Least-Squares Applications

Gabriel Hugh Elkaim

Winter 2016

Least-Squares Applications

- Least-squares data fitting
- Growing sets of regressors
- System identification
- Growing sets of measurements and recursive least-squares

Growing Sets of Measurements

$$\min_x \|Ax - y\|^2 = \sum_{i=1}^{m} (a_i^T x - y_i)^2$$

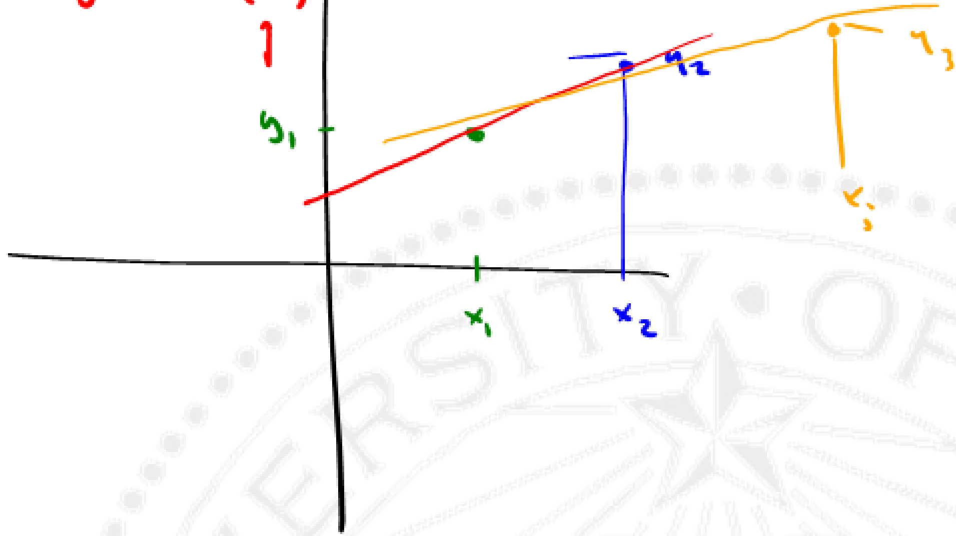
a_i^T are the rows of A. $(a_i \in \mathbb{R}^n)$

$x \in \mathbb{R}^n$ is vector to be estimated

each pair a_i, y_i corresponds to one additional measurement

Solution: $$x_{ls} = \underbrace{\left(\sum_{i=1}^{m} a_i a_i^T\right)}^{-1} \sum_{i=1}^{m} y_i a_i$$

$y = mx + b \qquad y = \begin{bmatrix} x & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix}$



Recursive Least-Squares

Compute $x_{ls}(m) = \left(\sum_{i=1}^{m} a_i a_i^T \right)^{-1} \sum_{i=1}^{m} y_i a_i$

$P(0) = 0 \in \mathbb{R}^{n \times n}$ $\quad q(0) = 0 \in \mathbb{R}^n$

for $m = 0, 1, \ldots$ $\quad P_{(m)}^{-1} \sim O(n^3)$

$$P(m+1) = P(m) + a_{m+1} a_{m+1}^T$$

$$q(m+1) = q(m) + y_{m+1} a_{m+1}^T$$

if $P(m)^{-1}$ exists $\longrightarrow$ $x_{ls}(m) = P^{-1}(m)\, q(m)$

$\{a_1 \ldots a_m\}$ span $\mathbb{R}^n$

Fast Update for *RLS* $\sim O(n^2)$

$$P_{(m+1)}^{-1} = \left(P_{(m)} + a_{m+1} a_{m+1}^T\right)^{-1}$$

$$(P + aa^T)^{-1} = P^{-1} - \frac{1}{1 + a^T P^{-1} a} (P^{-1} a)(P^{-1} a)^T$$

$P = P^T$ P and $P + aa^T$ are invertible

Verification of Rank One Update

$$(P + aa^T)\left(P^{-1} - \frac{1}{1+a^TP^{-1}a}(P^{-1}a)(P^{-1}a)^T\right)$$

$$I + aa^TP^{-1} - \frac{1}{1+a^TP^{-1}a}P(P^{-1}a)(P^{-1}a)^T - \frac{aa^T(P^{-1}a)(P^{-1}a)^T}{1+a^TP^{-1}a}$$

