Week 5 LEAST SQUARES PROBLEM (LS)

LS problem is a special case of Quadratic problems, which is very important.

In 1801, Carl Friedrich Gauss attempt to obtain trajectory of CERES behind the sun, without solving Kepler's equations which is non-linear and difficult to solve.

Gauss used L.S, which turned out to be the most accurate.

We have the problem,

$$\min_x \|Ax - b\|_2^2 = \min_x \underbrace{x^T A^T A x + 2b^T A x + b^T b}_{\text{Quadratic form}}$$

(QP) $\qquad A^T A \succeq 0$, which is PSD.

$x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$

This is convex.

(a) First case :

$b \in R(A) \Rightarrow$ There exists x such that linear combination of columns of A is x.

(ii) $\exists\, x$ s.t. $Ax = b$.

$\Rightarrow \min \|Ax - b\|_2^2 = 0$,

(b) Second case :

$b \notin R(A)$, $\nabla l(x) = 0$ (i.e.) $l(x) = \|Ax - b\|_2^2$

$$\Rightarrow 2A^TAx = 2A^Tb$$
$$\Rightarrow (A^TA)x = A^Tb$$
$$\Rightarrow x = \underbrace{(A^TA)^{-1}}A^Tb$$

when exists.

Eg $A = \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 2 \end{bmatrix}$, $A^TA = \begin{bmatrix} 3 & 6 \\ 6 & 12 \end{bmatrix}$ $R_2 = 2R_1$

So, this is a Rank-1 matrix. It is singular and not invertible. Therefore, Solution to $(A^TA)x = A^Tb$ is not unique. We can still solve it using SVD.

Approach : $A = U_{m\times m} \Sigma V^T_{n\times n}$, $A \in \mathbb{R}^{m\times n}$

then $\|b - Ax\|_2^2 = \|b - U\Sigma V^Tx\|_2^2$

$\tilde{x} = V^Tx$

$\Rightarrow x = V\tilde{x}$

$$= \min_{\tilde{x} \in \mathbb{R}^n} \|b - U\Sigma\tilde{x}\|_2^2$$

Orthogonal Matrix, U

vector, a

Aside: $\|Ua\|_2 = \|a\|_2$

since $\|Ua\|_2^2 = a^TU^TUa = a^Ta = \|a\|^2$

So, multiply with U^T,

$$= \|U^Tb - U^TU\Sigma\tilde{x}\|_2^2$$
$$= \|U^Tb - \Sigma\tilde{x}\|.$$

Denote $U^Tb = \tilde{b}$

$$\Rightarrow \min_{\tilde{x}} \|\tilde{b} - \Sigma\tilde{x}\|_2^2$$

where, $\tilde{x} \in \mathbb{R}^n$

$\tilde{b} \in \mathbb{R}^m$

$\Sigma \in \mathbb{R}^{m\times n}$

$$\Rightarrow \left\| \begin{bmatrix} \tilde{b}_1 - \sigma_1 \tilde{x}_1 \\ \tilde{b}_2 - \sigma_2 \tilde{x}_2 \\ \vdots \\ \tilde{b}_r - \sigma_r \tilde{x}_r \\ \tilde{b}_{r+1} \\ \vdots \\ \tilde{b}_m \end{bmatrix} \right\|_2^2$$

rank (A)

$$\Rightarrow \min_{\{\tilde{x}_i\}_{i=1}^r} \sum_{i=1}^{r} (\tilde{b}_i - \sigma_i \tilde{x}_i)^2 + \sum_{i=r+1}^{m} \tilde{b}_i^2$$

The Solution is

$$\tilde{x}_i = \begin{cases} \tilde{b}_i / \sigma_i, & i = 1, \dots, r \\ & \sigma_i > 0 \\ \text{arbitrary}, & i = r+1, \dots, m \end{cases}$$

$\tilde{x}_{r+1}, \dots, \tilde{x}_m$ do not effect objective.

$$\Rightarrow \tilde{x} = \Sigma^{+} \tilde{b}, \text{ where } [\Sigma^{+}]_{ii} = \begin{cases} 1/\sigma_i, & i = 1, \dots, r \\ 0, & \text{otherwise} \end{cases}$$

Final Solution

$$\boxed{x = \underbrace{V \Sigma^{+} U^{T}}_{\text{Pseudo-inverse}} b}$$

APPROXIMATION PROBLEMS

Having looked at the LS problem, where we minimize the l_2 norm, we can generalize the idea and look at the minimization of arbitrary norms, which gives us the general approximation problems.

