Lecture 7

Regularized least square and Gauss Newton method

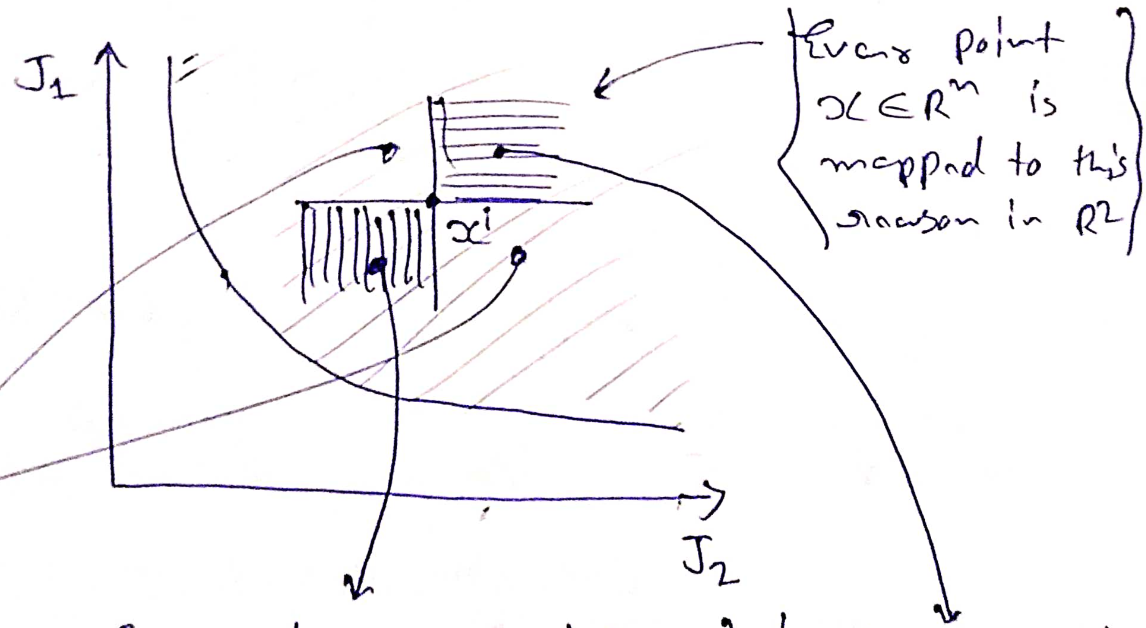
★ Multi-objective least-square

⇒ In many problem we have two (or more) objectives:

→ We want $J_1 = \|Ax - y\|^2$ small

→ and also $J_2 = \|Fx - g\|^2$ small

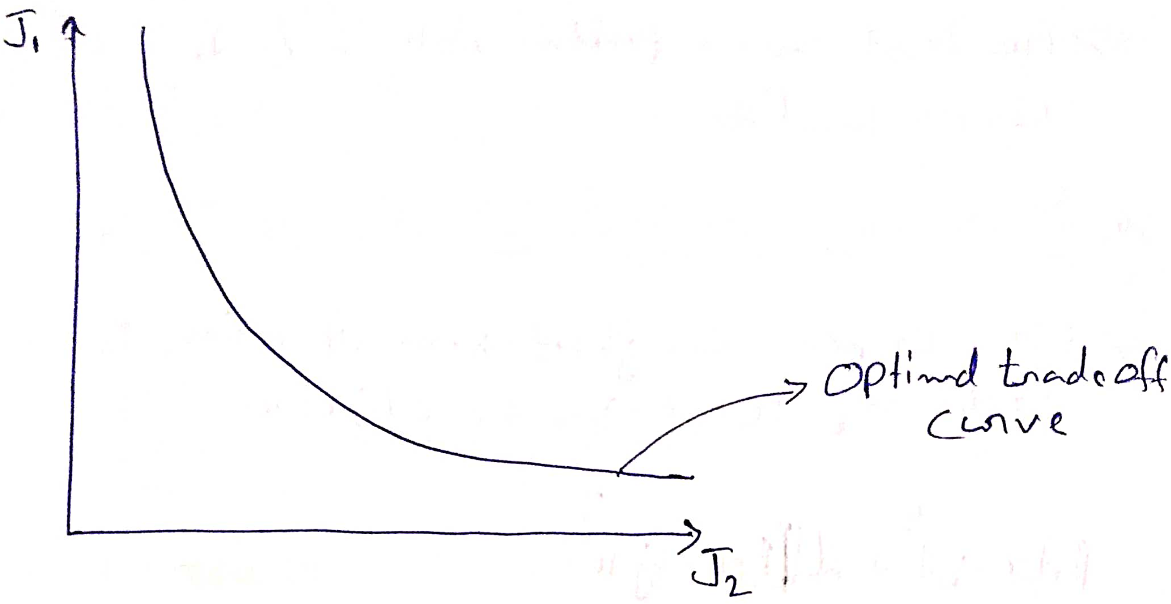
⇒ Plot (J_2, J_1) for every x;



All the points x^j mapped here are ambiguous, depends on the trade off

all the point x^j ~~that~~ Bare mapped to this reason are better the x^i in both objective.

all the points x^j that are mapped to this reason are worst then x^i in both objective



$\Rightarrow$ A point is called Pareto optimal if there is no other point in its 3rd quadrant.

