6. Multiobjective Least-Squares

- Example: finding an orthonormal basis
- Example: all solutions to a control problem
- Example: ill-conditioning
- Example: residuals
- Floating-point numbers
- Relative errors
- The condition number
- Example: condition number
- Multi-objective least squares
- Trade-off curves
- Weighted-sum objectives
- Example: trade-off curve

The Key Points of This Section

- for *physical problems*, absolute errors often matter
- when computing things *numerically*, it's *relative errors* that matter

- in practice, we don't just minimize error, we want to make *trade-offs*
  e.g., keep error small, but only if we can do so with a small input
- to find a trade-off, plot the trade-off curve
- do this by minimizing the *weighted-sum objective*

bounds on relative error

suppose we are computing $y = Ax$, and A is square and invertible

perturbing x to $x + \delta x$ results in y changing to $y + \delta y$

relative error is

$$\frac{\|\delta y\|}{\|y\|} \Big/ \frac{\|\delta x\|}{\|x\|} = \frac{\|A(x+\delta x) - Ax\|}{\|Ax\|} \Big/ \frac{\|\delta x\|}{\|x\|}$$

$$= \frac{\|A\delta x\|}{\|\delta x\|} \Big/ \frac{\|Ax\|}{\|x\|}$$

$$\leq \|A\|\|A^{-1}\|$$

because

$$\frac{\|x\|}{\|Ax\|} = \frac{\|A^{-1}Ax\|}{\|Ax\|} \leq \|A^{-1}\|$$

The quantity $\kappa(A) = \|A\|\|A^{-1}\|$ is called the *condition number* of A.

properties of the condition number

- $\kappa(A) = \|A\|\|A^{-1}\|$ if A is square and invertible
- otherwise $\kappa(A) = \|A\|\|A^{\dagger}\| = \dfrac{\sigma_{\max}(A)}{\sigma_{\min}(A)}$
- if $\kappa(A)$ is small we call A *well-conditioned*, otherwise we say A is *ill-conditioned*.
- $\kappa(A)$ is the *eccentricity* of the ellipse that is the image of the unit ball under A.

uses of the condition number

- relative error computing y using $y = Ax$ is $\dfrac{\|\delta y\|}{\|y\|} \Big/ \dfrac{\|\delta x\|}{\|x\|} \leq \kappa(A)$
- relative error computing x from $y = Ax$ is $\dfrac{\|\delta x\|}{\|x\|} \Big/ \dfrac{\|\delta y\|}{\|y\|} \leq \kappa(A)$
- relative error computing x from $y = Ax$ with *errors in A* is

$$\frac{\|\delta x\|}{\|x\|} \Big/ \frac{\|\delta A\|}{\|A\|} \leq \kappa(A)$$

Multi-Objective Least Squares

in many problems we have two (or more) objectives

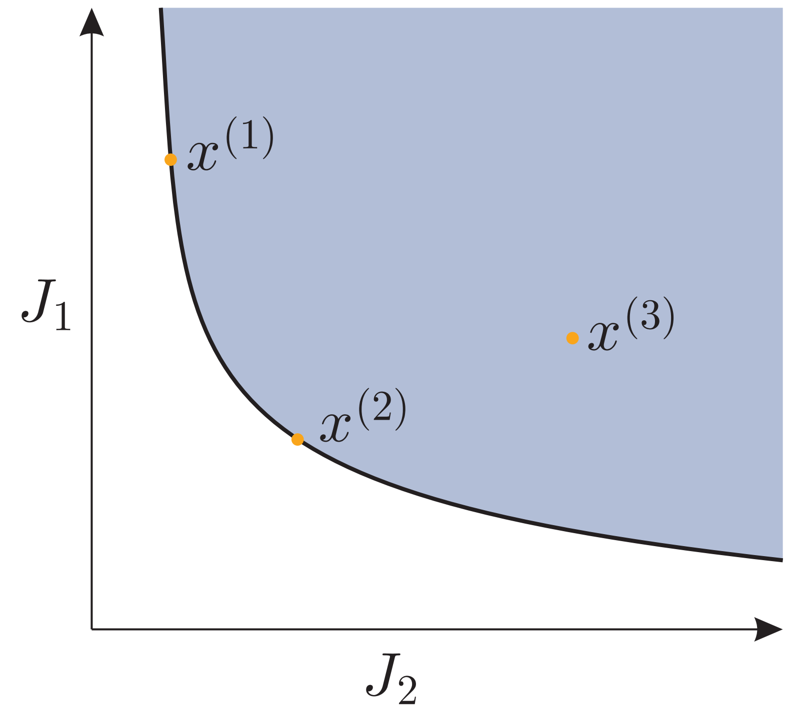
- we want $J_1 = \|Ax - y\|^2$ small
- and also $J_2 = \|Fx - g\|^2$ small

- 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

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 called *Pareto optimal* (for the objectives $\|Ax - y\|^2$, $\|Fx - g\|^2$)