$$I + aa^TP^{-1} - \frac{aa^TP^{-1}}{1+a^TP^{-1}a} - \frac{a^TP^{-1}a}{1+a^TP^{-1}a}\cdot aa^TP^{-1}$$

$$I + aa^TP^{-1} - \left[\underbrace{\cancel{\frac{1+a^TP^{-1}a}{1+a^TP^{-1}a}}}_{1}\right]aa^TP^{-1}$$

$$I$$

Questions?

11

Rank one down date

Regularized Least-Squares and Gauss-Newton Methods

Gabriel Hugh Elkaim

Winter 2016

Regularized Least-Squares and Gauss-Newton Methods

- Multi-objective Least-Squares
- Regularized Least-Squares
- Non-linear Least Squares
- Gauss-Newton Method

$\boxed{\text{TANSTAAFL}}$ ←

Multi-objective Least-Squares (1.3)

$J_1 = \|Ax - y\|^2$ small

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

$x \in \mathbb{R}^n$ ← same x to min. both.

→ objectives are competing

→ we can only make J_1 small at the cost of J_2 ↑

$F = I \quad g = 0$

min $\|Ax - y\|$

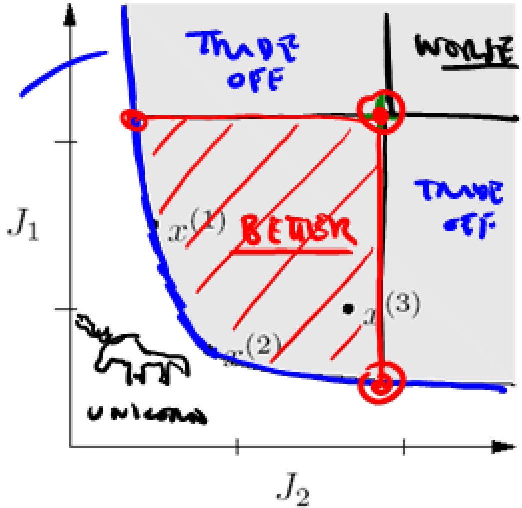
subj min $\|x\|$.

Multi-objective Least-Squares (2.3)

