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

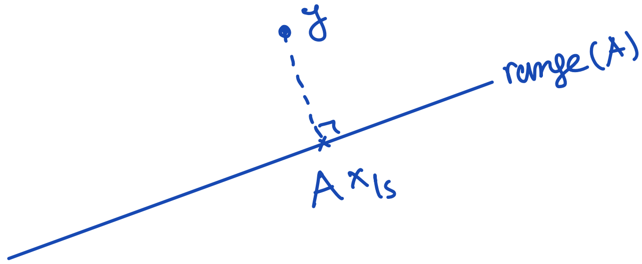
least-squares

$$\min_x \|Ax - y\|$$

A : skinny (& full-rank)

$$x_{ls} = (A^TA)^{-1}A^Ty$$

projection on the range

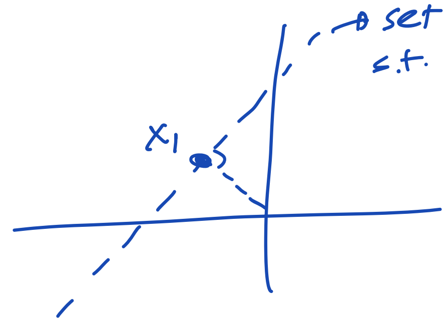


least-norm

$$\min_x \|x\| \quad \text{s.t.} \quad Ax = y$$

A : fat (& full-rank)

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



* multi objective : $\min \; J_1 + \mu J_2$

$J_1 \to \|Ax-y\|^2$, $\quad J_2 \to \|Fx-g\|^2$

$\Downarrow$

solve the LS problem

for $\tilde{A}x = \tilde{y}$

$$\tilde{A} = \begin{bmatrix} A \\ \sqrt{\mu}F \end{bmatrix}, \quad \tilde{y} = \begin{bmatrix} y \\ \sqrt{\mu}g \end{bmatrix}$$

$y = Ax$

$A \in \mathbb{R}^{m \times n}$

Possible # of solutions

$m > n$	$m = n$	$m < n$
$0, 1, \infty$	$0, 1, \infty$	$0, \infty$

0: $y \notin \text{range}(A)$

1: $y \in \text{range}(A)$, $\text{null}(A) = \{0\}$

∞: $y \in \text{range}(A)$, $\text{null}(A) \neq \{0\}$

$\underbrace{\text{rank}(A)}_{\leq m} + \underbrace{\dim \text{null}(A)}_{\geq n-m>0} = n$

$0 \neq z \in \text{null}(A)$

$Ax = y = A(x+z)$

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

$A(x_p + z) = \underbrace{Ax_p}_{y} + \underbrace{Az}_{0} = y$

? How to show any solution $\tilde{x}$ belongs to this set?

$A\tilde{x} = y \rightarrow z = \tilde{x} - x_p \rightarrow Az = A\tilde{x} - Ax_p = y - y = 0$

$\tilde{x} = x_p + z \rightarrow z \in \text{null}(A)$

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.*, $Ax_p = y$

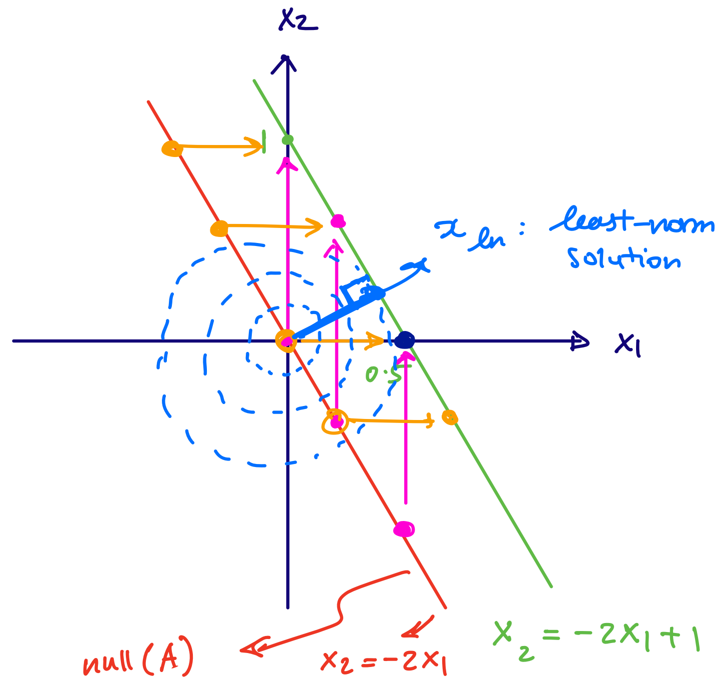
- z characterizes available choices in solution
- solution has $\mathbf{dim\,null}(A) = n - m$ 'degrees of freedom'
- can choose z to satisfy other specs or optimize among solutions

? What if we construct the set using a different particular solution $\tilde{x}_p$?