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

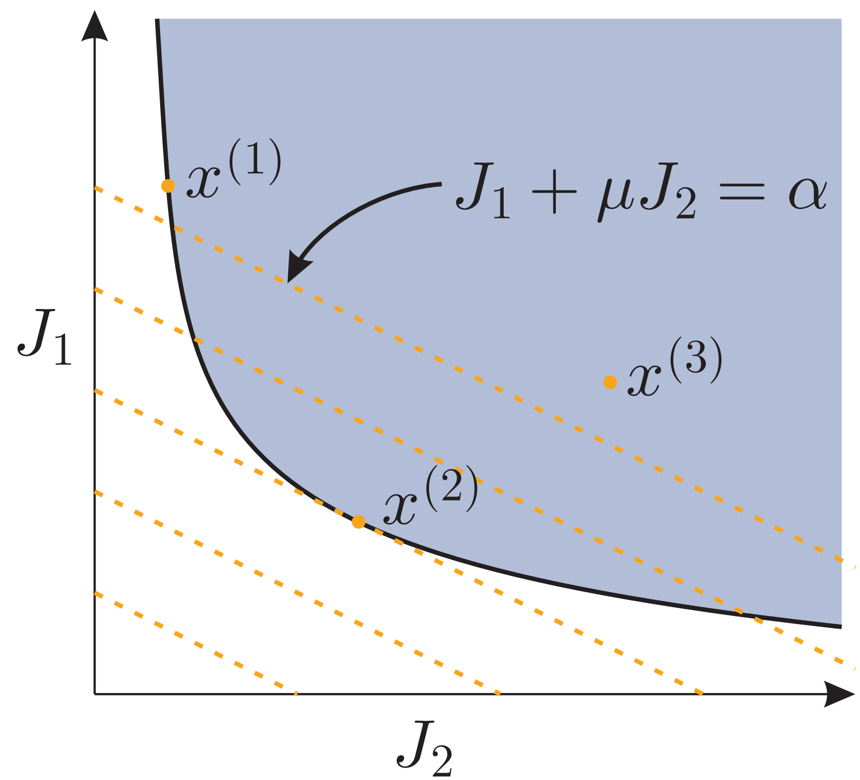
- $x^{(3)}$ is worse than $x^{(2)}$ on both counts (J_2 and J_1)
- $x^{(1)}$ is better than $x^{(2)}$ in J_2, but worse in J_1

weighted-sum objective

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

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

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 $-\mu$

- $x^{(2)}$ minimizes the weighted-sum objective for μ shown
- by varying μ from 0 to $+\infty$, we can sweep out the entire *optimal trade-off curve*

minimizing weighted-sum objective

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

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

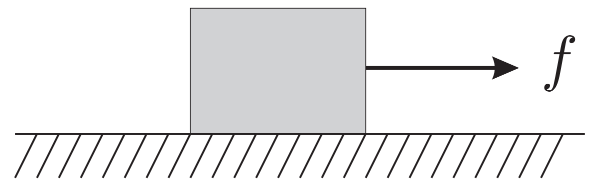
where

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

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

$$\begin{aligned} x &= \tilde{A}^\dagger \tilde{y} \\ &= (A^TA+\mu F^TF)^{-1}(A^Ty+\mu F^Tg)\end{aligned}$$

example: forces on mass



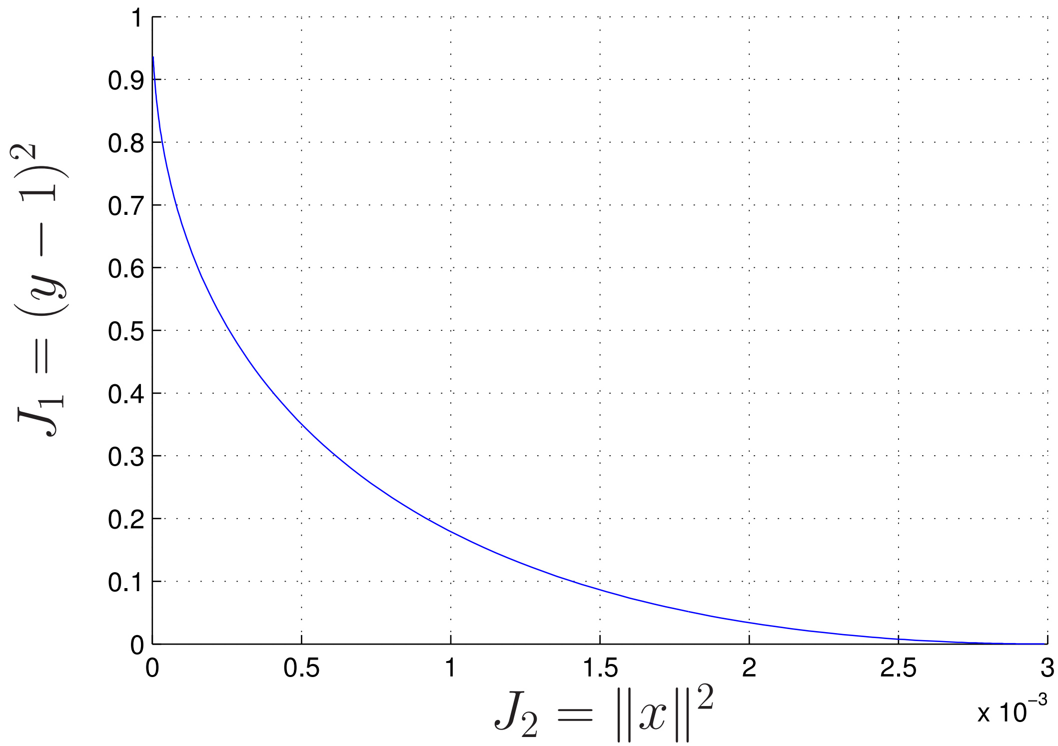
- unit mass at rest, subject to forces x_i for $i-1 \leq t \leq i$, for $i = 1, \ldots, 10$
- $y \in \mathbb{R}$ is position at $t = 10$; $y = a^T x$ where $a \in \mathbb{R}^{10}$
- $J_1 = (y-1)^2$ (final position error squared)
- $J_2 = \|x\|^2$ (sum of squares of forces)

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

optimal x:

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

optimal trade-off curve



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