Plot $J_1(x)$ and $J_2(x)$

efficient boundary

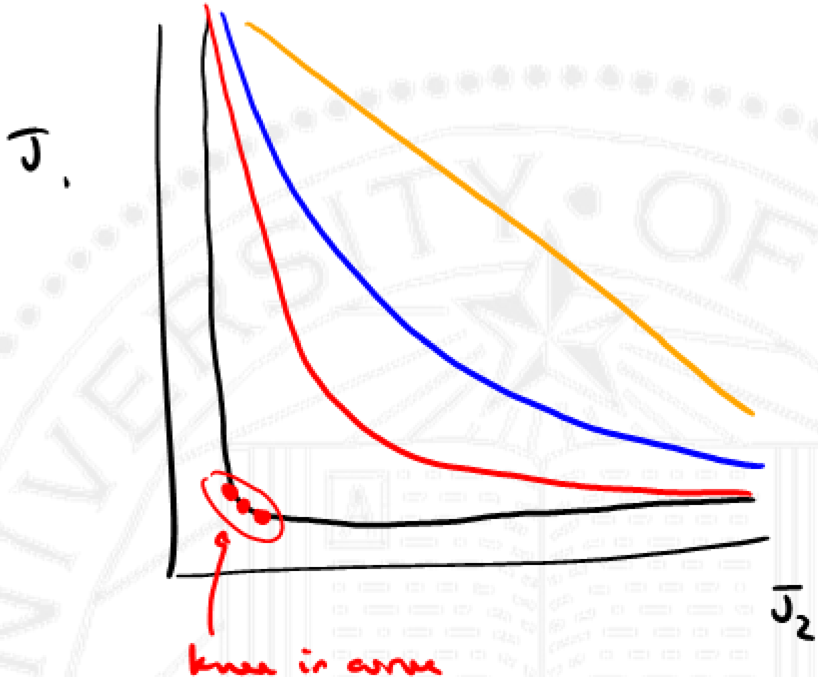
pareto optimal boundary



Multi-objective Least-Squares (3.3)

$J_1 \quad \|Ax - y\|^2$

$J_2 \quad \|Fx - g\|^2$



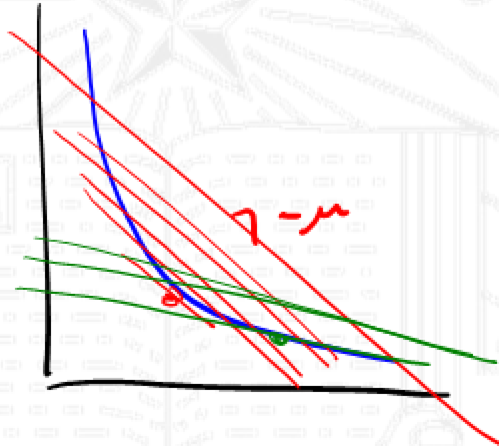
Weighted-sum Objective (1.3)

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

$\mu \geq 0$ gives the relative weight between J_1, J_2

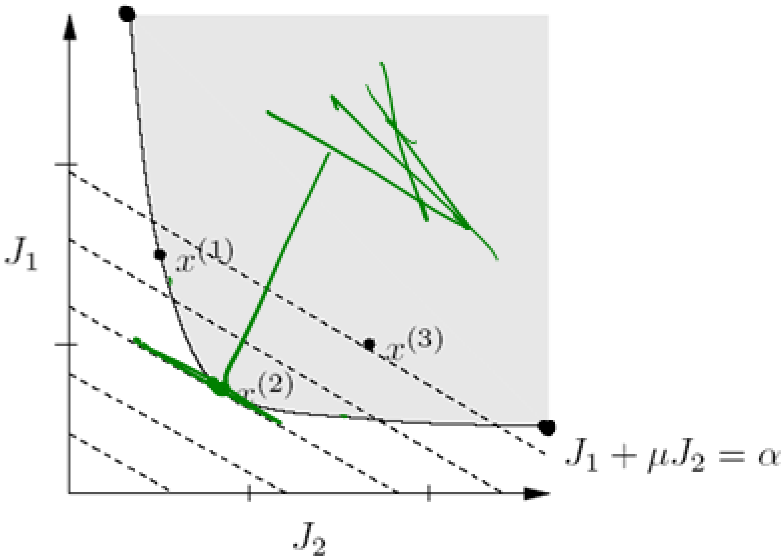
local exchange rate

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



Weighted-sum Objective (2.3)

Vary $\mu \to 0 : \infty$



Minimizing Weighted-Sum Objective

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

$$\left\| \begin{bmatrix} a \\ b \end{bmatrix} \right\|^2 = \|a\|^2 + \|b\|^2$$

$$\min_x \|\tilde{A}x - \tilde{y}\|^2 \longrightarrow x_{ls} = (\tilde{A}^T\tilde{A})^{-1}\tilde{A}^T\tilde{y}$$

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

$$\mu = 0 \longrightarrow x_{ls} = (A^TA)^{-1}A^Ty \quad x_{ls} \to A, y$$

$$\mu \to \infty \longrightarrow x_{ls} = (F^TF)^{-1}F^Tg \quad x_{ls} \to F, g$$

