Lecture 7
Regularized least-squares and Gauss-Newton method

- multi-objective least-squares
- regularized least-squares
- nonlinear least-squares
- Gauss-Newton method

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

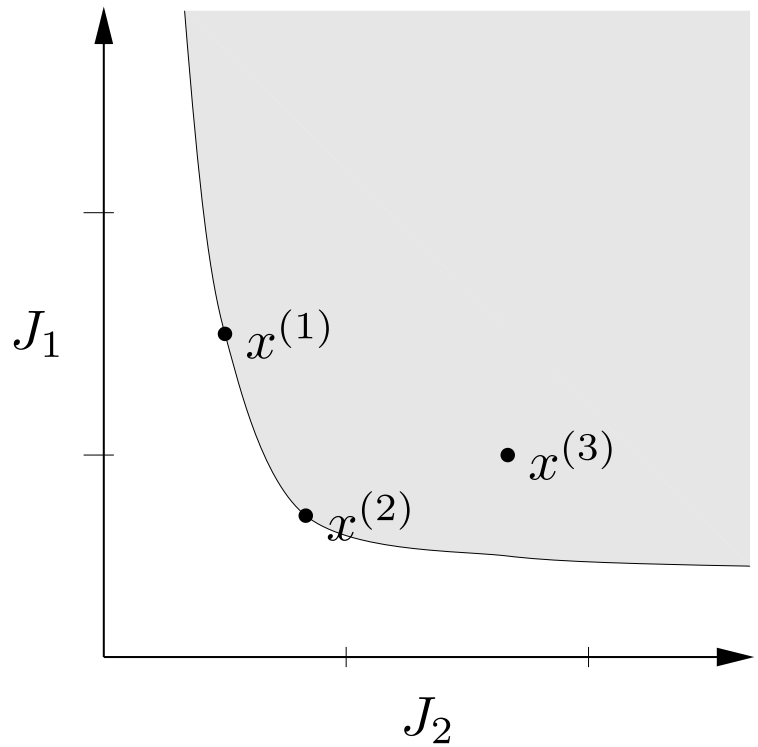
($x \in \mathbf{R}^n$ is the variable)

- 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

plot (J_2, J_1) for every x:



note that $x \in \mathbf{R}^n$, but this plot is in $\mathbf{R}^2$; point labeled $x^{(1)}$ is really $\left(J_2(x^{(1)}), J_1(x^{(1)})\right)$

- shaded area shows (J_2, J_1) achieved by some $x \in \mathbf{R}^n$
- clear area shows (J_2, J_1) not achieved by any $x \in \mathbf{R}^n$
- 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

