9. Equality constraints and tradeoffs

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

Solving the least squares optimization problem:

\[
\minimize_x \| Ax - b \|^2
\]

Is equivalent to solving the normal equations:

\[
A^T A \hat{x} = A^T b
\]

- If \( A^T A \) is invertible (\( A \) has linearly independent columns)
  \[
  \hat{x} = (A^T A)^{-1} A^T b
  \]
- \( A^\dagger := (A^T A)^{-1} A^T \) 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:

\[
y_t \approx w_1 u_t + w_2 u_{t-1} + \cdots + w_k u_{t-k+1}
\]

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

\[
y_t \approx w_1 u_t + w_2 u_{t-1} + \cdots + w_k u_{t-k+1}
\]

- find weights $w_1, \ldots, w_k$ that best agree with the data.
Example: moving average model

- Moving average model:

\[ y_t \approx w_1 u_t + w_2 u_{t-1} + w_3 u_{t-2} \quad \text{for all } t \]

- 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 $\hat{x}^T(\hat{x} - w) = 0$ for all $w$ satisfying $Aw = b$. 
Minimum-norm least squares

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

In other words:

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

Therefore, $\hat{x}$ is a linear combination of the rows of $A$.

Stated another way, $\hat{x} = A^T z$ for some $z$.

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

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

(this also follows from $\mathcal{R}(A)^\perp = \mathcal{N}(A^T)$)
**Minimum-norm least squares**

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

\[
\text{minimize} \quad \|x\|^2 \\
\text{subject to:} \quad Ax = b
\]

**Proof:** Suppose \( A\hat{x} = b \) and \( A^Tz = \hat{x} \). For any \( x \) that satisfies \( Ax = b \), we have:

\[
\|x\|^2 = \|x - \hat{x} + \hat{x}\|^2 \\
= \|x - \hat{x}\|^2 + \|\hat{x}\|^2 + 2\hat{x}^T(x - \hat{x}) \\
= \|x - \hat{x}\|^2 + \|\hat{x}\|^2 + 2z^T A(x - \hat{x}) \\
= \|x - \hat{x}\|^2 + \|\hat{x}\|^2 \\
\geq \|\hat{x}\|^2
\]
Minimum-norm least squares

Solving the minimum-norm least squares problem:

\[
\begin{align*}
\text{minimize} & \quad \|x\|^2 \\
\text{subject to:} & \quad Ax = b
\end{align*}
\]

Is equivalent to solving the linear equations:

\[
A\hat{x} = b \quad \text{and} \quad A^Tz = \hat{x} \quad \implies \quad AA^Tz = b
\]

- If \(AA^T\) is invertible (\(A\) has linearly independent rows)
  \[
  \hat{x} = A^T(AA^T)^{-1}b
  \]
- \(A^\dagger := A^T(AA^T)^{-1}\) is also called the \textbf{pseudoinverse} of \(A\).
Equality-constrained least squares

A more general optimization problem:

\[
\begin{align*}
\text{minimize} & \quad \| Ax - b \|^2 \\
\text{subject to:} & \quad Cx = d \\
\end{align*}
\]

(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:

\[
\begin{align*}
\text{minimize} & \quad \|Ax - b\|^2 \\
\text{subject to:} & \quad Cx = d
\end{align*}
\]

Is equivalent to solving the linear equations:

\[
A^TA\hat{x} + C^Tz = A^Tb \quad \text{and} \quad C\hat{x} = d
\]
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,

\[
\|Ax - b\|^2 = \|A(x - \hat{x}) + (A\hat{x} - b)\|^2 \\
= \|A(x - \hat{x})\|^2 + \|A\hat{x} - b\|^2 + 2(x - \hat{x})^T A^T (A\hat{x} - b) \\
= \|A(x - \hat{x})\|^2 + \|A\hat{x} - b\|^2 - 2(x - \hat{x})^T C^T z \\
= \|A(x - \hat{x})\|^2 + \|A\hat{x} - b\|^2 - 2(Cx - C\hat{x})^T z \\
= \|A(x - \hat{x})\|^2 + \|A\hat{x} - b\|^2 \\
\geq \|A\hat{x} - b\|^2
\]

Therefore \( \hat{x} \) is an optimal choice.
Several different variants of least squares problems are \textbf{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**: \[
\min_x \| x \|^2 \\
\text{s.t. } Ax = b
\]
- **Equality constrained**: \[
\min_x \| Ax - b \|^2 \\
\text{s.t. } Cx = d
\]
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:

\[
\text{minimize} \quad J_1 + \lambda J_2
\]

where $\lambda > 0$ is a (fixed) tradeoff parameter.

- Then tune $\lambda$ 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!

\[ J_1 + \lambda J_2 = \|Ax - b\|^2 + \lambda \|Cx - d\|^2 \]

\[ = \left\| \begin{bmatrix} Ax - b \\ \sqrt{\lambda}(Cx - d) \end{bmatrix} \right\|^2 \]

\[ = \left\| \begin{bmatrix} A \\ \sqrt{\lambda}C \end{bmatrix} x - \begin{bmatrix} b \\ \sqrt{\lambda}d \end{bmatrix} \right\|^2 \]

- An ordinary least squares problem!
- Equivalent to solving

\[ (A^TA + \lambda C^TC)\hat{x} = (A^Tb + \lambda C^Td) \]
Tradeoff analysis

1. Choose values for $\lambda$ (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 $\lambda$ value, find $\hat{x}_\lambda$ that minimizes $J_1 + \lambda J_2$.

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

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

\[ J_1 \rightarrow \infty \]

\[ J_2 \rightarrow \infty \]

\[ \lambda \rightarrow 0 \]

\[ \lambda \rightarrow \infty \]

better \( J_1 \)
better \( J_2 \)
worse \( J_1 \)
worse \( J_2 \)

candidate point
Pareto curve

\[ J_2 \]
\[ J_1 \]

\[ \lambda \to 0 \]
\[ \lambda \to \infty \]

feasible, but strictly suboptimal

Pareto-optimal points

infeasible
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: \( t = 0, 1, 2, \ldots, T \).
- Important variables: position \( x_t \), velocity \( v_t \), thrust \( u_t \).
- Simplified model of the dynamics:
  \[
  x_{t+1} = x_t + v_t \\
  v_{t+1} = v_t + u_t
  \]
  for \( t = 0, 1, \ldots, T - 1 \)

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

**First model:** hit the waypoints exactly

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

Julia model: Hovercraft.ipynb
Example: hovercraft

Second model: allow waypoint misses

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

- $\lambda$ controls the tradeoff between making $u$ small and hitting all the waypoints.