EECS 127 Lecture 11.

Today:

- Convex Opt.
- Transformations
  - ↳ LASSO
  - ↳ logistic regression.

Admin

- Midterm: March 12. 5-7 pm.
- Review. March 9.

$$p^* = \min_x f_0(x)$$

s.t. $f_i(x) \leq 0 \quad i = 1, \ldots m.$

$Ax = b$

Convex problem: f_i's $i = 0, 1, \ldots m$ to be convex

Implicit constraint: $D = \bigcap_{i=0}^{m} \operatorname{dom} f_i$'s

Classic example: LP: Linear program.

$$\min_{\vec{x}} \vec{c}^T \vec{x}$$

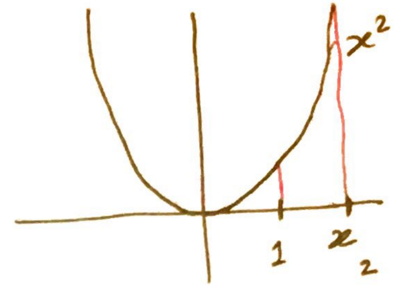
s.t. $A\vec{x} = \vec{b}$

example:

$$\min_{x \in \mathbb{R}} x^2 = p^*$$

$x \geq 1$

$x \leq 2$



In general, if $f(x)$ is convex

to find optimum point: $\nabla f(x) = 0$.

Constraint: $x \geq 1$ is called an "active constraint"

$$\boxed{\begin{array}{l} \min \ x^2 \\ x \geq 1 \\ x \leq 0 \end{array}}$$

Problem is infeasible.

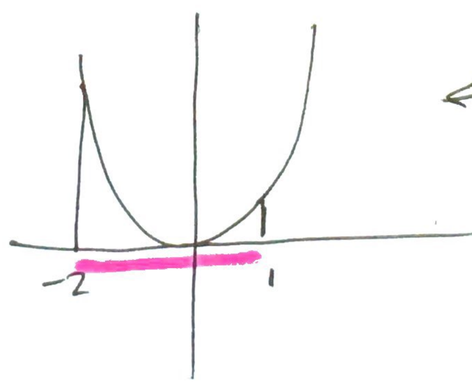
& Feasible set $= \phi$

$p^* = \infty$

$\min x^2$

$x \geq -2$

$x \leq 1$



← No active constraints.

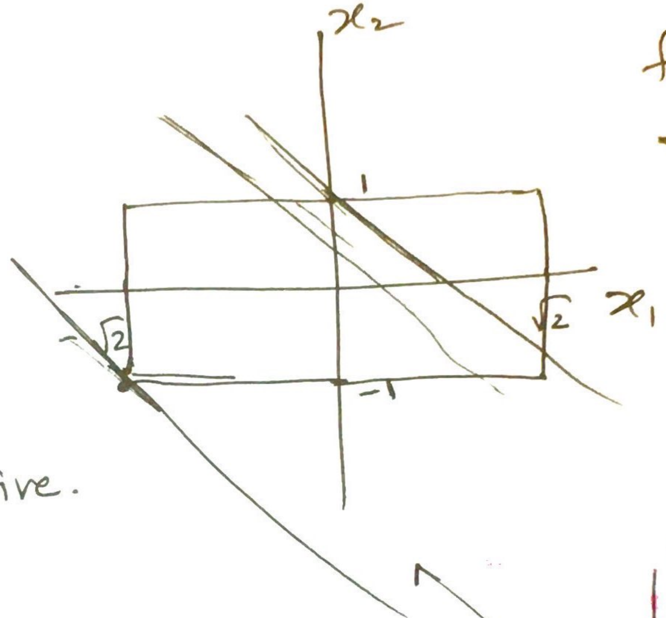
example:

$$\min \quad x_1 + x_2$$
$$\vec{x} \in \mathbb{R}^2$$
$$x_1^2 \leq 2$$
$$x_2^2 \leq 1$$

$\vec{c}^T \vec{x}$ $\quad \vec{c} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$f_0(x) = x_1 + x_2.$

