Least-norm solutions of underdetermined 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

Underdetermined linear equations

set of all solutions has form

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

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

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

Least-norm solution

one particular solution is

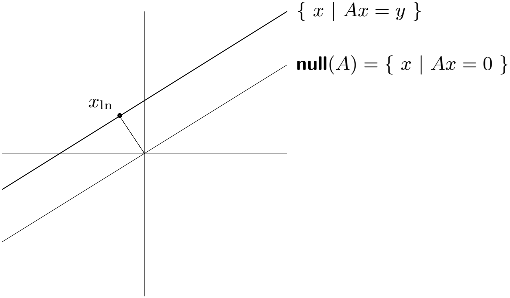
$$x_{\mathrm{ln}} = A^{\mathsf{T}}(AA^{\mathsf{T}})^{-1}y$$

- AA^{T} 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 $A(x - x_{\rm ln}) = 0$ and

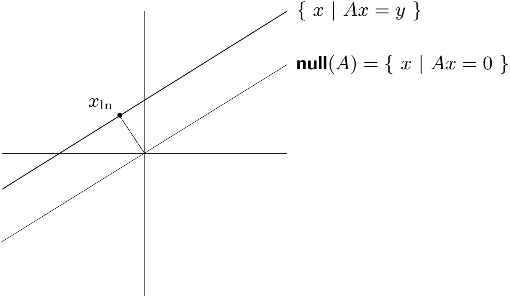
$$\begin{aligned} (x - x_{\rm ln})^T x_{\rm ln} &= (x - x_{\rm ln})^T A^T (AA^T)^{-1} y \\ &= (A(x - x_{\rm ln}))^T (AA^T)^{-1} y \\ &= 0 \end{aligned}$$

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

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

i.e., $x_{\rm ln}$ has smallest norm of any solution

Orthogonality



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

Comparison with least-squares

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

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

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

Least-norm solution via QR factorization

find QR factorization of A^T, *i.e.*, $A^\mathsf{T} = QR$, with

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

then

- $x_{\text{ln}} = A^\mathsf{T}(AA^\mathsf{T})^{-1}y = QR^{-T}y$
- $\|x_{\text{ln}}\| = \|R^{-T}y\|$

Derivation via Lagrange multipliers

- least-norm solution solves optimization problem

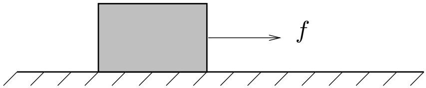
$$\begin{array}{ll} \text{minimize} & x^\mathsf{T} x \\ \text{subject to} & Ax = y \end{array}$$

- introduce Lagrange multipliers: $L(x, \lambda) = x^\mathsf{T} x + \lambda^\mathsf{T}(Ax - y)$
- optimality conditions are

$$\nabla_x L = 2x + A^\mathsf{T}\lambda = 0, \qquad \nabla_\lambda L = Ax - y = 0$$

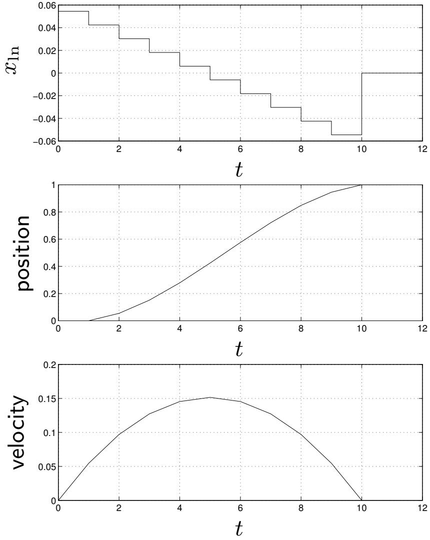
- from first condition, $x = -A^\mathsf{T}\lambda/2$
- substitute into second to get $\lambda = -2(AA^\mathsf{T})^{-1}y$
- hence $x = A^\mathsf{T}(AA^\mathsf{T})^{-1}y$

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.*, $y = (1, 0)$

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

$$x_\mu = \left(A^\mathsf{T} A + \mu I\right)^{-1} A^\mathsf{T} y$$

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

$$\left(A^\mathsf{T} A + \mu I\right)^{-1} A^\mathsf{T} \to A^\mathsf{T} \left(A A^\mathsf{T}\right)^{-1}$$

  (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

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

- optimality conditions are

$$\nabla_x L = A^\mathsf{T}Ax - A^\mathsf{T}b + C^\mathsf{T}\lambda = 0, \qquad \nabla_\lambda L = Cx - d = 0$$

- write in block matrix form as

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

- if the block matrix is invertible, we have

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

Explicit formulae

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

- from first block equation we get

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

- substitute into $Cx = d$ to get

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

so

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

- recover x from equation above (not pretty)

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