Consider the l_1-norm minimization (a) $\min_{x \in \mathbb{R}^n} \|Ax - b\|_1$, $b \in \mathbb{R}^m$.

$$= \min_{x} \sum_{i=1}^{m} |a_i^T x - b_i|$$

Is this LP?

Applying Epigraph trick,

$$\Rightarrow \min_{(x, r)} \sum_{i=1}^{m} r_i \quad , \quad i = 1, 2, \dots, m$$

$$\text{s.t. } |a_i^T x - b_i| \le r_i$$

$r \in \mathbb{R}^m$

$(m+n)$ dimensional problem

$-r_i \le a_i^T x - b_i \le r_i$

In compact,

$$\boxed{\begin{aligned} &\min_{(x, r)} \quad \mathbb{1}^T r \\ &-r \le Ax - b \le r \end{aligned}}$$

$\mathbb{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$

Entrywise inequalities.

This is an LP.

In contrast with LS, (ii) l_2 - norm minimization

(LS)

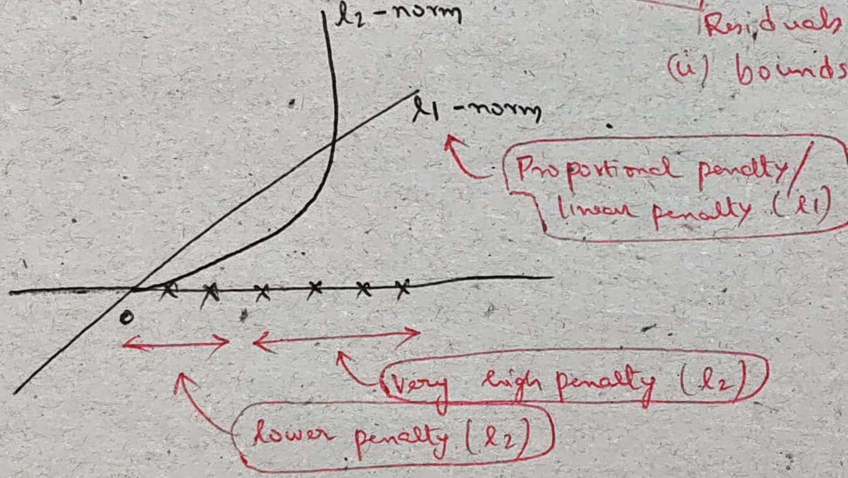
$$\min \sum_{i=1}^{m} r_i^2$$

$$-r_i \le a_i^T x - b_i \le r_i$$

Same constraint

Residuals

(ii) bounds on error.



- l_1 norm allows large residuals.
  l_2 norm does not allow large residuals
- l_2 norm minimize NO. of points with large residuals

---

l_∞ - norm minimization

$$\min_{x} \|Ax - b\|_\infty = \min_{x} \max_{1 \le i \le m} |a_i^T x - b_i|$$

Applying epigraph trick to convert into LP.

$$\Rightarrow \min_{(x,r)} r$$

$$\text{s.t.} \max_{1 \le i \le m} |a_i^T x - b_i| \le r$$

$$\Rightarrow \min r$$

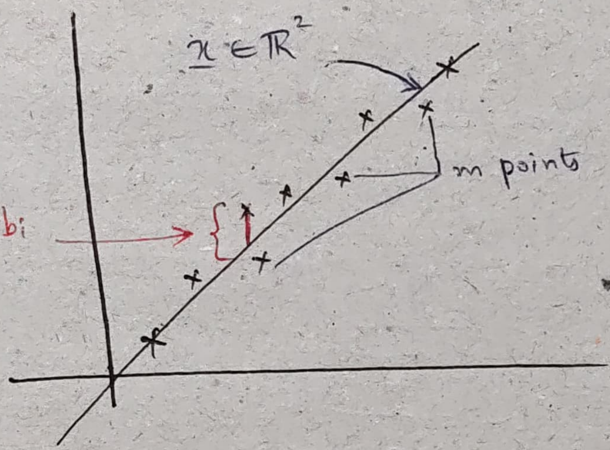
$$|a_i^T x - b_i| \le r \qquad \forall\, i = 1, \dots, m.$$

