Multi-objective least-squares

Stephen Boyd and Sanjay Lall

EE263
Stanford University

Multi-objective least-squares

in many problems we have two (or more) objectives

- we want $J_1 = ||Ax - y||^2$ small
- and also $J_2 = ||Fx - g||^2$ small

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

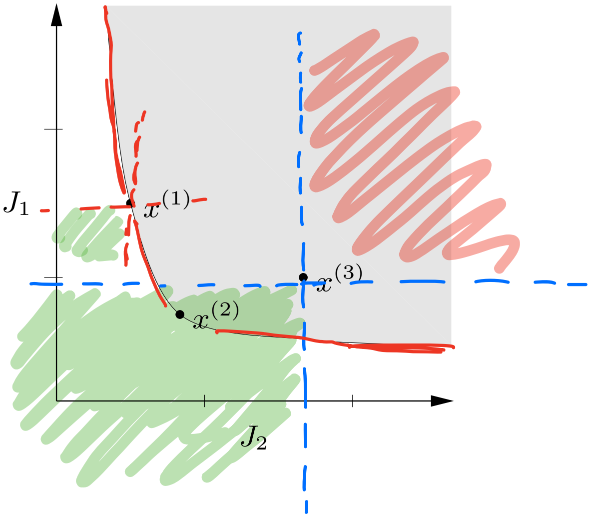
- usually the objectives are *competing*
- we can make one smaller, at the expense of making the other larger

common example: $F = I$, $g = 0$; we want $||Ax - y||$ small, with small x

Plot of achievable objective pairs

$J_1 = \| Ax - y \|$

$J_2 = \| Fx - g \|$



- plot (J_2, J_1) for every x
- $x \in \mathbb{R}^n$, but this plot is in $\mathbb{R}^2$
- point labeled $x^{(1)}$ is really $\big(J_2(x^{(1)}), J_1(x^{(1)})\big)$

$\underline{x \in \mathbb{R}}$

$$J_1(x) = (x-1)^2$$

$$J_2(x) = x^2$$

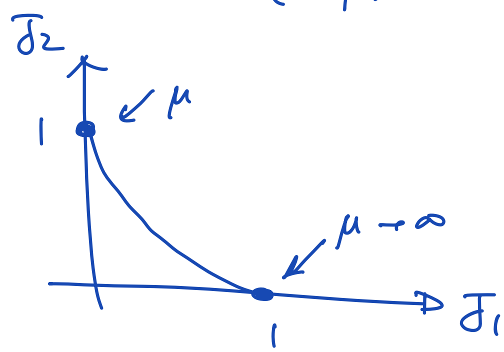
$$J_\mu(x) = J_1 + \mu J_2 = (x-1)^2 + \mu x^2$$

$$= (1+\mu)\,x^2 - 2x + 1$$

$$\min J_\mu \longrightarrow x_\mu = \frac{1}{1+\mu} \longrightarrow J_1(x_\mu) = \left(\frac{1}{1+\mu} - 1\right)^2 = \frac{1}{(1+1/\mu)^2}$$

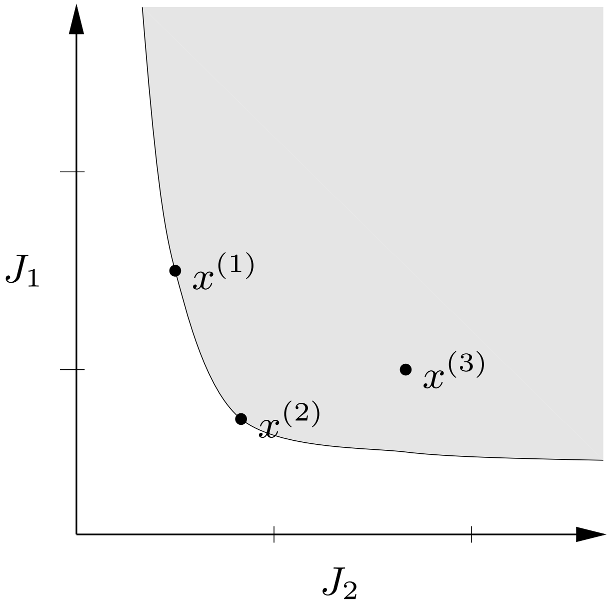
$$2(1+\mu)x - 2 = 0$$

$$J_2(x_\mu) = \frac{1}{(1+\mu)^2}$$



Optimal trade-off curve

- shaded area shows (J_2, J_1) achieved by some $x \in \mathbb{R}^n$
- clear area shows (J_2, J_1) not achieved by any $x \in \mathbb{R}^n$
- boundary of region is called *optimal trade-off curve*
- corresponding x are called *Pareto optimal* for the two objectives $||Ax - y||^2$, $||Fx - g||^2$



