Today's Lecture

- Multi-objective least-squares
- Regularized least-squares
- Nonlinear least-squares and Gauss-Newton method
- Minimum-norm solution of underdetermined equations

Multi-objective Least-squares

- We discussed minimizing the error norm using least-squares.

  In many problems, we have other goals; we have two (or more) objectives.

  We want to find $x \in R^n$ such that
  - $J_1 = \|Ax - y\|^2$ is small and
  - $J_2 = \|Fx - g\|^2$ is also small.

- Usually the objectives are competing.

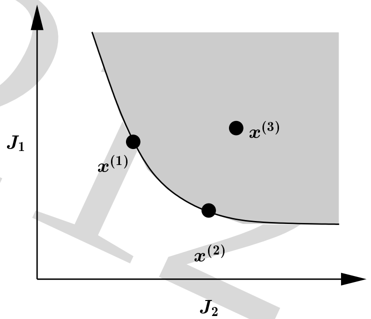
  We can make one smaller at the expense of making the other larger.

Multi-objective Least-squares

- Example. Left $F = I$ and $g = 0$.

  We want $\|Ax - y\|$ small and at the same time small x.

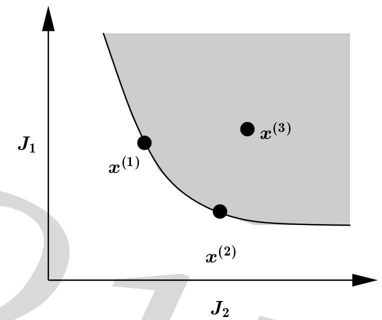
- Plot (J_2, J_1) for every x.
  - shaded area shows (J_2, J_1) achieved by some $x \in R^n$.
  - clear area shows (J_2, J_1) not achieved by any $x \in R^n$.



Multi-objective Least-squares

- Boundary of the region is called optimal trade-off curve.

  Points x along the boundary are called Pareto optimal (for the two objective functions J_1 and J_2).

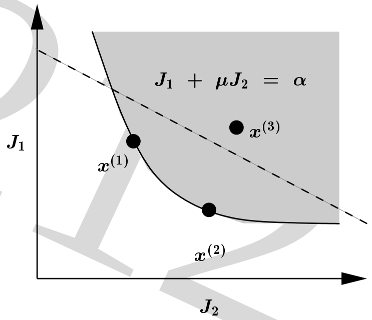


- Consider the choices of x : $x^{(1)}$, $x^{(2)}$, $x^{(3)}$.
  - $x^{(3)}$ is worse than $x^{(2)}$ based on both J_1 and J_2.
  - $x^{(1)}$ is better than $x^{(2)}$ in J_2 but worse in J_1.

- **To find Pareto optimal points (x on the optimal trade-off curve), minimize the weighted-sum objective**

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

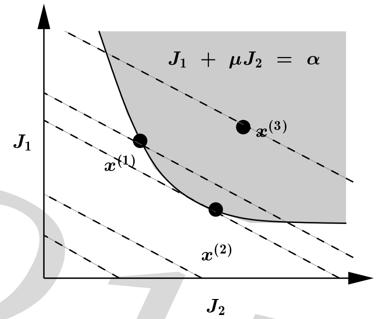
- **Parameter $\mu \geq 0$ gives the relative weight (importance) between J_1 and J_2.**



- **Points where the weighted sum is constant, i.e.,**

$$J_1 + \mu J_2 = \alpha$$

correspond to the line with slope $-\mu$.



- **Point $x^{(2)}$ minimizes the weighted-sum for some μ and gives us a point on the optimal trade-off curve.**

To find other points on the curve, vary μ from 0 to $+\infty$ and minimize weighted-sum.

- **Express the weighted-sum objective as an ordinary least-squares objective.**

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

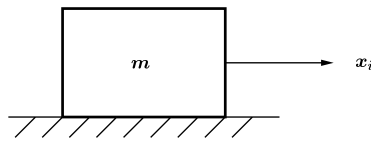
$$= \left\| \underbrace{\begin{bmatrix} A \\ \sqrt{\mu}F \end{bmatrix}}_{\tilde{A}} x + \underbrace{\begin{bmatrix} y \\ \sqrt{\mu}g \end{bmatrix}}_{\tilde{y}} \right\|^2 = \|\tilde{A}x - \tilde{y}\|^2$$

