

Least-norm solutions of underdetermined equations

- ▶ least-norm solution of underdetermined equations
- ▶ 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

Underdetermined linear equations

set of all solutions has form

$$\{ x \mid Ax = y \} = \{ x_p + z \mid z \in \mathbf{null}(A) \}$$

where x_p is any ('particular') solution, *i.e.*, $Ax_p = y$

- ▶ z characterizes available choices in solution
- ▶ solution has $\dim \mathbf{null}(A) = n - m$ 'degrees of freedom'
- ▶ can choose z to satisfy other specs or optimize among solutions

Least-norm solution

one particular solution is

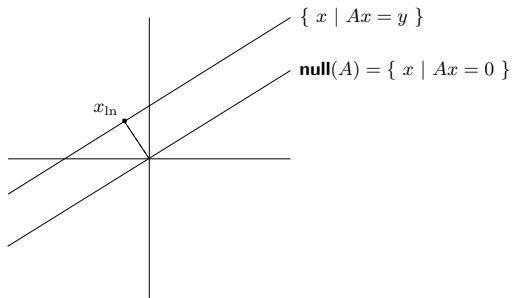
$$x_{\text{ln}} = A^{\top}(AA^{\top})^{-1}y$$

- ▶ $AA^{\top}$ is invertible since A full rank
- ▶ in fact, x_{ln} is the solution of $y = Ax$ that minimizes $\|x\|$
- ▶ *i.e.*, x_{ln} is solution of optimization problem

$$\begin{array}{ll}\text{minimize} & \|x\| \\ \text{subject to} & Ax = y\end{array}$$

(with variable $x \in \mathbb{R}^n$)

Orthogonality



suppose $Ax = y$, so $A(x - x_{ln}) = 0$ and

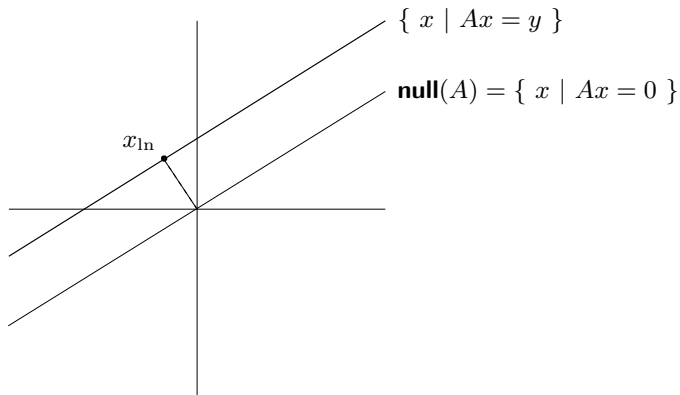
$$\begin{aligned}(x - x_{ln})^T x_{ln} &= (x - x_{ln})^T A^T (AA^T)^{-1} y \\ &= (A(x - x_{ln}))^T (AA^T)^{-1} y \\ &= 0\end{aligned}$$

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

Orthogonality



- *orthogonality condition:* $x_{\text{ln}} \perp \text{null}(A)$
- *projection interpretation:* x_{ln} is projection of 0 on solution set $\{ x \mid Ax = y \}$

Comparison with least-squares

- ▶ $A^\dagger = A^\top(AA^\top)^{-1}$ is called the *pseudo-inverse* of full rank, fat A
- ▶ $A^\top(AA^\top)^{-1}$ is a *right inverse* of A
- ▶ $I - A^\top(AA^\top)^{-1}A$ gives projection onto **null**(A)

cf. analogous formulas for full rank, **skinny** matrix A :

- ▶ $A^\dagger = (A^\top A)^{-1}A^\top$
- ▶ $(A^\top A)^{-1}A^\top$ is a *left inverse* of A
- ▶ $A(A^\top A)^{-1}A^\top$ gives projection onto **range**(A)

Derivation via Lagrange multipliers

- ▶ least-norm solution solves optimization problem

$$\begin{array}{ll}\text{minimize} & x^\top x \\ \text{subject to} & Ax = y\end{array}$$