three example choices of x: $x^{(1)}$, $x^{(2)}$, $x^{(3)}$

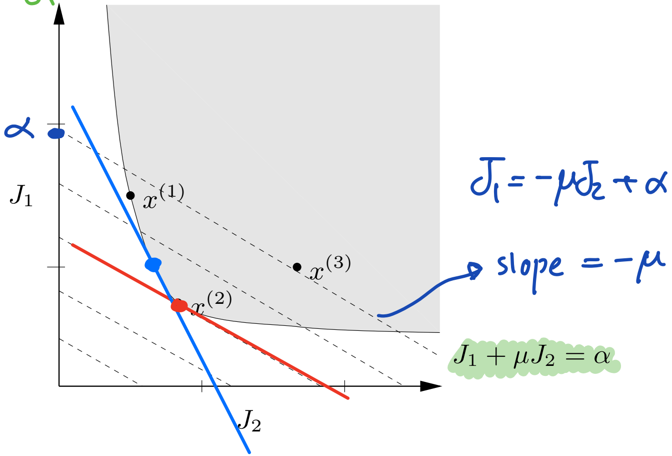
- $x^{(3)}$ is worse than $x^{(2)}$ on both counts (J_2 and J_1)
- $x^{(1)}$ is better than $x^{(2)}$ in J_2, but worse in J_1

Weighted-sum objective

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

$$\|y\|^2 = \underbrace{y_1^2 + y_2^2}_{\|z_1\|^2} + \underbrace{\cdots + y_4^2}_{+\|z_2\|^2} \qquad y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \begin{matrix} \to z_1 \\ \\ \to z_2 \\ \end{matrix}$$

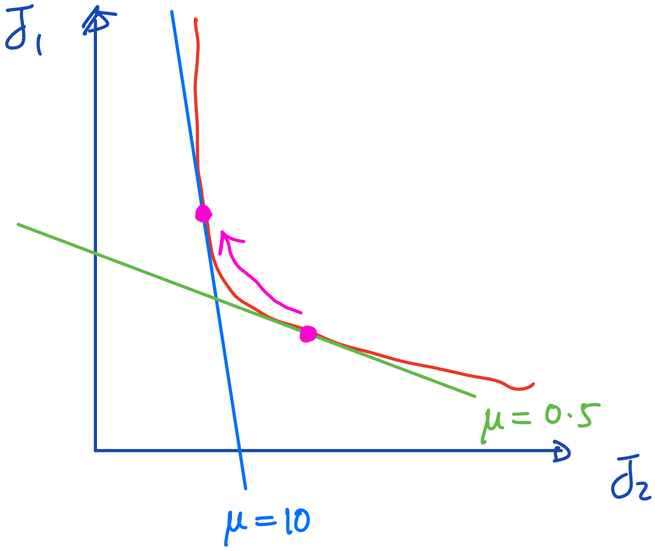
- to find Pareto optimal points,
  $i.e.$, x's on optimal trade-off curve,
  we minimize *weighted-sum objective*

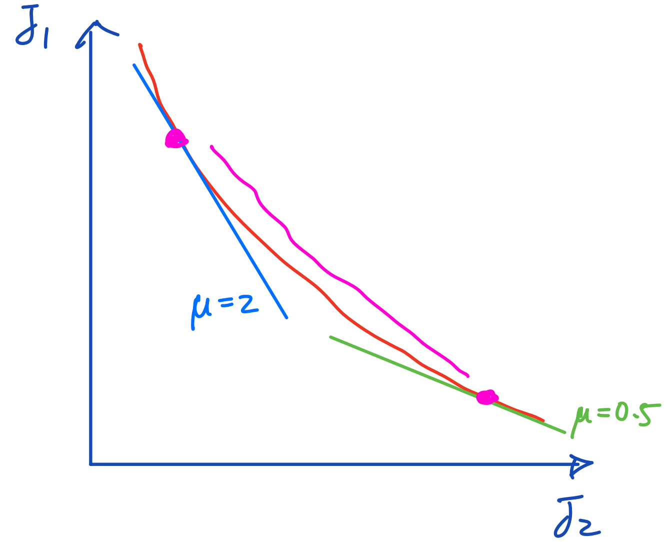


$$J_1 + \mu J_2 = \|Ax - y\|^2 + \mu\|Fx - g\|^2$$

$$\tilde{A}x = \begin{bmatrix} Ax \\ \sqrt{\mu}\, Fx \end{bmatrix} \qquad = \|\tilde{A}x - \tilde{y}\|^2$$