- **Thus, assuming $\tilde{A}$ is full rank,**

$$x = (\tilde{A}^T\tilde{A})^{-1}\tilde{A}^T\tilde{y}$$

$$= (A^TA + \mu F^TF)^{-1}(A^Ty + \mu F^Tg)$$

- **Example. Minimizing position error and effort.**



- **mass is initially at rest.**
- **mass is subjected to a piecewise-constant force profile**

$$x_i \qquad i - 1 < t \leq i,\ i = 1, \ldots, 10$$

- **$y \in R$ is the position at $t = 10$. With $A \in R^{1\times 10}$,**

$$y = Ax \qquad \text{where } A_i = \frac{1}{m}\cdot\frac{1}{2}[1 + 2(10 - i)]$$

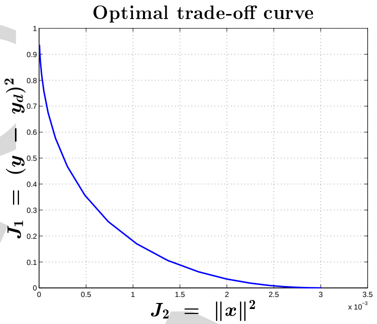
- **$J_1 = (y - y_d)^2$ (square of final position error).**
- **$J_2 = \|x\|^2$ (sum of the square of forces).**

Minimizing Weighted-sum Objective

- Weighted-sum objective : $(Ax - 1) + \mu\|x\|^2$

  Optimal x (for a certain μ) : $x = (A^TA + \mu I)^{-1}A^Ty_d$

- Let $y_d = 1$.
  - left portion of the curve corresponds to $x \rightarrow 0$ (small effort).
  - right portion corresponds to $y \rightarrow y_d$ (small position error).



Regularized Least-squares

- Consider the case when $F = I$ and $g = 0$. The objectives are

$$J_1 = \|Ax - y\|^2 \text{ and } J_2 = \|x\|^2$$

- Optimal x : $x = (A^TA + \mu I)^{-1}(A^Ty)$.

  The solution x is the regularized least-squares solution of $Ax \approx y$.
  - also termed as the Tychonov regularization.
  - for $\mu > 0$, works for any A (no shape or rank restriction).

Regularized Least-squares

- Applications to estimation (inversion).
  - $Ax - y$ is sensor residual.
  - prior information : x is small.
  - model only accurate for x small.
  - regularized solution trades off between sensor fit and size of x.

Nonlinear Least-squares

- Most physical systems (models) are nonlinear. We linearize first to get our model into the $Ax = y$ form.

- Nonlinear least-squares (NLLS) problem. Find $x \in R^n$ that minimizes

$$\|r(x)\|^2 = \sum_{i=1}^{m}[r_i(x)]^2$$

  where $r : R^n \rightarrow R^m$.
  - $r(x)$ is a vector of residuals.
  - reduces to (linear) least-squares if $r(x) = Ax - b$.

Nonlinear Least-squares

- **Example.** Estimate the position $x \in R^2$ from the range measurments to the beacons at locations $b_1, \ldots, b_m \in R^2$ without linearizing.
  - we measure $\rho_i = \|x - b_i\| + v_i$ (v_i is the unknown sensor error, assumed small).
  - NLLS estimate : choose $\hat{x}$ to minimize

$$\sum_{i=1}^{m} [r_i(x)]^2 = \sum_{i=1}^{m} [\rho_i - \|x - b_i\|]^2$$

- Several ways to do this.

  One way is using the Gauss-Newton method.

Gauss-Newton Method for NLLS

- NLLS : Find $x \in R^n$ that minimizes

$$\|r(x)\|^2 = \sum_{i=1}^{m} [r_i(x)]^2$$

  where $r : R^n \rightarrow R^m$.

- In general, very hard to get the exact solution.

- Many algorithms are available to compute optimal solution (at least locally).

Gauss-Newton Method for NLLS

- Gauss-Newton method.

  given a starting guess for x.
  repeat
  - linearize r near current guess.
  - new guess is linear least-squares solution using the linearized r.

  until convergence

- Other algorithms use different ways of guessing the next iterate.

In Detail, Gauss-Newton Method for NLLS

- Linearize r near current iterate $x^{(k)}$.

$$r(x) \approx r(x^{(k)}) + Dr(x^{(k)})(x - x^{(k)})$$

  where Dr is the Jacobian of $r(x)$ given by $(Dr)_{ij} = \partial r_i / \partial x_j$.