$y = Ax \qquad A \in \mathbb{R}^{1\times 2}$

$A = \begin{bmatrix} 2 & 1 \end{bmatrix}$

$y = \begin{bmatrix} 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2x_1 + x_2 = 1$



$\text{null}(A) = \{ x \mid Ax = 0 \}$

$= \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \,\middle|\, 2x_1 + x_2 = 0 \right\}$

$x_p = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

$\tilde{x}_p = \begin{bmatrix} 0.5 \\ 0 \end{bmatrix}$

$y = A x$

observe find

$x = By$, B a right inverse of A

$\Rightarrow Ax = \underbrace{AB}_{I}y = y$ ✓

$\Rightarrow$ any right inverse of A $\rightarrow$ a solution

Least-norm solution

$B = A^T (AA^T)^{-1} \rightarrow AB = AA^T(AA^T)^{-1} = I$

one particular solution is

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

$x_{ls} = (A^T A)^{-1} A^T y$

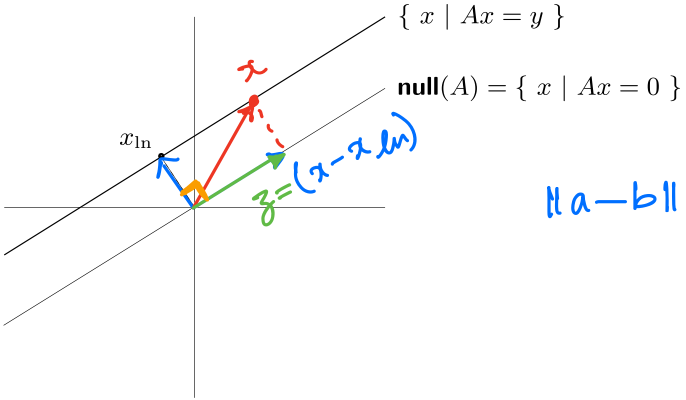
- AA^T is invertible since A full rank
- in fact, x_{ln} is the solution of $y = Ax$ that minimizes $||x||$
- *i.e.*, x_{ln} is solution of optimization problem

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

minimize $\|Ax - y\|$

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

Orthogonality



$$\|a - b\|^2 = \|a\|^2 + \|b\|^2 - 2\underbrace{a^T b}_{0}$$

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

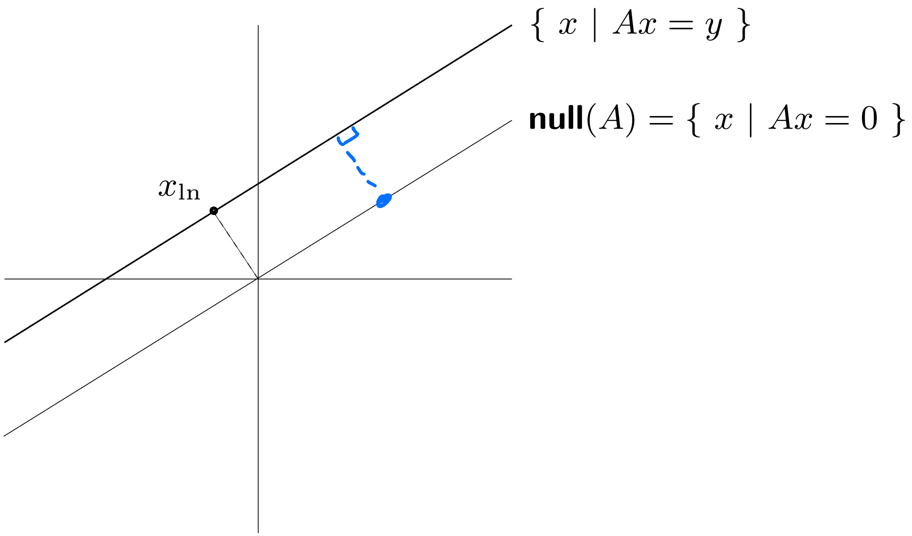
$$\begin{aligned}
(x - x_{\text{ln}})^{\mathsf{T}} x_{\text{ln}} &= (x - x_{\text{ln}})^{\mathsf{T}} A^{\mathsf{T}} (AA^{\mathsf{T}})^{-1} y \\
&= (A(x - x_{\text{ln}}))^{\mathsf{T}} (AA^{\mathsf{T}})^{-1} y \\
&= 0
\end{aligned}$$