- parameter $\mu \geq 0$ gives relative weight between J_1 and J_2
- points where weighted sum is constant, $J_1 + \mu J_2 = \alpha$, correspond to line with slope $-\mu$ on (J_2, J_1) plot
- $x^{(2)}$ minimizes weighted-sum objective for μ shown
- by varying μ from 0 to $+\infty$, can sweep out entire *optimal tradeoff curve*





which one is more sensitive to μ?

Minimizing weighted-sum objective

$\|v\|^2 + \|u\|^2 = \left\| \begin{bmatrix} v \\ u \end{bmatrix} \right\|^2$

can express weighted-sum objective as ordinary least-squares objective:

$$\|Ax - y\|^2 + \mu\|Fx - g\|^2 = \left\| \begin{bmatrix} A \\ \sqrt{\mu}F \end{bmatrix} x - \begin{bmatrix} y \\ \sqrt{\mu}g \end{bmatrix} \right\|^2$$

$$= \|\tilde{A}x - \tilde{y}\|^2$$

where

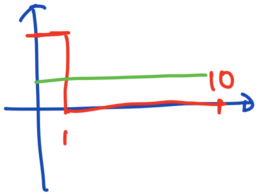
$$\tilde{A} = \begin{bmatrix} A \\ \sqrt{\mu}F \end{bmatrix}, \qquad \tilde{y} = \begin{bmatrix} y \\ \sqrt{\mu}g \end{bmatrix}$$

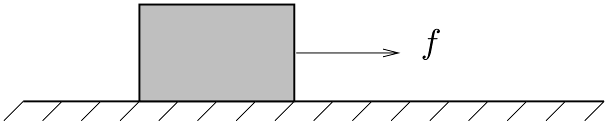
$\begin{bmatrix} A^T & \sqrt{\mu}F^T \end{bmatrix} \begin{bmatrix} A \\ \sqrt{\mu}F \end{bmatrix}$

hence solution is (assuming $\tilde{A}$ full rank)

$$x = \left(\tilde{A}^T \tilde{A}\right)^{-1} \tilde{A}^T \tilde{y}$$
$$= \left(A^T A + \mu F^T F\right)^{-1} \left(A^T y + \mu F^T g\right) \quad = x_\mu$$

Example





$a^T = A = \begin{bmatrix} 19/2 & 17/2 & \cdots & 3/2 & 1/2 \end{bmatrix}$

$\|y - Ax\|^2$

- unit mass at rest subject to forces x_i for $i - 1 < t \leq i$, $i = 1, \ldots, 10$
- $y \in \mathbb{R}$ is position at $t = 10$; $y = a^T x$ where $a \in \mathbb{R}^{10}$
- $J_1 = (y - 1)^2$ (final position error squared)
- $J_2 = ||x||^2$ (sum of squares of forces)

$J_1 = \|Ax - y\|^2 = \|a^T x - 1\|^2$

$J_2 = \|Fx - g\|^2 = \|x\|^2$

$$x = \left(\tilde{A}^T \tilde{A}\right)^{-1} \tilde{A}^T \tilde{y}$$
$$= \left(A^T A + \mu F^T F\right)^{-1} \left(A^T y + \mu F^T g\right)$$