- Rewrite the linear approximation as

$$r(x^{(k)}) + Dr(x^{(k)})(x - x^{(k)}) = A^{(k)}x - b^{(k)}$$

  where

$$A^{(k)} = Dr(x^{(k)})$$
$$b^{(k)} = Dr(x^{(k)})x^{(k)} - r(x^{(k)})$$

In Detail, Gauss-Newton Method for NLLS

- At the kth iteration, we approximate the NLLS problem by the linear LS problem.

$$\|r(x)\|^2 \approx \left\|A^{(k)}x - b^{(k)}\right\|^2$$

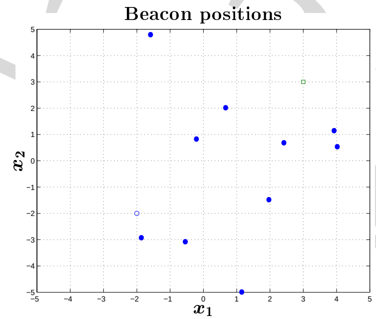
- The solution to the linearized LS problem updates the iteration.

$$x^{(k+1)} = \left[{A^{(k)}}^T A^{(k)}\right]^{-1} {A^{(k)}}^T b^{(k)}$$

- Note that there are other ways of getting $x^{(k+1)}$.

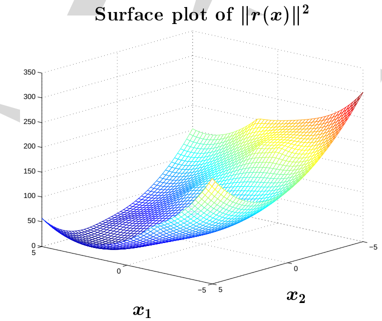
In Detail, Gauss-Newton Method for NLLS

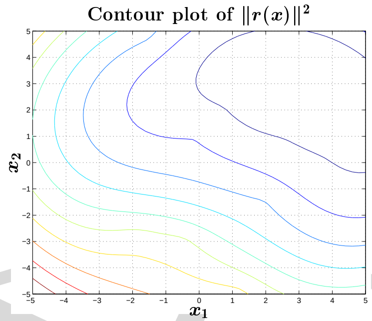
- Example. Navigation problem using 10 beacons.
  - actual position : $(3, 3)$
  - initial guess : $(-2, -2)$
  - range measurement accuracy : ± 0.5



In Detail, Gauss-Newton Method for NLLS

- The objective function $\|r(x)\|^2$ looks like

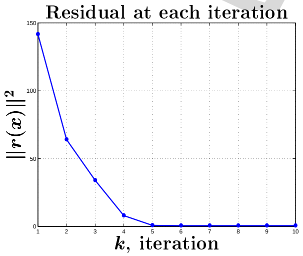


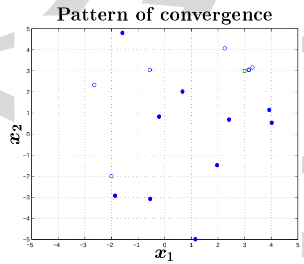


  - for a linear least-squares problem, the objective would be a quadratic bowl.
  - "bumps" are due to the nonlinearity of r.

In Detail, Gauss-Newton Method for NLLS

- Using Gauss-Newton method, we find
  - $x^{(k)}$ converges to a minimum (in this case, globally).
  - convergence takes a few steps (less than 10).
  - final estimate is $\hat{x} = (3.158, 2.0424)$.
  - estimation error $\|\hat{x} - x\| = 0.1636$. (better than the range accuracy).





Underdetermined Linear Equations

- Consider

$$y = Ax$$

where $A \in R^{m \times n}$ is fat ($m < n$).

  - there are more variables than equations.
  - x is underspecified, i.e., many possible x's give the same y.

- Assume that A is full rank (rank $= m$). Each $y \in R^m$ has a corresponding solution.

Underdetermined Linear Equations

- The set of all solutions has the form

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

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

- Remarks.
  - z characterizes the available solutions.
  - solution has $\dim \mathcal{N}(A) = n - m$ degrees of freedom.
  - can choose z to satisfy secondary specification or optimize based on other objective(s).

Minimum-norm (Least-norm) Solution

- A particular solution is

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

which is a solution to $y = Ax$ that minimizes $\|x\|$. AA^T is nonsingular since A is full rank.

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

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

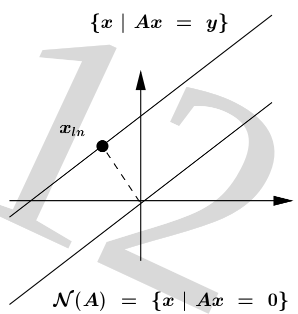
Minimum-norm (Least-norm) Solution

- This implies that $(x - x_{ln}) \perp x_{ln}$. Thus,

$$\begin{aligned}\|x\|^2 = \|x - x_{ln} + x_{ln}\|^2 &= \|x_{ln}\|^2 + \|x - x_{ln}\|^2 \\ &\geq \|x_{ln}\|^2\end{aligned}$$

i.e., x_{ln} has the smallest norm among the possible solutions.