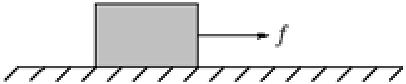
Minimizing Weighted-Sum Objective

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

$$\begin{bmatrix} 1 & -1 & & \\ & \ddots & \ddots & \\ & & \ddots & \ddots \\ & & & \end{bmatrix}$$

smoothen

Example: Mass-Force Program



$i = 1 \ldots 10$

$y \in \mathbb{R}$ posit'n @ $t=10$ $\quad y = a^T x \quad a \in \mathbb{R}^{10}$

$J_1 = (y-1)^2$ final posit'n error squared

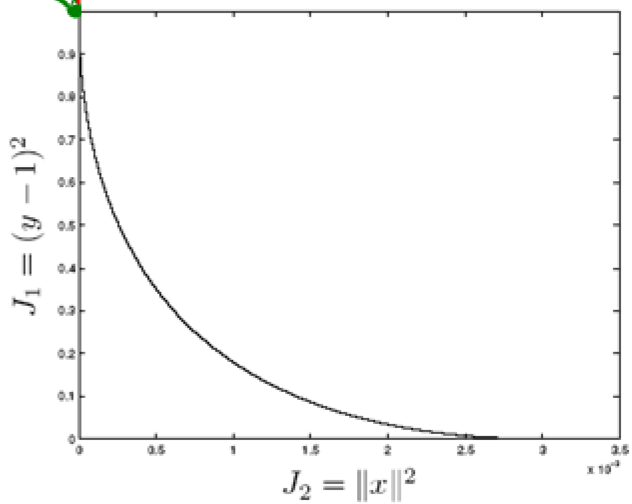
$J_2 = \|x\|^2$ sum square of force.

$$(a^T x - 1)^2 + \mu x^T x$$

$$\boxed{x_{opt} = (a a^T + \mu I)^{-1} a \cdot 1}$$

Example: Mass-Force Program

DO NOTHING



Regularized Least-Squares (1.3)

when $F = I$ $\quad g = 0$ "ridge regression"

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

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

regularized least squares soln $Ax \approx y$.

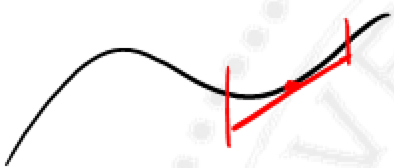
Tychonov Regularization

$\mu > 0$ works with any A

Regularized Least-Squares (2.3)

$Ax - y$ sensor residual

prior information x is small.



linearization, error only valid for small δx

image processing → Laplacian Regularization

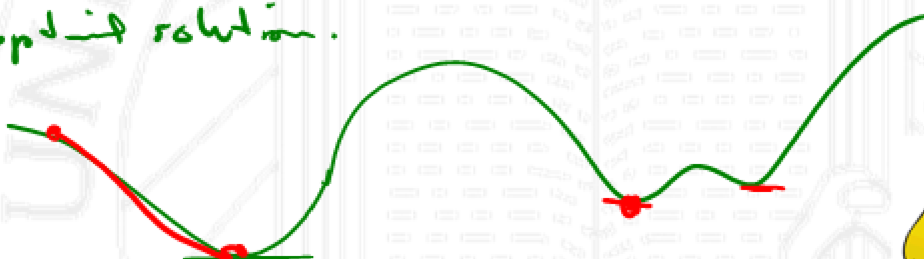
$$\|Ax - y\|^2 + \mu \left\| \begin{matrix} D_v x \\ D_v y \end{matrix} \right\|^2$$

Non-Linear Least Squares (1.3)

find $x \in \mathbb{R}^n$ $\quad \min_x \|r(x)\|^2 = \sum_{i=1}^{m} r_i(x)^2$

$r : \mathbb{R}^n \to \mathbb{R}^m$

- In general, very hard to solve exactly.
- Reduces to (linear) LS if $r(x) = Ax - y$.
- Lots of good heuristics compute a locally optimal solution.



Non-Linear Least Squares (2.3)

① Guess x or x_0 initial condition

② $\boxed{\text{Linearize } r \text{ near } x}$

③ Solve lin LS based on ②.