$$-r \le a_i^T x - b_i \le r \leftarrow \text{Affine constraint}$$

(ie) minimize largest residuals.

$\therefore$ This is an LP.

<u>Example</u>: Line fitting in $\mathbb{R}^2$ $(n=2)$



(slope, abscissa) $\rightarrow x \in \mathbb{R}^2$

$$A = \begin{bmatrix} 1 & a_1 \\ 1 & \vdots \\ 1 & \vdots \\ 1 & a_m \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

Residue, $r = A\underline{x} - b$, $r \in \mathbb{R}^m$
$A \in \mathbb{R}^{m \times 2}$

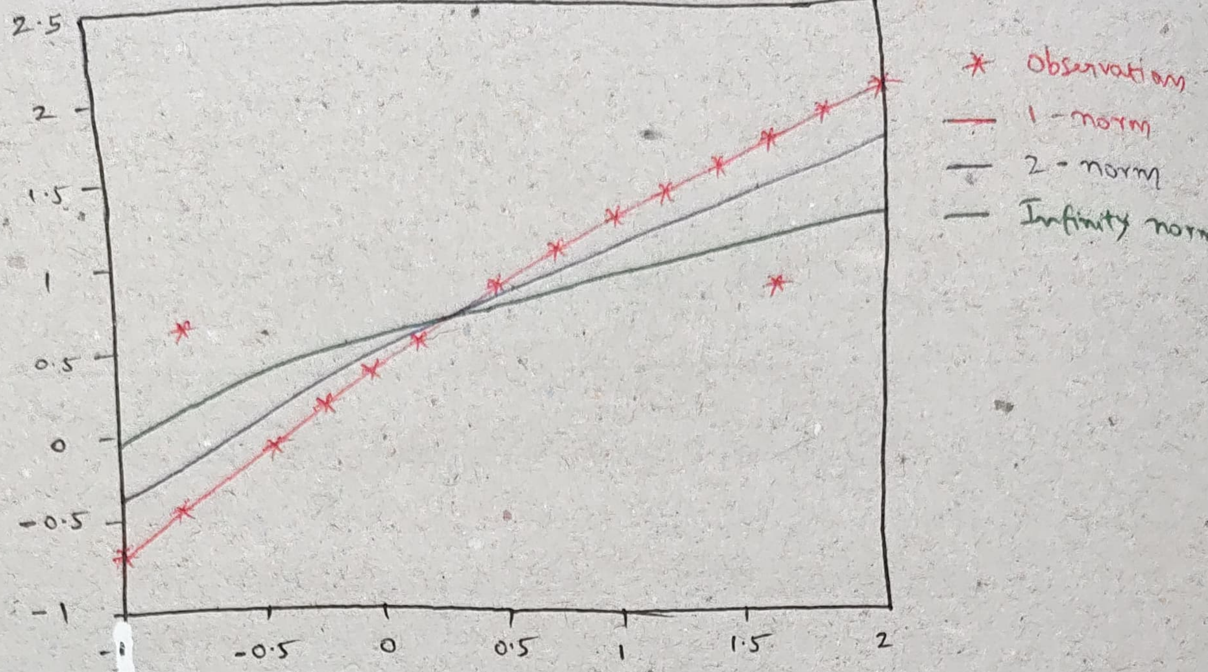
Given m points (A, b)

Goal: Find $x \in \mathbb{R}^2$, $p = 1, 2, \infty$

$$\min_x \|Ax - b\|_p$$

l_1	l_2	l_∞
$\sum r_i$	$\sum r_i^2$	$\max_i \lvert r_i \rvert$

$$\underline{r} = Ax - b$$



Second order Cone Program (SOCP)

Let us look at even more generalized form of optimization problem called SOCP.

