Lecture 8
Least-norm solutions of undetermined equations

- least-norm solution of underdetermined equations
- minimum norm solutions via $QR$ factorization
- derivation via Lagrange multipliers
- relation to regularized least-squares
- general norm minimization with equality constraints
Underdetermined linear equations

we consider

\[ y = Ax \]

where \( A \in \mathbb{R}^{m \times n} \) is fat \((m < n)\), i.e.,

- there are more variables than equations
- \( x \) is underspecified, i.e., many choices of \( x \) lead to the same \( y \)

we’ll assume that \( A \) is full rank \((m)\), so for each \( y \in \mathbb{R}^m \), there is a solution

set of all solutions has form

\[
\{ x \mid Ax = y \} = \{ x_p + z \mid z \in \mathcal{N}(A) \}
\]

where \( x_p \) is any (‘particular’) solution, i.e., \( Ax_p = y \)
• $z$ characterizes available choices in solution

• solution has $\text{dim } \mathcal{N}(A) = n - m$ ‘degrees of freedom’

• can choose $z$ to satisfy other specs or optimize among solutions
Least-norm solution

one particular solution is

\[ x_{\text{ln}} = A^T(AA^T)^{-1}y \]

\((AA^T)\) is invertible since \(A\) full rank\)

in fact, \(x_{\text{ln}}\) is the solution of \(y = Ax\) that minimizes \(\|x\|\)

\(i.e., \ x_{\text{ln}}\) is solution of optimization problem

\[
\begin{align*}
\text{minimize} & \quad \|x\| \\
\text{subject to} & \quad Ax = y \\
\end{align*}
\]

(with variable \(x \in \mathbb{R}^n\))
suppose $Ax = y$, so $A(x - x_{ln}) = 0$ and

$$(x - x_{ln})^T x_{ln} = (x - x_{ln})^T A^T (AA^T)^{-1} y$$

$$= (A(x - x_{ln}))^T (AA^T)^{-1} y$$

$$= 0$$

i.e., $(x - x_{ln}) \perp x_{ln}$, so

$$\|x\|^2 = \|x_{ln} + x - x_{ln}\|^2 = \|x_{ln}\|^2 + \|x - x_{ln}\|^2 \geq \|x_{ln}\|^2$$

i.e., $x_{ln}$ has smallest norm of any solution
\[ \left\{ x \mid Ax = y \right\} \]

\[ \mathcal{N}(A) = \left\{ x \mid Ax = 0 \right\} \]

- **orthogonality condition:** \( x_{ln} \perp \mathcal{N}(A) \)

- **projection interpretation:** \( x_{ln} \) is projection of 0 on solution set \( \left\{ x \mid Ax = y \right\} \)
• $A^\dagger = A^T(AA^T)^{-1}$ is called the \textbf{pseudo-inverse} of full rank, fat $A$

• $A^T(AA^T)^{-1}$ is a \textit{right inverse} of $A$

• $I - A^T(AA^T)^{-1}A$ gives projection onto $\mathcal{N}(A)$

cf. analogous formulas for full rank, \textbf{skinny} matrix $A$:

• $A^\dagger = (A^TA)^{-1}A^T$

• $(A^TA)^{-1}A^T$ is a \textit{left inverse} of $A$

• $A(A^TA)^{-1}A^T$ gives projection onto $\mathcal{R}(A)$
Least-norm solution via QR factorization

find $QR$ factorization of $A^T$, i.e., $A^T = QR$, with

- $Q \in \mathbb{R}^{n \times m}$, $Q^T Q = I_m$
- $R \in \mathbb{R}^{m \times m}$ upper triangular, nonsingular

then

- $x_{ln} = A^T (A A^T)^{-1} y = Q R^{-T} y$
- $\|x_{ln}\| = \|R^{-T} y\|$
Derivation via Lagrange multipliers

- least-norm solution solves optimization problem

\[
\begin{align*}
\text{minimize} & \quad x^T x \\
\text{subject to} & \quad Ax = y
\end{align*}
\]