④ Update x

calculus
Particle filters
Perturbation

$$r(x) \simeq Ax - b \text{ near } x_0 .$$

Non-Linear Least Squares (3.3)

linearize r near around $x^{(k)}$

$$r(x) \simeq r(x^{(k)}) + Dr(x^{(k)})\underbrace{(x - x^{(k)})}_{\Delta x}$$

$$Dr \triangleq \text{jacobian} = Dr_{ij} = \frac{\partial r_i}{\partial x_j} \quad m \times n$$

$$r(x) \simeq r(x^{(k)}) + Dr(x^{(k)})\,\Delta x \triangleq A^{(k)}\,\Delta x - b^{(k)}$$

$$\boxed{\begin{aligned} A^{(k)} &= Dr(x^{(k)}) \\ b^{(k)} &= Dr(x^{(k)})\,x^{(k)} - r(x^{(k)}) \end{aligned}}$$

NLLS Example: Position from Ranges

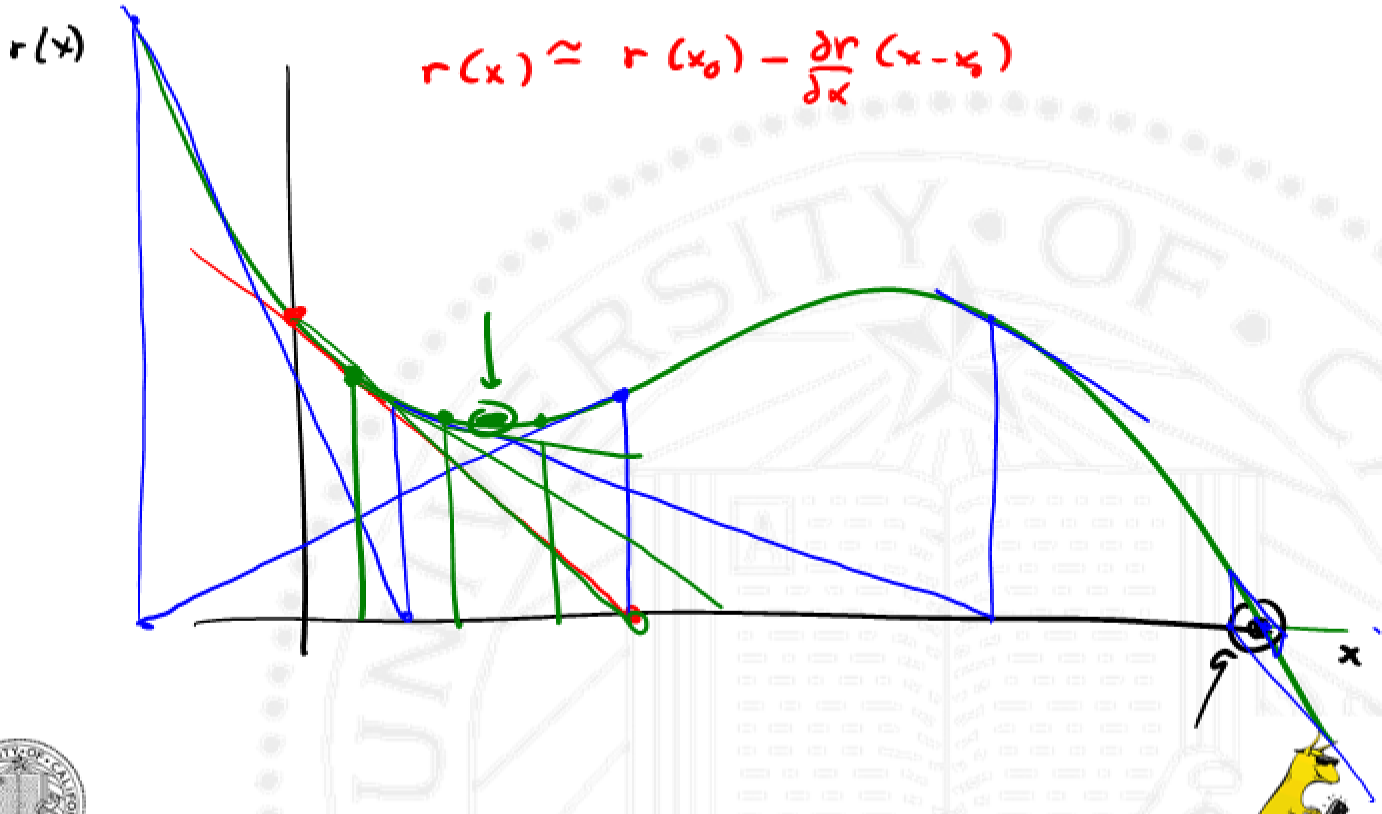
$$\|r(x)\|^2 \simeq \|A^{(k)}\delta x - b^{(k)}\|^2$$