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

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

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

Orthogonality



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

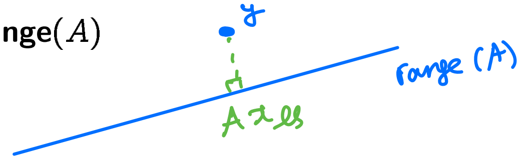
Comparison with least-squares

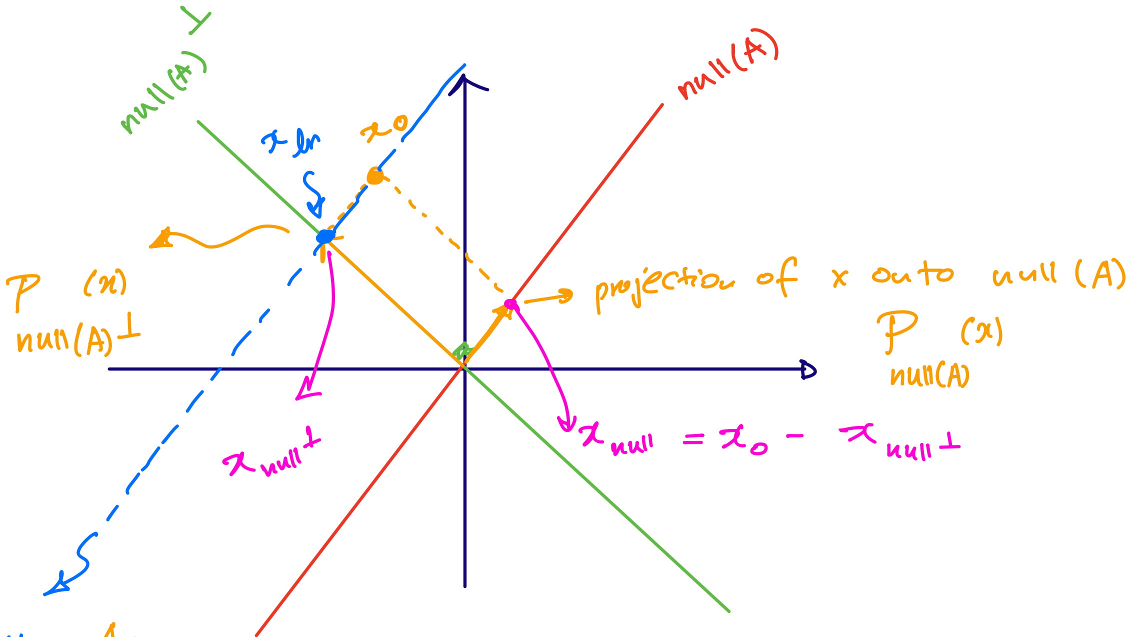
square : $A^T(AA^T)^{-1} = A^T A^{-T} A^{-1} = A^{-1}$

- $A^\dagger = A^\mathsf{T}(AA^\mathsf{T})^{-1}$ 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 **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 **range**(A)