- introduce Lagrange multipliers: \( L(x, \lambda) = x^T x + \lambda^T (Ax - y) \)

- optimality conditions are

\[
\nabla_x L = 2x + A^T \lambda = 0, \quad \nabla_\lambda L = Ax - y = 0
\]

- from first condition, \( x = -A^T \lambda / 2 \)

- substitute into second to get \( \lambda = -2(AA^T)^{-1}y \)

- hence \( x = A^T(AA^T)^{-1}y \)
Example: transferring mass unit distance

- unit mass at rest subject to forces $x_i$ for $i - 1 < t \leq i$, $i = 1, \ldots, 10$

- $y_1$ is position at $t = 10$, $y_2$ is velocity at $t = 10$

- $y = Ax$ where $A \in \mathbb{R}^{2 \times 10}$ ($A$ is fat)

- find least norm force that transfers mass unit distance with zero final velocity, i.e., $y = (1, 0)$
Relation to regularized least-squares

- suppose $A \in \mathbb{R}^{m \times n}$ is fat, full rank
- define $J_1 = \|Ax - y\|^2$, $J_2 = \|x\|^2$
- least-norm solution minimizes $J_2$ with $J_1 = 0$
- minimizer of weighted-sum objective $J_1 + \mu J_2 = \|Ax - y\|^2 + \mu \|x\|^2$ is
  
  $$x_\mu = (A^T A + \mu I)^{-1} A^T y$$

- **fact:** $x_\mu \rightarrow x_{ln}$ as $\mu \rightarrow 0$, i.e., regularized solution converges to least-norm solution as $\mu \rightarrow 0$
- in matrix terms: as $\mu \rightarrow 0$,
  
  $$\left( A^T A + \mu I \right)^{-1} A^T \rightarrow A^T \left( AA^T \right)^{-1}$$

  (for full rank, fat $A$)
General norm minimization with equality constraints

consider problem

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

with variable \( x \)

• includes least-squares and least-norm problems as special cases

• equivalent to

\[
\begin{align*}
\text{minimize} & \quad (1/2)\|Ax - b\|^2 \\
\text{subject to} & \quad Cx = d
\end{align*}
\]

• Lagrangian is

\[
L(x, \lambda) = (1/2)\|Ax - b\|^2 + \lambda^T(Cx - d)
\]

\[
= (1/2)x^T A^T A x - b^T A x + (1/2)b^T b + \lambda^T C x - \lambda^T d
\]
• optimality conditions are

\[ \nabla_x L = A^T A x - A^T b + C^T \lambda = 0, \quad \nabla_\lambda L = C x - d = 0 \]

• write in block matrix form as

\[
\begin{bmatrix}
A^T A & C^T \\
C & 0
\end{bmatrix}
\begin{bmatrix}
x \\
\lambda
\end{bmatrix}
=
\begin{bmatrix}
A^T b \\
d
\end{bmatrix}
\]

• if the block matrix is invertible, we have

\[
\begin{bmatrix}
x \\
\lambda
\end{bmatrix}
=
\begin{bmatrix}
A^T A & C^T \\
C & 0
\end{bmatrix}^{-1}
\begin{bmatrix}
A^T b \\
d
\end{bmatrix}
\]
if $A^TA$ is invertible, we can derive a more explicit (and complicated) formula for $x$

- from first block equation we get

$$x = (A^TA)^{-1}(A^Tb - C^T\lambda)$$

- substitute into $Cx = d$ to get

$$C(A^TA)^{-1}(A^Tb - C^T\lambda) = d$$

so

$$\lambda = (C(A^TA)^{-1}C^T)^{-1}(C(A^TA)^{-1}A^Tb - d)$$

- recover $x$ from equation above (not pretty)

$$x = (A^TA)^{-1} \left( A^Tb - C^T (C(A^TA)^{-1}C^T)^{-1} (C(A^TA)^{-1}A^Tb - d) \right)$$