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

$$\rightarrow \|A^{(k)}x^{(k)} - b^{(k)}\|^2 + \overbrace{\mu\|x - x^{(k)}\|^2}^{\text{TRUST REGION}}$$

$$F = I \quad g = \phi.$$

Gauss-Newton method for NLLS (1.3)



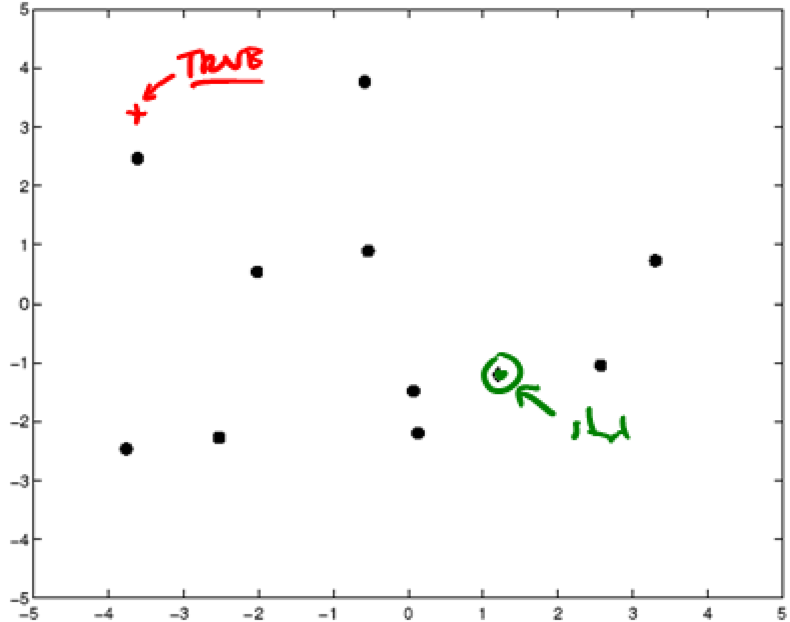
Gauss-Newton: Example (1.4)