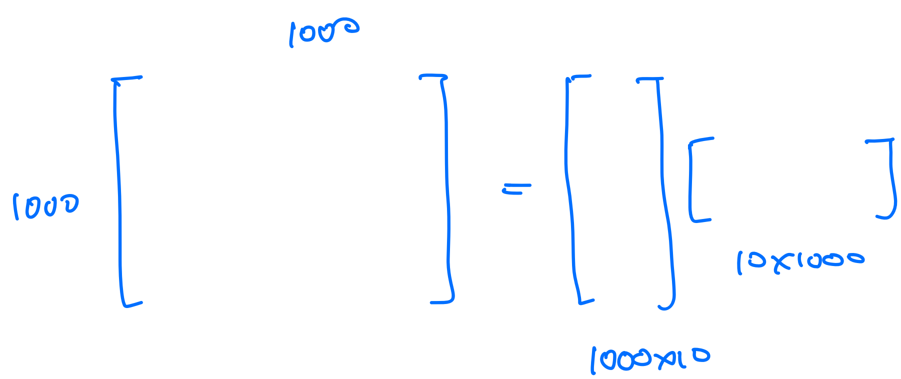


$$Ax = y = Ax_0$$

$$x_{lr} = A^T(AA^T)^{-1}y = \underbrace{\boxed{A^T(AA^T)^{-1}A}}_{\text{Projection matrix}}\, x_0 = P_{null(A)^\perp}(x_0)$$

$$P_{null(A)}(x) = x - A^T(AA^T)^{-1}Ax = \underbrace{\boxed{\left(I - A^T(AA^T)^{-1}A\right)}}_{\text{Projection Matrix}}\, x$$

$$P \rightarrow I - P$$



Derivation via Lagrange multipliers

- least-norm solution solves optimization problem

$$\begin{array}{ll} \text{minimize} & \overbrace{x^\top x}^{\|x\|^2} \\ \text{subject to} & Ax = y \end{array}$$

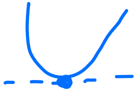
- introduce Lagrange multipliers: $L(x, \lambda) = x^\top x + \lambda^\top(Ax - y)$
- optimality conditions are

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

$$-AA^T\lambda/2 - y = 0$$

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

min $f(x)$



$f'(x) = 0$

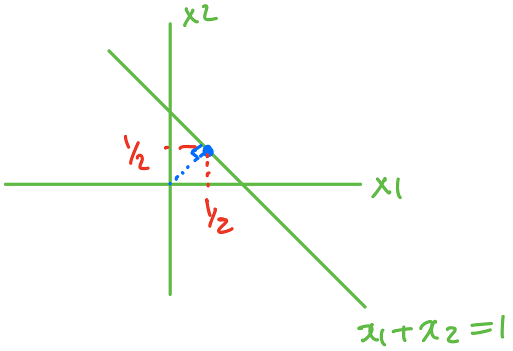
$\nabla f(x) = 0$

min $f(x)$

s.t. $g(x) = 0$

$f = a^T x$

$\nabla f = a$



$$\min \ x_1^2 + x_2^2 \quad f(x)$$

$$\text{s.t.} \ x_1 + x_2 = 1$$

$$\underbrace{x_1 + x_2 - 1 = 0}_{g(x)=0}$$

$$\min \ \underbrace{x_1^2 + x_2^2}_{\|x\|^2}$$

$$\rightarrow x^* = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\lambda_1 g_1 + \lambda_2 g_2 + \cdots$$

$$\begin{bmatrix} \lambda_1 & \lambda_2 & \cdots \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ \vdots \end{bmatrix}$$

$$\lambda^T g$$

$$\mathcal{L}(x,\lambda) = f(x) + \lambda\, g(x) \rightarrow \begin{cases} \nabla_x \mathcal{L} = 0 \\ \nabla_\lambda \mathcal{L} = 0 \end{cases}$$

Lagrangian

$$\mathcal{L}(x,\lambda) = (x_1^2 + x_2^2) + \lambda(x_1 + x_2 - 1)$$

$$\nabla_x \mathcal{L} = \begin{bmatrix} 2x_1 + \lambda \\ 2x_2 + \lambda \end{bmatrix} = 0 \quad , \quad \nabla_\lambda \mathcal{L} = x_1 + x_2 - 1 = 0$$

