$

  where $\lambda > 0$ is a (fixed) **tradeoff parameter**.
- Then tune λ to explore possible results.
  - When $\lambda \to 0$, we place more weight on J_1
  - When $\lambda \to \infty$, we place more weight on J_2

Optimal tradeoffs

This problem is also equivalent to solving linear equations!

$$\begin{aligned} J_1 + \lambda J_2 &= \|Ax - b\|^2 + \lambda \|Cx - d\|^2 \\ &= \left\| \begin{bmatrix} Ax - b \\ \sqrt{\lambda}(Cx - d) \end{bmatrix} \right\|^2 \\ &= \left\| \begin{bmatrix} A \\ \sqrt{\lambda} C \end{bmatrix} x - \begin{bmatrix} b \\ \sqrt{\lambda} d \end{bmatrix} \right\|^2 \end{aligned}$$

- An ordinary least squares problem!
- Equivalent to solving

$$(A^\mathsf{T} A + \lambda C^\mathsf{T} C)\,\hat{x} = (A^\mathsf{T} b + \lambda C^\mathsf{T} d)$$

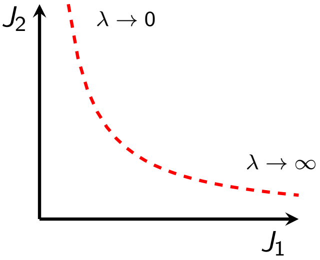
Tradeoff analysis

1. Choose values for λ (usually log-spaced). A useful command: `lambda = logspace(p,q,n)` produces n points logarithmically spaced between 10^p and 10^q.

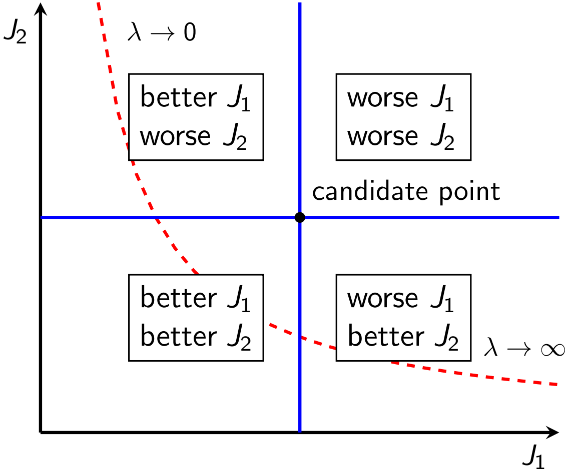
2. For each λ value, find $\hat{x}_\lambda$ that minimizes $J_1 + \lambda J_2$.

3. For each $\hat{x}_\lambda$, also compute the corresponding J_1^λ and J_2^λ.

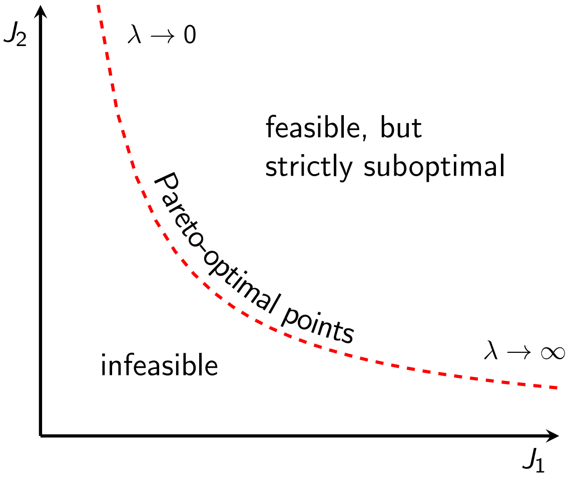
4. Plot $(J_1^\lambda, J_2^\lambda)$ for each λ and connect the dots.



Pareto curve

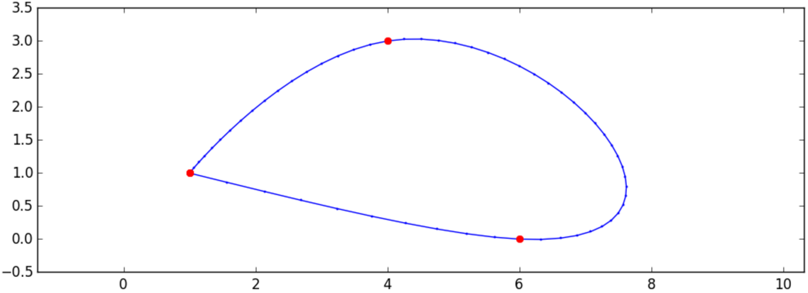


Pareto curve



Example: hovercraft

We are in command of a hovercraft. We are given a set of k waypoint locations and times. The objective is to hit the waypoints at the prescribed times while minimizing fuel use.



Goal is to choose appropriate thruster inputs at each instant.

Example: hovercraft

We are in command of a hovercraft. We are given a set of k waypoint locations and times. The objective is to hit the waypoints at the prescribed times while minimizing fuel use.

- Discretize time: $t = 0, 1, 2, \dots, T$.
- Important variables: position x_t, velocity v_t, thrust u_t.
- Simplified model of the dynamics:

$$\begin{aligned} x_{t+1} &= x_t + v_t \\ v_{t+1} &= v_t + u_t \end{aligned} \qquad \text{for } t = 0, 1, \dots, T-1$$

- We must choose $u_0, u_1, \dots, u_T$.
- Initial position and velocity: $x_0 = 0$ and $v_0 = 0$.
- Waypoint constraints: $x_{t_i} = w_i$ for $i = 1, \dots, k$.
- Minimize fuel use: $\|u_0\|^2 + \|u_1\|^2 + \cdots + \|u_T\|^2$

Example: hovercraft

First model: hit the waypoints exactly

$$
\begin{aligned}
\underset{x_t, v_t, u_t}{\text{minimize}} \quad & \sum_{t=0}^{T} \|u_t\|^2 \\
\text{subject to:} \quad & x_{t+1} = x_t + v_t \quad \text{for } t = 0, 1, \dots, T-1 \\
& v_{t+1} = v_t + u_t \quad \text{for } t = 0, 1, \dots, T-1 \\
& x_0 = v_0 = 0 \\
& x_{t_i} = w_i \qquad\qquad \text{for } i = 1, \dots, k
\end{aligned}
$$

Julia model: Hovercraft.ipynb

Example: hovercraft

Second model: allow waypoint misses

$$\begin{aligned} \underset{x_t, v_t, u_t}{\text{minimize}} \quad & \sum_{t=0}^{T} \|u_t\|^2 + \lambda \sum_{i=1}^{k} \|x_{t_i} - w_i\|^2 \\ \text{subject to:} \quad & x_{t+1} = x_t + v_t \quad \text{for } t = 0, 1, \ldots, T-1 \\ & v_{t+1} = v_t + u_t \quad \text{for } t = 0, 1, \ldots, T-1 \\ & x_0 = v_0 = 0 \end{aligned}$$

- λ controls the tradeoff between making u small and hitting all the waypoints.