10 beacons @ known locations

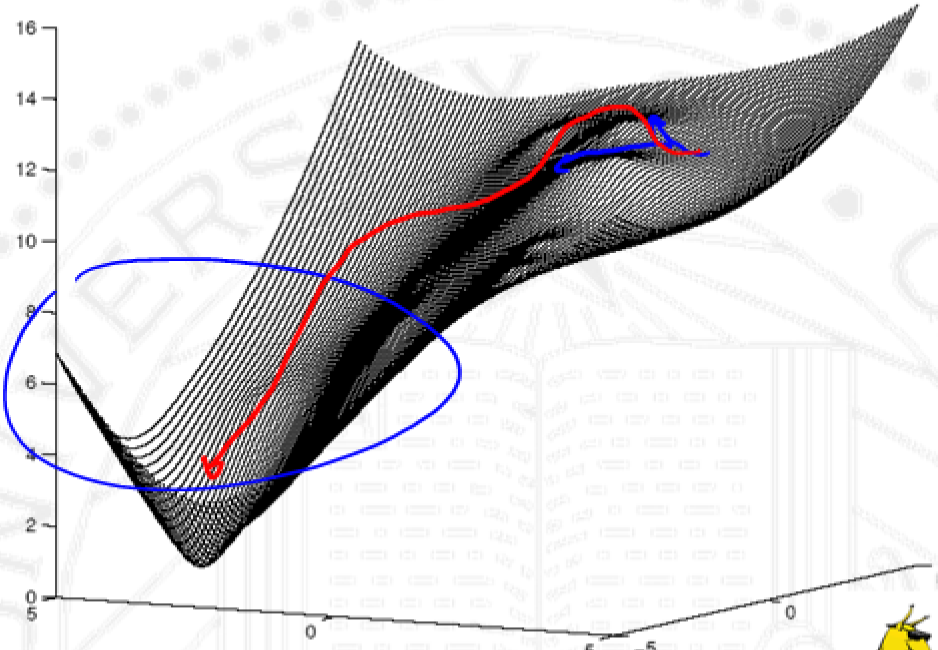
(−3.6, 3.2)

(+1.2, −1.2) green

ranges are ± 0.5 m



Gauss-Newton: Example (2.4)

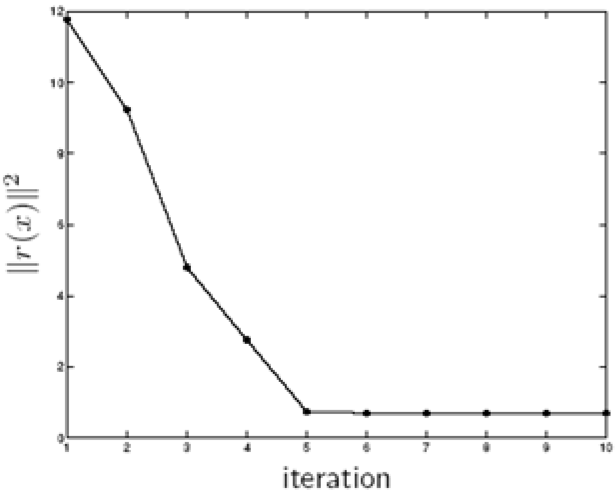


Gauss-Newton: Example (3.4)

converge to global min

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

$\|x - \hat{x}\|^2 = 0.31$



Gauss-Newton: Example (4.4)

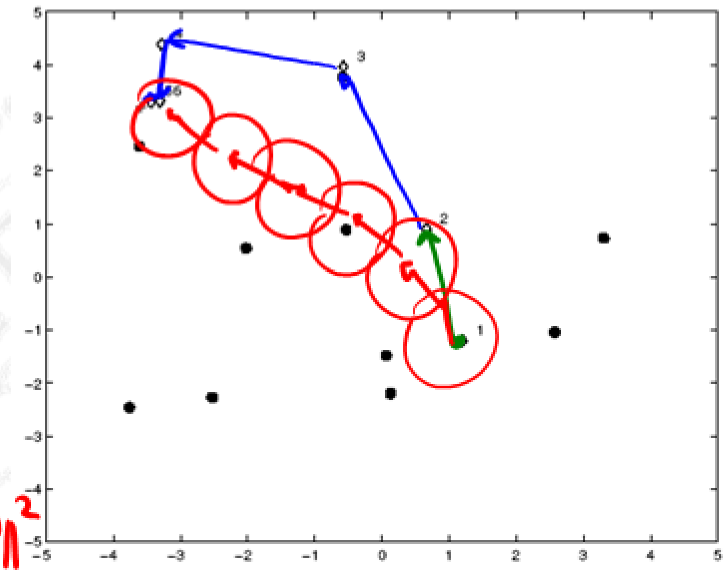
Use trust region

$\underline{\Delta x}$

$x := x + \underset{J}{\Delta x}$

$x := x + \frac{\Delta x}{10}$

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



Questions?