  - orthogonality condition. $x_{ln} \perp \mathcal{N}(A)$.
  - projection interpretation. x_{ln} is the projection of 0 on the solution set $\{x \mid Ax = y\}$.



Minimum-norm (Least-norm) Solution

- $A^T(AA^T)^{-1}$ is called the pseudoinverse of A (for a full rank and fat A).
  $A^T(AA^T)^{-1}$ is a right inverse of A.

- Least-norm solution using QR factorization.
  Decompose A^T into $A^T = QR$. Thus,

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

  where $R^{-T} = (R^{-1})^T$ and $\|x_{ln}\| = \|R^{-T}y\|$.

Least-norm Solution Using Lagrange Multipliers

- The minimum-norm problem can be cast as an optimization problem.

$$\begin{aligned}&\text{minimize } x^Tx\\ &\text{subject to } Ax = y\end{aligned}$$

- Solve using Lagrange multipliers. Define the Lagrangian function

$$L(x, \lambda) = x^Tx + \lambda^T(Ax - y)$$

  We want to minimize the Langrangian function with respect to x and λ.

Least-norm Solution Using Lagrange Multipliers

- Optimality conditions are

$$\frac{\partial L}{\partial x} = 2x^T + \lambda^T A = 0, \qquad \frac{\partial L}{\partial \lambda} = (Ax - y)^T = 0$$

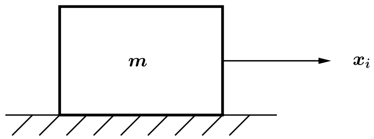
- First condition gives $x = -A^T\lambda/2$.
  Substituting into the second condition gives

$$\lambda = -2(AA^T)^{-1}y$$

  Thus, $x = A^T(AA^T)^{-1}y$
  (same as the previous least-norm solution).

Least-norm Solution Using Lagrange Multipliers

- Example. Moving a mass.



  – mass is initially at rest.
  – mass is subjected to a piecewise-constant force profile

$$x_i \qquad i - 1 < t \leq i,\ i = 1, \ldots, 10$$

  – $y_1 \in R$ is the position at $t = 10$.
  $y_2 \in R$ is the final velocity at $t = 10$.

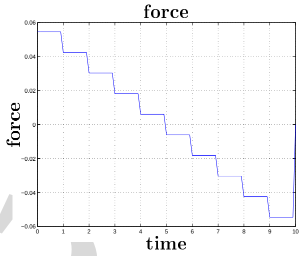
$$y = Ax \qquad\qquad \text{where } A \in R^{2\times 10}$$

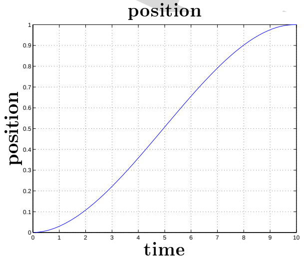
  – find the minimum norm force that moves the mass a unit distance with zero final velocity, i.e., $y_d = (1, 0)$.

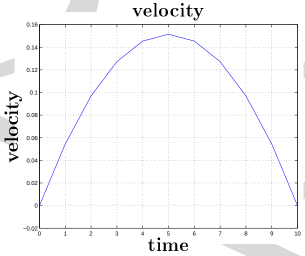
Least-norm Solution Using Lagrange Multipliers

- Solution is

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







Relation to Regularized Least-squares

- Suppose $A \in R^{m\times n}$ is fat and full rank.

  Define

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

  Least-norm solution minimizes J_2 with $J_1 = 0$.

- Optimal solution to weighted-sum objective problem

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

  is

$$x_\mu = (A^TA + \mu I)^{-1}A^Ty$$

Relation to Regularized Least-squares

- Fact. $x_\mu \rightarrow x_{ln}$ as $\mu \rightarrow 0$,
  i.e., regularized solution converges to the least-norm solution as $\mu \rightarrow 0$.

  In matrix form, for a full rank and fat A, as $\mu \rightarrow 0$,

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

Summary

- Multi-objective least-squares
- Regularized least-squares
- Nonlinear least-squares and Gauss-Newton method
- Minimum-norm solution of underdetermined equations
- Relation to regularized least-squares