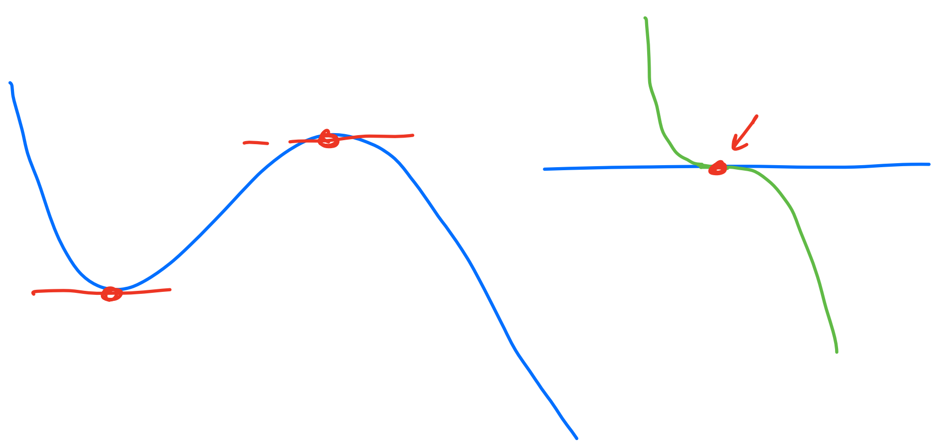
$$\begin{cases} x_1 = -\frac{\lambda}{2} \\ x_2 = -\frac{\lambda}{2} \end{cases}$$

$$\frac{-\lambda}{2} - \frac{\lambda}{2} - 1 = 0$$

$$\boxed{\lambda = -1}$$

$$x = \frac{-\lambda}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\rightarrow \boxed{x = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}}$$



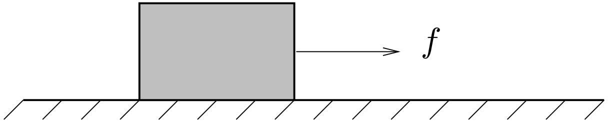
Example: transferring mass unit distance

$$\begin{bmatrix} p \\ v \end{bmatrix} = A x$$

$$\begin{bmatrix} 19/2 & 17/2 & \cdots & 3/2 & 1/2 \\ \hline 1 & 1 & \cdots & 1 & 1 \end{bmatrix}$$

$x \in \mathbb{R}^{10}$

$$\begin{bmatrix} x_1 \\ \vdots \\ x_{10} \end{bmatrix}$$



$Ax = y = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$A^T (AA^T)^{-1} y$

- 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

$$\min \; \|x\|_2^2 \quad \longleftrightarrow$$

s.t. final $p=1$

final $v=0$

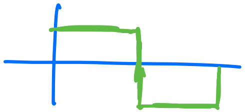
$$Ax = \begin{bmatrix}1\\0\end{bmatrix}$$

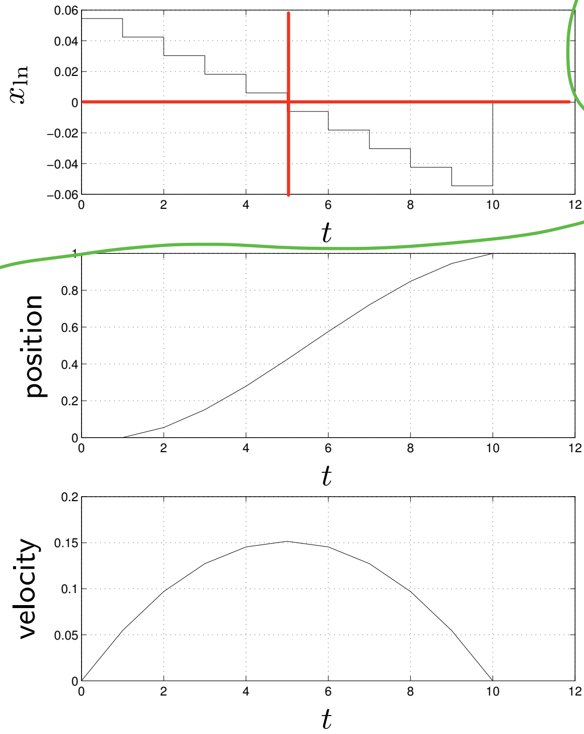
$$\min \; \|x\|_1$$

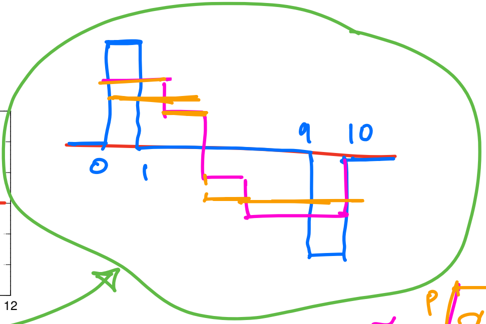
$$\text{s.t.} \quad Ax = \begin{bmatrix}1\\0\end{bmatrix}$$