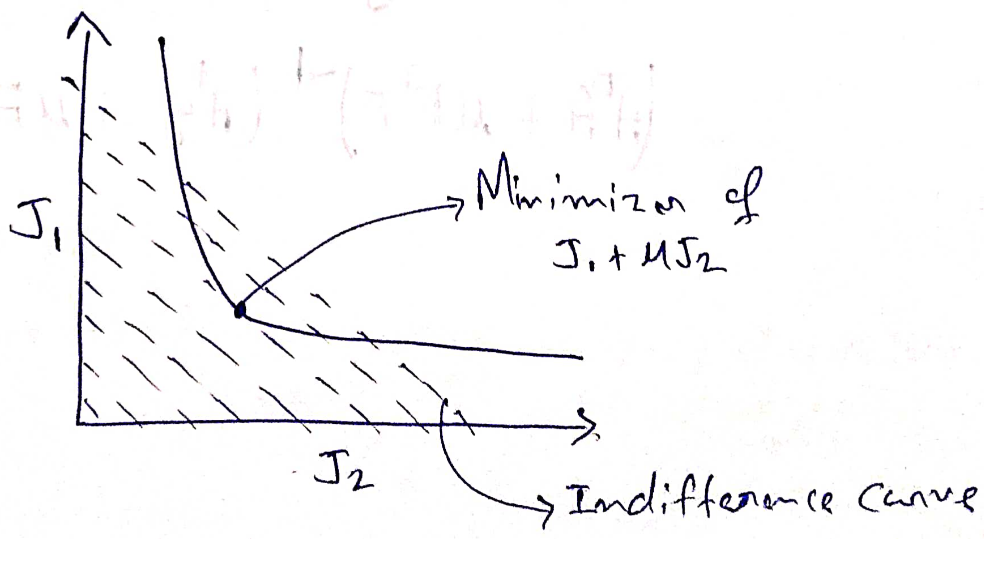
$\Rightarrow$ All points on optimal tradeoff curve is Pareto optimal.

★ Weighted-sum objective

$\Rightarrow$ To find Pareto optimal points, we minimize weighted-sum objective.

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

$\mu \geq 0$ gives relative weight between J_1 and J_2



⇒ For least square problem both J_1 & J_2 are convex function.

★ Minimizing weighted-sum objective

⇒ Can express weighted-sum objective as ordinary least-square objective:

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

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

$$= \|\bar{A}x-\bar{y}\|^2 \quad \text{where, } \bar{A} = \begin{bmatrix} A \\ \sqrt{\mu}F \end{bmatrix}$$

$$\bar{y} = \begin{bmatrix} y \\ \sqrt{\mu}g \end{bmatrix}$$

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

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

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

★ Re

⇒ C

★ Regularized least-square

⇒ When $F=I$, $g=0$ the objectives are

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

minimizer of weighted-sum objective,

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

is called regularized ~~sto~~ least-square solution of $y \approx Ax$.

{ In statistic it is called Ridge regression }

This formulae makes sense for any A
(Skinny, Fat, full rank, not full rank)

⇒ $A^TA + \mu I$ is invertable iff $\mu > 0$

⇒ Let Z be a non-zero element of null space of $A^TA + \mu I$

$$(A^TA + \mu I) Z = 0$$

$$Z^T (A^TA + \mu I) Z = 0$$

$$Z^T A^T A Z + \mu Z^T Z = 0$$

$$\|AZ\|^2 + \mu \|Z\| = 0$$

$\mu\|Z\| = 0$ → only if z is zero
&
$\|AZ\| = 0$

So assumption is contradicted

⇒ So, $A^TA + \mu I$ is invertable.

⇒ Estimation / inversion application

- → $Ax - y$ is sensor residual
- → model only accurate ∀ x small
- → regularized solution trades off sensor fit, size of x.

★ **Nonlinear least-Square (NLLS)**

⇒ find $x \in \mathbb{R}^n$ that minimizes

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

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

→ $r(x)$ is vector of residuals.

→ In general very hard to solve exactly.

→ Many good heuristics to complete locally optimal solution.

★ **Gauss-Newton method:**

⇒ given starting guess of x

repeat

→ linearize r near current guess ①

→ new guess is linear LS solution ②, using linearized r

until convergence

① Linearize r near current iterate $x^{(K)}$

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

where, Dr is the Jacobian $(Dr)_{ij} = \frac{\delta r_i}{\delta x_j}$

Provided x is near $x^{(K)}$

⇒ Write linearized approximation as

$$r(x) \approx A^{(K)}x - b^{(K)}$$

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

$$\| r(x) \|^2 \approx \| A^{(K)}x - b^{(K)} \|^2.$$

② Next iterate solves this linearized LS problem:

$$x^{(K+1)} = (A^{(K)T}A^{(K)})^{-1} A^{(K)T} b^{(K)}$$

⇒ Useful variation on Gauss-Newton:
add regularization term

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

So that next iterate is not too far from the previous one.

(hence, linearized model still pretty accurate)