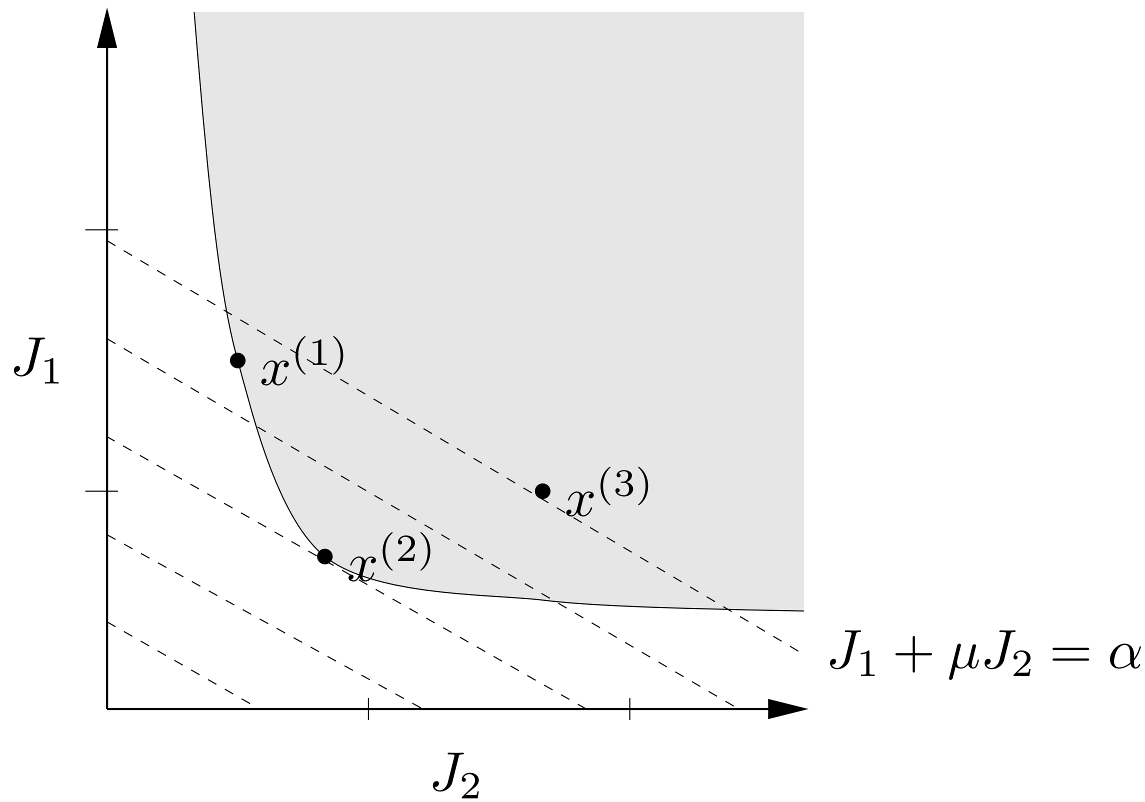
Weighted-sum objective

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

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

- 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*

Minimizing weighted-sum objective

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

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

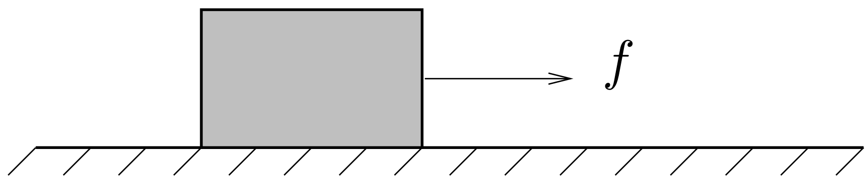
where

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

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

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

Example



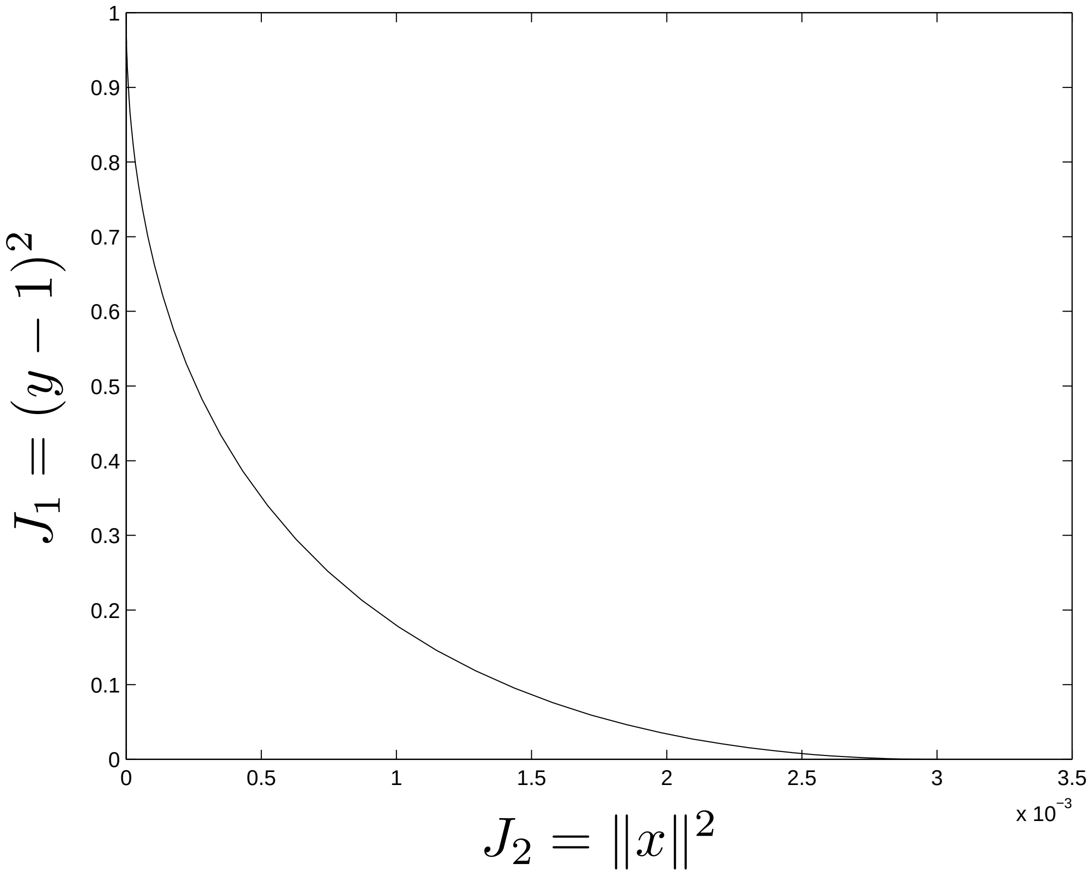
- unit mass at rest subject to forces x_i for $i-1 < t \leq i$, $i = 1, \ldots, 10$
- $y \in \mathbf{R}$ is position at $t = 10$; $y = a^T x$ where $a \in \mathbf{R}^{10}$
- $J_1 = (y-1)^2$ (final position error squared)
- $J_2 = \|x\|^2$ (sum of squares of forces)