SOCP takes the form,

$$\min_{x} \; f^T x$$

s.t. ① $\|A_i x - b_i\| \le \underbrace{c_i^T x + d_i}_{\geq 0}$, $i = 1, \dots, m$

② $Fx = g$

$$\underbrace{\|A_i x - b_i\|}_{\text{Norm (convex)}} - \underbrace{c_i^T x - d_i}_{\text{Affine}} \le 0 \quad = \text{CONVEX.}$$

Note: The form it is being expressed is very important.

Example: $\|x\| \le t$, $x \in \mathbb{R}^n$, $t \in \mathbb{R}$

$\Rightarrow \|x\|^2 \le t^2$

$\Rightarrow \underbrace{\|x\|^2 - t^2}_{} \le 0$ non-convex quadratic constraint.

$n = 1$, $x^2 - t^2 \le 0$

Hessian, $\nabla^2 (x^2 - t^2) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$

One Eigen value is Positive.

Another Eigen value is Negative. Therefore, this is not convex.

In terms of Algorithms, we have the following order.

$$LP \subseteq QP \subseteq \underline{QCQP} \subseteq SOCP$$

(i) LP is a special case of QP
QP is a special case of QCQP
QCQP is a special case of SOCP

In QP, we have $\underbrace{\frac{1}{2} x^T P x + q^T x}$,

$\downarrow$ converting QP into LP by setting $P = 0$.

$$q^T x$$

In QCQP, we have $\underbrace{\frac{1}{2} x^T P_i x + q_i^T x + r_i \leq 0}$

$\downarrow$ Converting QCQP into QP by setting $P_i = 0$.

$$q_i^T x + r_i \leq 0$$

LP is Fastest (Easy to Solve) $\longleftarrow$ Faster Algorithms — SOCP is Slowest (Difficult To solve)

<u>Claim</u>: $SOCP \subseteq QCQP$

$$\left.\begin{aligned} \min\ & \frac{1}{2} x^T P_0 x + q_0^T x \\ s.t\ & \frac{1}{2} x^T P_i x + q_i^T x + r_i \leq 0 \end{aligned}\right\} \quad P_0, P_i \succeq 0 \text{ PSD}$$

$$\| \sqrt{P_i}\, x + \sqrt{P_i}^{-1} q_i \|_2^2 - q_i^T P_i^{-1} q_i + r_i \leq 0$$

(or) $$\| \sqrt{P_i}\, x + \sqrt{P_i}^{-1} q_i \|_2 \leq \boxed{\underbrace{\sqrt{q_i^T P_i q_i - r_i}}_{\geq 0 \text{ (Problem infeasible otherwise)}}}$$

This is the Norm Ball / SOC constraint.

We write the objective function as

$$\min \| \sqrt{P_0}\, x + \sqrt{P_0}^{-1} q_0 \|_2^2 - \cancel{q_0^T P_0^{-1} q_0}$$ Skip the constant

$$\Rightarrow \min \| \sqrt{P_0}\, x + \sqrt{P_0}^{-1} q_0 \|_2$$

$$s.t.\ \| \sqrt{P_i}\, x + \sqrt{P_i}^{-1} q_i \|_2 \leq \sqrt{q_i^T P_i^{-1} q_i - r_i}$$

Applying Epigraph trick,

$$\min \; t$$

$$\left\| \sqrt{P_0}\, x + \sqrt{P_0}^{-1} q_0 \right\|_2 \le t$$ SOCP constraints.

$$\left\| \sqrt{P_i}\, x + \sqrt{P_i}^{-1} q_i \right\|_2 \le \sqrt{q_i^T P_i^{-1} q_i - r_i}$$

norm ball constraints (also SOCP)

Therefore, we have converted QCQP into SOCP.
But the opposite is not possible.

Robustness to parameter variations

Let us look into Robust optimization, in particular we are concerned with Robustness to parameter variations.