$A = a^T$

$F = I$

$y = 1$

$g = 0$

weighted-sum objective: $(a^T x - 1)^2 + \mu ||x||^2$

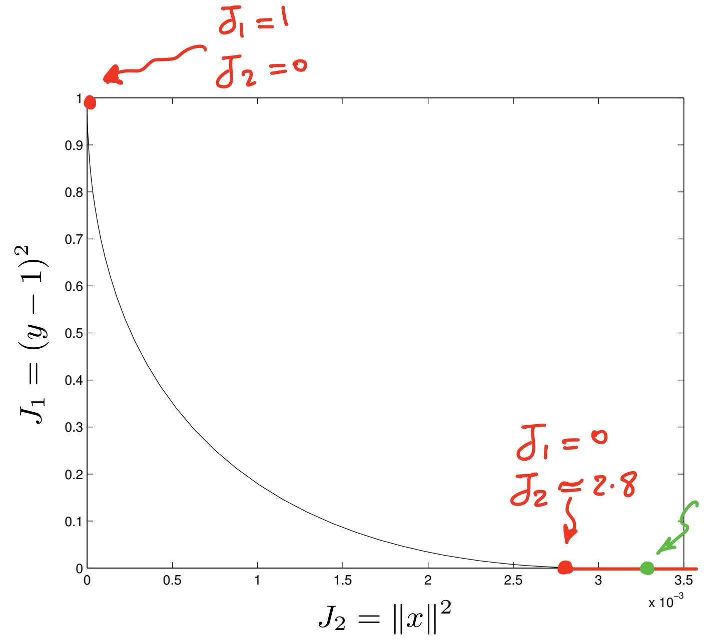
optimal x:

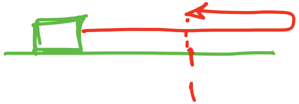
$$x = \left(a a^T + \mu I\right)^{-1} a$$

Optimal trade-off curve

$J_1 = (y-1)^2$

$J_2 = \|x\|^2$





- upper left corner of optimal trade-off curve corresponds to $x = 0$
- bottom right corresponds to input that yields $y = 1$, *i.e.*, $J_1 = 0$

Regularized least-squares

$$\|Ax - y\|^2 + \mu \|x\|^2$$

when $F = I$, $g = 0$ the objectives are

$$J_1 = \|Ax - y\|^2, \qquad J_2 = \|x\|^2$$

$$f(t) = -100\,t^9 + 85\,t^8 + \cdots$$

$x_9 \quad x_8 \quad \cdots$

minimizer of weighted-sum objective,

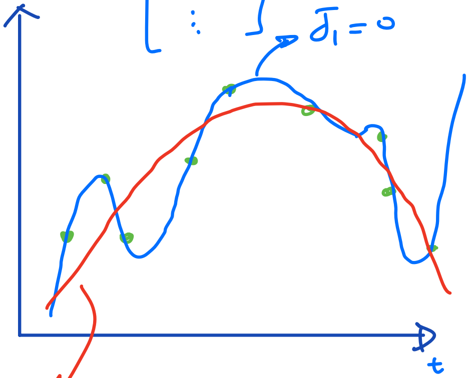
$$x = \left(A^\top A + \mu I\right)^{-1} A^\top y,$$

$$x = \begin{bmatrix} -100 \\ 85 \\ \vdots \end{bmatrix} \rightarrow J_2 : \text{large}$$

$J_1 = 0$

is called *regularized* least-squares (approximate) solution of $Ax \approx y$

- also called *Tychonov regularization*
- for $\mu > 0$, works for *any* A (no restrictions on shape, rank ...)



estimation/inversion application:

- $Ax - y$ is sensor residual
- prior information: x small
- regularized solution trades off sensor fit, size of x

$$x = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ -2 \\ 1 \\ 3 \end{bmatrix} \rightarrow J_2 : \text{small}$$

J_1 : small

$$x = (\tilde{A}^\top \tilde{A})^{-1} \tilde{A}^\top \tilde{y}$$
$$= (A^\top A + \mu F^\top F)^{-1} (A^\top y + \mu F^\top g)$$

$$\tilde{A} = \begin{bmatrix} A \\ \sqrt{\mu}\, I \end{bmatrix}$$

indep. columns:

$Bx = 0 \longrightarrow x = 0$

reg. least squares: $F = I$, $g = 0$

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

$$Bx = 0 \longrightarrow A^T A x + \mu x = 0$$

x^T

$$\Rightarrow \underbrace{x^T A^T}_{(Ax)^T} \underbrace{Ax} + \mu x^T x = 0$$

$$\Rightarrow \underbrace{\|Ax\|^2}_{\geq 0} + \mu \underbrace{\|x\|^2}_{\geq 0} = 0$$

$$\Rightarrow \begin{cases} \|Ax\| = 0 \\ \|x\| = 0 \end{cases} \longrightarrow \boxed{x = 0}$$