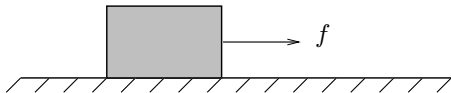
- ▶ introduce Lagrange multipliers: $L(x, \lambda) = x^\top x + \lambda^\top (Ax - y)$

- ▶ optimality conditions are

$$\nabla_x L = 2x + A^\top \lambda = 0, \quad \nabla_\lambda L = Ax - y = 0$$

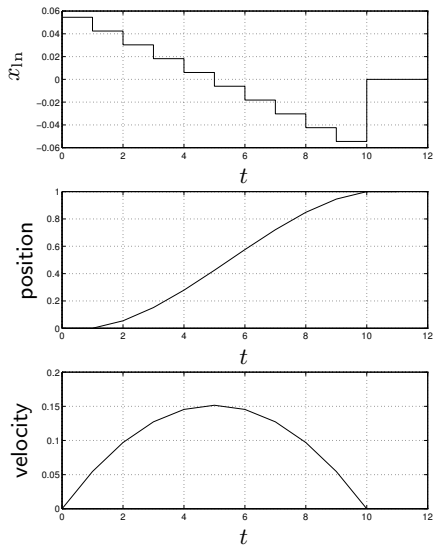
- ▶ from first condition, $x = -A^\top \lambda / 2$
- ▶ substitute into second to get $\lambda = -2(AA^\top)^{-1}y$
- ▶ hence $x = A^\top (AA^\top)^{-1}y$

Example: transferring mass unit distance



- ▶ unit mass at rest subject to forces x_i for $i - 1 < t \leq i$, $i = 1, \dots, 10$
- ▶ 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.*, $y = (1, 0)$

Example: transferring mass unit distance



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

$$x_\mu = (A^\top A + \mu I)^{-1} A^\top y$$

- ▶ **fact:** $x_\mu \rightarrow x_{\text{ln}}$ as $\mu \rightarrow 0$, *i.e.*, regularized solution converges to least-norm solution as $\mu \rightarrow 0$
- ▶ in matrix terms: as $\mu \rightarrow 0$,

$$(A^\top A + \mu I)^{-1} A^\top \rightarrow A^\top (AA^\top)^{-1}$$

(for full rank, fat A)

General norm minimization with equality constraints

consider problem

$$\begin{array}{ll}\text{minimize} & \|Ax - b\| \\ \text{subject to} & Cx = d\end{array}$$

with variable x

- ▶ includes least-squares and least-norm problems as special cases
- ▶ equivalent to

$$\begin{array}{ll}\text{minimize} & (1/2)\|Ax - b\|^2 \\ \text{subject to} & Cx = d\end{array}$$

General norm minimization with equality constraints

- Lagrangian is

$$\begin{aligned}L(x, \lambda) &= (1/2)\|Ax - b\|^2 + \lambda^\top(Cx - d) \\&= (1/2)x^\top A^\top Ax - b^\top Ax + (1/2)b^\top b + \lambda^\top Cx - \lambda^\top d\end{aligned}$$

- optimality conditions are

$$\nabla_x L = A^\top Ax - A^\top b + C^\top \lambda = 0, \quad \nabla_\lambda L = Cx - d = 0$$

- write in block matrix form as

$$\begin{bmatrix} A^\top A & C^\top \\ C & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} A^\top b \\ d \end{bmatrix}$$

- if the block matrix is invertible, we have

$$\begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} A^\top A & C^\top \\ C & 0 \end{bmatrix}^{-1} \begin{bmatrix} A^\top b \\ d \end{bmatrix}$$

Explicit formulae

if $A^T A$ is invertible, we can derive a more explicit (and complicated) formula for x

- ▶ from first block equation we get

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

- ▶ substitute into $Cx = d$ to get

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

so

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

- ▶ recover x from equation above (not pretty)

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