Let us look into an Example (Problems in Logistics)

How much goods to ship,

but allow minor variations in traffic / demand, etc.,

Uncertain quantities

Another example is communication

Variations in channel gains.

How to allow for a margin of error?

consider a LP,

$$\min \; c^T x$$
$$\text{s.t. } Ax \le b$$
$$\text{(i.e.) } a_i^T x \le b_i, \; i = 1, \dots, m.$$

Discrete Set

$A \in \mathcal{A} = \{A_1, A_2, \dots, A_k\}$

k is not too large.

Formulation: Constraint should be satisfied always.

$$\boxed{\begin{aligned} &\min \; c^T x \\ &A_i x \le b, \; i = 1, \dots, k \end{aligned}}$$

Constraint is satisfied in the worst case.

Note: More constraints $\Rightarrow$ Minimum objective value increases

Very Robust Formulation $\Rightarrow$ Cost increases.

L.S. Example.

$$\min \|Ax - b\|_2$$

$$A \in \mathcal{A} = \{A_1, \dots, A_k\}$$ Calligraphic A.

Let us denote worst case error (Maximum error that could incur)

$$e_{wc}(x) = \max_{A \in \mathcal{A}} \|Ax - b\|_2$$

$$= \max_{i} \|A_i x - b\|_2$$

$$= \min_{x} \max_{1 \le i \le k} \|A_i x - b\|_2$$

Therefore,

$$\min_x e_{wc}(x) \overset{\text{(epigraphic trick)}}{=\!=\!=} \min_x t$$

$$\max_{1 \le i \le k} \|A_i x - b\|_2 \le t$$

$$\text{(SOCP)} \quad = \quad \min_{(x,t)} t$$

$$\|A_i x - b\|_2 \le t, \quad i = 1 \dots k.$$

This is the Robust LS problem.

Since, the objective is linear, constraint involved is

Norm $\le$ Affine

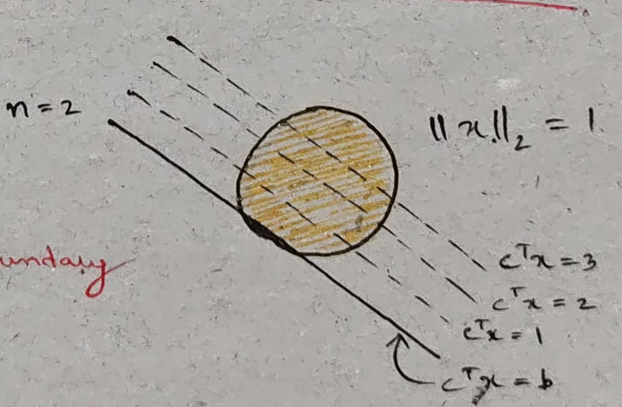
Therefore, this is called SOCP.

Minimising Linear objective over a norm ball

$$\min c^T x$$
$$\|x\|_2 \le 1$$

Solution is on the boundary of the norm ball.

So, $\|x^*\|_2 = 1$



From Cauchy-Schwarz inequality, we have

$$c^T x \geq -\|c\|_2 \, \|x\|_2$$

$$\Rightarrow \underbrace{c^T x^* \geq -\|c\|_2 \, \|x^*\|_2 = -\|c\|_2}_{\text{This is Inequality}}$$

The equality happens when $x^* = -\alpha c$, $\alpha \in \mathbb{R}_+$

$$\Rightarrow c^T(x^*) = c^T(-\alpha c) = -\alpha\|c\|_2^2 = -\|c\|_2$$

$$\Rightarrow \alpha = \frac{1}{\|c\|_2}$$

Thus, $\left.\begin{aligned} x^* &= \frac{-c}{\|c\|_2} \\ \Rightarrow c^T x^* &= -\|c\|_2 \end{aligned}\right\}$ This is also called as closed-form solution. (ie) Solution expressible in analytical form.

<u>Key steps</u>:

- Establish a bound
- Show that it is achieved

Ellipsoidal Uncertainity

So far, we've seen discrete uncertainities. Now we'll see ellipsoidal uncertainity.