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

optimal x:

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

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.*, $J_1 = 0$

Regularized least-squares

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

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

minimizer of weighted-sum objective,

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

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
- or, model only accurate for x small
- regularized solution trades off sensor fit, size of x

Nonlinear least-squares

nonlinear least-squares (NLLS) problem: find $x \in \mathbf{R}^n$ that minimizes

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

where $r : \mathbf{R}^n \to \mathbf{R}^m$

- $r(x)$ is a vector of 'residuals'
- reduces to (linear) least-squares if $r(x) = Ax - y$

Position estimation from ranges

estimate position $x \in \mathbf{R}^2$ from approximate distances to beacons at locations $b_1, \ldots, b_m \in \mathbf{R}^2$ *without* linearizing

- we measure $\rho_i = \|x - b_i\| + v_i$
  (v_i is range error, unknown but assumed small)

- NLLS estimate: choose $\hat{x}$ to minimize

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

Gauss-Newton method for NLLS

NLLS: find $x \in \mathbf{R}^n$ that minimizes $\|r(x)\|^2 = \sum_{i=1}^{m} r_i(x)^2$, where $r : \mathbf{R}^n \to \mathbf{R}^m$

- in general, very hard to solve exactly
- many good heuristics to compute *locally optimal* solution

Gauss-Newton method:

given starting guess for x
repeat
    linearize r near current guess
    new guess is linear LS solution, using linearized r
until convergence

Gauss-Newton method (more detail):

- linearize r near current iterate $x^{(k)}$:

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

  where Dr is the Jacobian: $(Dr)_{ij} = \partial r_i / \partial x_j$

- write linearized approximation as

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

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

- at kth iteration, we approximate NLLS problem by linear LS problem:

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

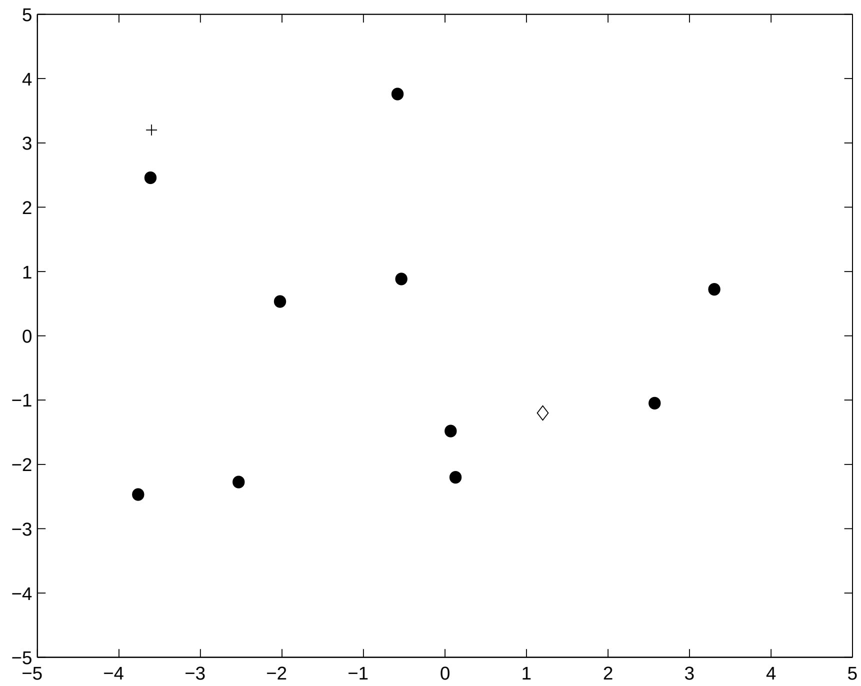
- next iterate solves this linearized LS problem:

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

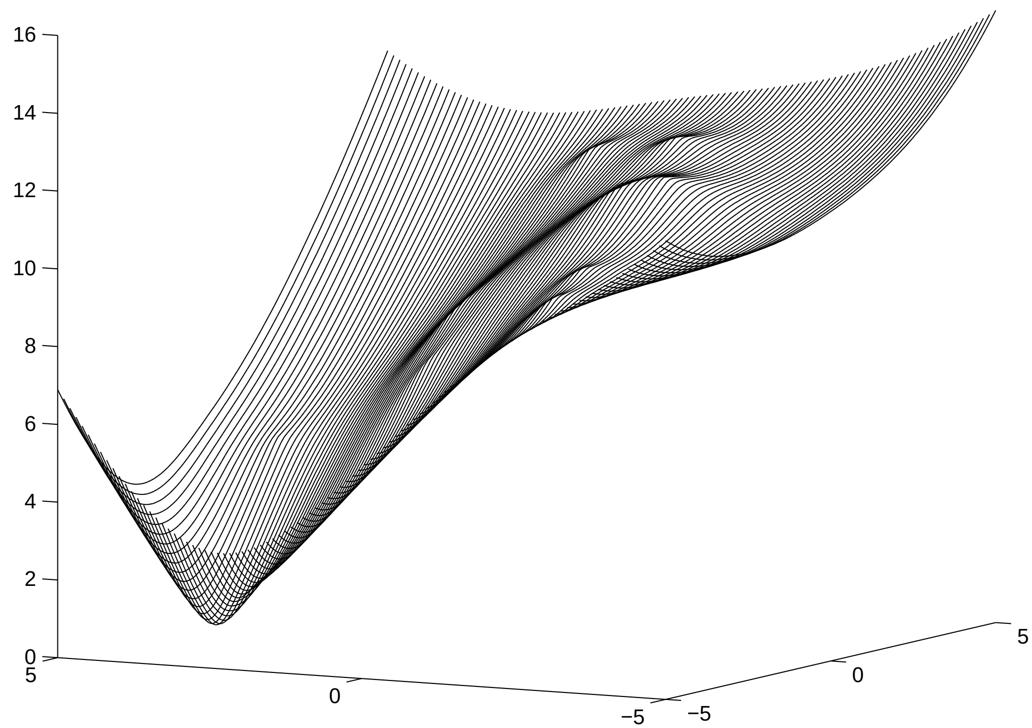
- repeat until convergence (which *isn't* guaranteed)

Gauss-Newton example

- 10 beacons
- $+$ true position $(-3.6, 3.2)$; $\diamondsuit$ initial guess $(1.2, -1.2)$
- range estimates accurate to ± 0.5