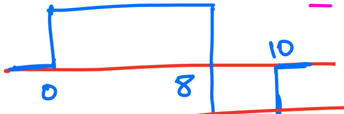
$$\min \; \|x\|_\infty$$

$$\text{s.t.} \quad Ax = \begin{bmatrix}1\\0\end{bmatrix}$$









$$\|x\|_p = \sqrt[p]{x_1^p + \cdots + x_n^p} \simeq \sqrt[p]{x_n^p}$$

$p \to \infty$

$x_n > x_i$

$$\|x\|_2 = \sqrt{x_1^2 + \cdots + x_{10}^2}$$

$$\|x\|_1 = |x_1| + \cdots + |x_{10}|$$

$$\|x\|_\infty = \lim_{p\to\infty} \|x\|_p$$

$$= \max_i |x_i|$$

Relation to regularized least-squares

$$\lim_{\mu \to 0} (A^T A + \mu I)^{-1} A^T y$$

- 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

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

$$\boxed{\begin{gathered}\lim_{\mu \to 0} \mu^2 + 1 = 1 \\ \lim_{\mu \to 1} \frac{\mu^2 - 1}{\mu - 1} = 2\end{gathered}}$$

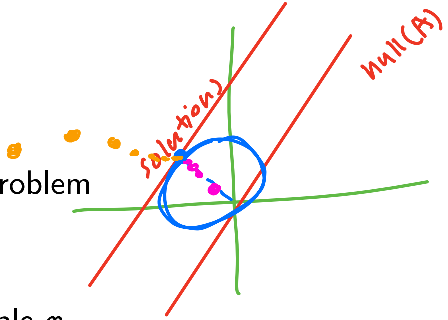
- **fact:** $x_\mu \to x_{\text{ln}}$ as $\mu \to 0$, *i.e.*, regularized solution converges to least-norm solution as $\mu \to 0$
- in matrix terms: as $\mu \to 0$,

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

(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

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

$$\begin{array}{ll} \min & \|x\|^2 \\ \text{s.t.} & Ax = y \end{array}$$

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & g(x) = 0 \end{array}$$

General norm minimization with equality constraints

- Lagrangian is

$$L(x,\lambda) = (1/2)||Ax-b||^2 + \lambda^T(Cx-d)$$
$$= (1/2)x^TA^TAx - b^TAx + (1/2)b^Tb + \lambda^TCx - \lambda^Td$$

- optimality conditions are

$$\nabla_x L = A^TA\underset{?}{x} - A^Tb + C^T\underset{?}{\lambda} = 0, \qquad \nabla_\lambda L = C\underset{?}{x} - d = 0$$

- write in block matrix form as

$$\boxed{\begin{array}{c} \\ n \\ m \end{array}\begin{array}{c} \begin{array}{cc} n & m \end{array} \\ \begin{bmatrix} A^TA & C^T \\ C & 0 \end{bmatrix} \end{array} \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} A^Tb \\ d \end{bmatrix} \begin{array}{l} \to n \\ \to m \end{array}}$$

n $\quad m\times m \quad$ m

- if the block matrix is invertible, we have

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

min $f(x)$
s.t. $\boxed{g(x) = 0}$

$\mathcal{L} = f + \lambda g$

$\nabla_x \mathcal{L} = \nabla_x f + \lambda \nabla_x g$

$\nabla_\lambda \mathcal{L} = \boxed{g = 0}$

minimize	$(1/2)\|\|Ax-b\|\|^2$
subject to	$Cx = d$

$x \in \mathbb{R}^n$

$C : m\times n \to d \in \mathbb{R}^m$

$A : k\times n \to b \in \mathbb{R}^k$

Explicit formulae

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

- from first block equation we get

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

- substitute into $Cx = d$ to get

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

so

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

- recover x from equation above (not pretty)

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

I $\quad$ 0

$$-C^\top(CC^\top)^{-1}(-d) \xrightarrow{0} \boxed{C^\top(CC^\top)^{-1}d}$$

minimize	$(1/2)\|\|Ax - b\|\|^2$
subject to	$Cx = d$

$C = 0, d = 0 \rightarrow$ LS

$A = I, b = 0 \rightarrow$ LN