Consider the problem

$$\text{(LP)} \qquad \min \; c^T x \quad \text{s.t.} \; a^T x \leq b$$

(simple hypoplane halfspace constraint)

$$a \in \overbrace{\mathcal{E}(a_0, P)}^{\text{Ellipsoid with center } a_0 \text{ and Matrix } P.} = \{ a \in \mathbb{R}^n \mid (a-a_0)^T P^{-1} (a-a_0) \leq 1 \}$$

'a' could be any point in this ellipsoid. We want to solve the optimization problem such that the worst possible error that will incur is minimized, regardless of what value 'a' takes.

This is the ellipsoidal uncertainity for Robust optimization problems.

P is symmetric positive definite (ie) $P \succ 0$.

$n=1$ case, $a \in [-1, 1]$ interval

For large n, we can work out similarly.

⊙ $n=1$

$$\min \; cx$$
$$ax \le b, \quad a \in [-1, 1]$$

amounts to infinite number of constraints.

(or)

$$\min \; cx$$
$$\left(\max_{a \in [-1,1]} ax \right) \le b.$$

<u>Aside</u>:

$$\max_{a \in [-1,1]} ax \quad (\text{given } x) \quad = |x|$$

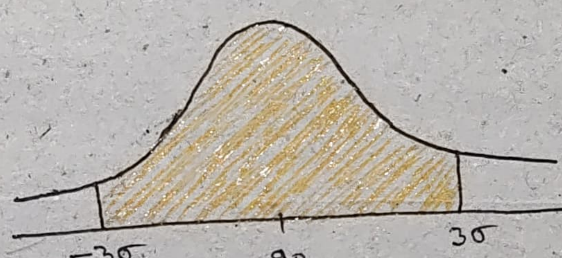
optimal value, $a^* = \begin{cases} 1 & \text{when } x > 0 \\ -1 & \text{when } x < 0 \end{cases}$

$$\Rightarrow a^* x = |x|$$

$$\min_{|x| \le b} cx \iff \min_{-b \le x \le b} cx$$

<u>Eg.</u> $a \sim \mathcal{N}(a_0, \sigma^2)$

$a_0 \in [a_0 - 3\sigma \,;\, a_0 + 3\sigma]$



$$a - 3\sigma \le a \le a + 3\sigma$$

Now, let us see how to solve the general problem.

$$\min \; c^T x \qquad \iff \qquad \min_x \; c^T x$$
$$a^T x \le b, \quad a \in \mathcal{E}(a_0, P) \qquad \qquad \underbrace{\left(\max_{a \in \mathcal{E}(a_0, P)} a^T x \right)}_{\text{w.r.t } a} \le b$$

Find LHS such that a lies in this uncertainity region.

<u>Aside</u>:

given x: $\max_a \; a^T x$

$$\text{s.t.} \; (a - a_0)^T P^{-1} (a - a_0) \le 1$$

$$a \in \mathcal{E}(a_0, P) = \left\{ a_0 + \sqrt{P}\, u \;\middle|\; \|u\|_2 \le 1 \right\}$$

$\therefore$ The optimization problem can be written as

$$\max_{u} \ (a_0 + \sqrt{P}\, u)^T x \ = \ a_0^T x + \boxed{\begin{array}{l} \max_{u} \ u^T \sqrt{P}\, x \\ \text{s.t. } \|u\|_2 \le 1 \end{array}}$$

$$\text{s.t. } \|u\|_2 \le 1$$

This is the revised optimization problem

WKT.

If $w = \sqrt{P}\, x$, then $\max_{u} \ u^T w$

$$\text{s.t. } \|u\|_2 \le 1$$

$$\Rightarrow u^* = \frac{w}{\|w\|}$$

$$\Rightarrow u^{*T} w = \|w\|_2$$

$$\min \ c^T x$$

Linear objective

$$\|\sqrt{P}\, x\|_2 \le b - a_0^T x$$

affine function

which is SOCP.

This is an example application of convex optimization and all its manipulations that we need to do in order to get a formulation in a form that is reasonable, that we can input to any software and solve.