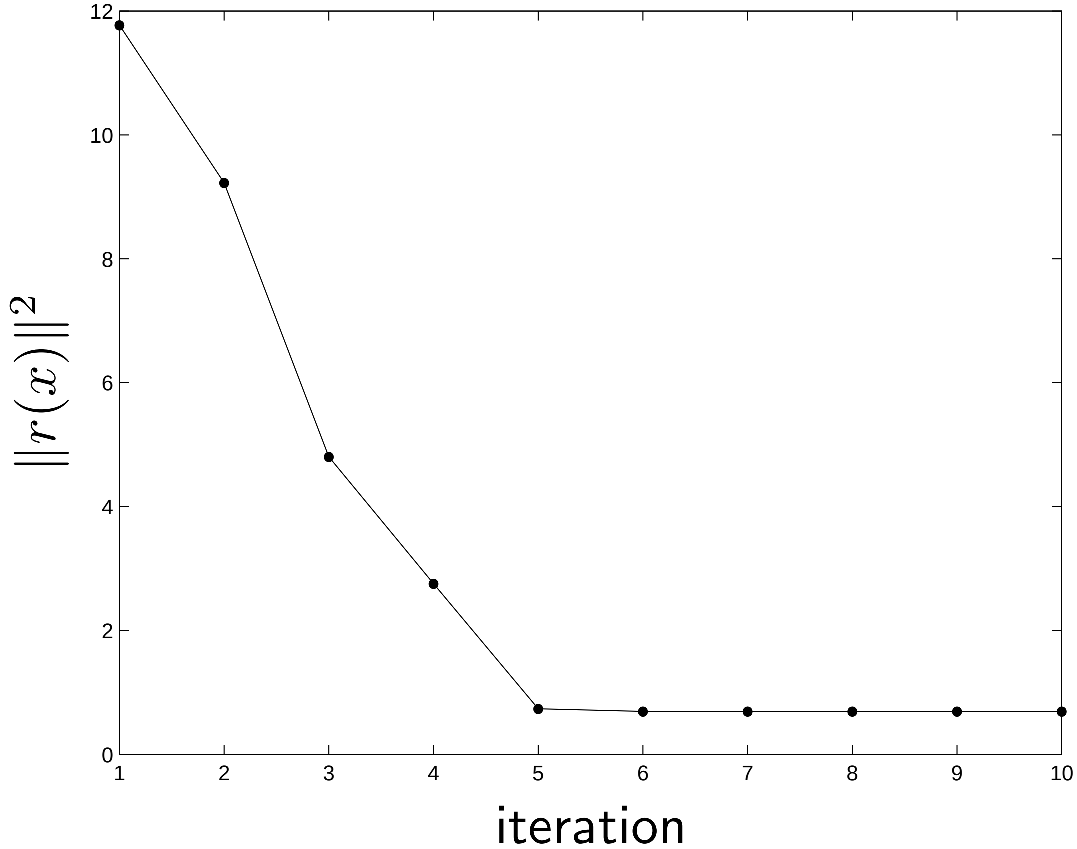


NLLS objective $\|r(x)\|^2$ versus x:



- for a linear LS problem, objective would be nice quadratic 'bowl'
- bumps in objective due to strong nonlinearity of r

objective of Gauss-Newton iterates:

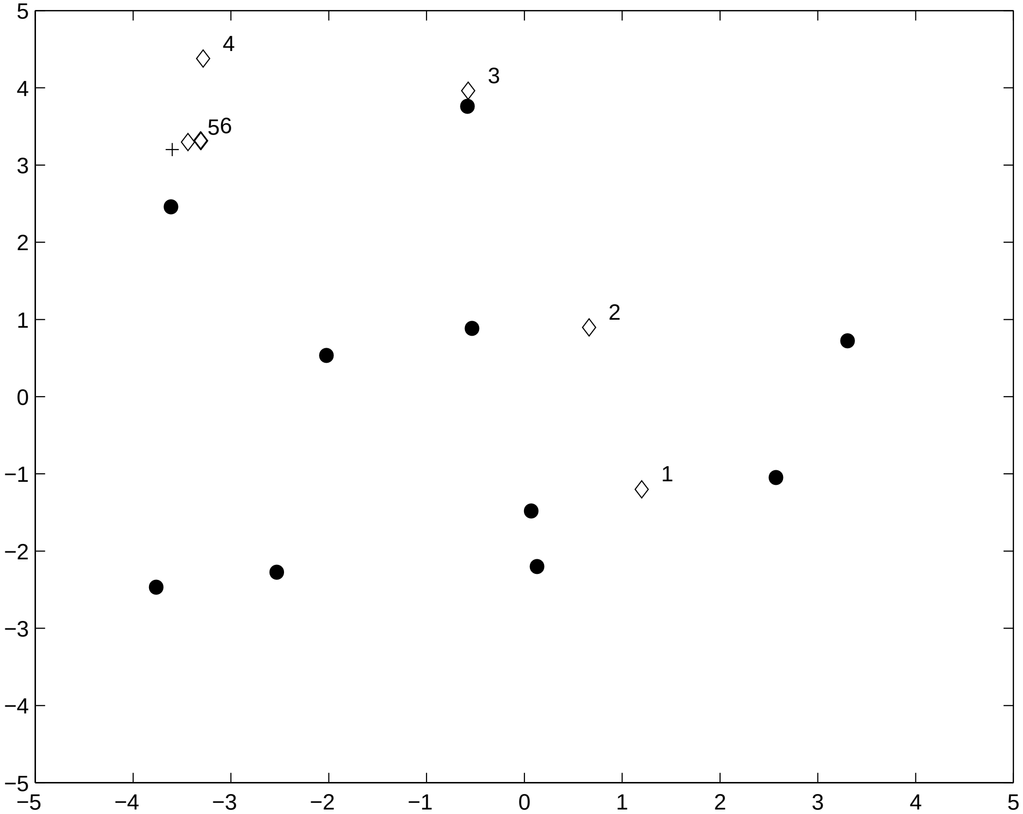


- $x^{(k)}$ converges to (in this case, global) minimum of $\|r(x)\|^2$
- convergence takes only five or so steps

- final estimate is $\hat{x} = (-3.3, 3.3)$

- estimation error is $\|\hat{x} - x\| = 0.31$
  (substantially smaller than range accuracy!)

convergence of Gauss-Newton iterates:



useful varation on Gauss-Newton: add regularization term

$$\|A^{(k)}x - b^{(k)}\|^2 + \mu\|x - x^{(k)}\|^2$$

so that next iterate is not too far from previous one (hence, linearized model still pretty accurate)