$\nabla f_0(x) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$



$x_1 = -\sqrt{2}$

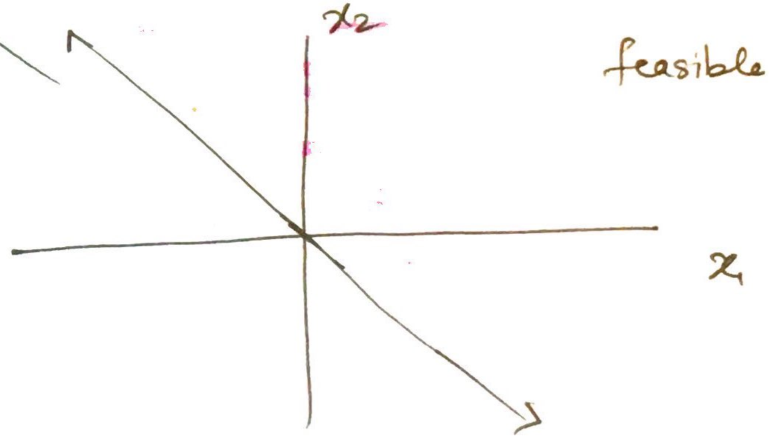
$x_2 = -1$

Both constraints active.

example: Consider: $\min \quad x_1 = p^*$

s.t. $x_1 + x_2 \geq 0$

$p^* = -\infty$

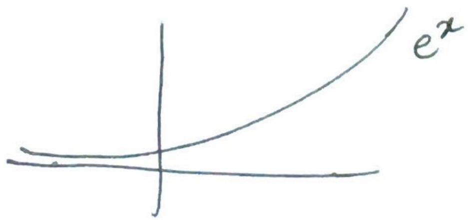


feasible half plane

Infimum v/s minimum.

$\min \quad e^x$

$x > 0$



Thm:

$$\min_{x \in \mathcal{X}} \vec{c}^{\,T} x$$

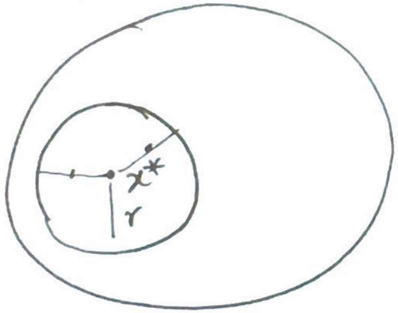
$\mathcal{X}$: convex set.

Closed set = i.e. $\mathcal{X}$ includes its boundary.

If x^* is an optimal solution to this problem, then x^* belongs to the boundary of $\mathcal{X}$.

Proof: If possible let $x^* \in$ interior of $\mathcal{X}$

Then there exists some radius $r > 0$ such that a ball of radius r is in the interior of $\mathcal{X}$.



$$\forall \vec{z} \;\; \|\vec{x} - \vec{z}\|_2 \leq r \quad , \quad \vec{z} \in \mathcal{X}.$$

$$\nabla f_0(x) = \vec{c}$$

$$\alpha = \frac{r}{\|\vec{c}\|_2}$$

Consider: $\vec{z} = \alpha \cdot \vec{c}$

Compute: $\vec{c}^{\,T} (\vec{z}) = c^T \cdot \alpha \cdot \vec{c} = \alpha \|\vec{c}\|_2^2$

Consider : $\vec{x}^* - \vec{z}$

$$f_0(\vec{x}^* - \vec{z}) = \vec{c}^{\,T} \cdot x^* - \alpha \cdot \vec{c}^{\,T} \cdot c = \vec{c}^{\,T} \cdot \vec{x}^* - \frac{r}{\|\vec{c}\|_2} \cdot \|\vec{c}\|_2^2.$$

$$= \vec{c}^{\,T} \cdot \vec{x}^* - r \|\vec{c}\|_2$$

Problem Transformations

① Monotone transformations.

② Addition of slack variables.

---

② $\min\limits_{\vec{x} \in \mathcal{X}} f_0(\vec{x})$. $\Longleftrightarrow$ $\min\limits_{t} t$
$f(x) \leq t$.

"epigraph reformulation"

t: slack variable.

USE:

$$p^* = \min_{x} \|A\vec{x} - \vec{y}\|_2^2 + \|\vec{x}\|_1$$

LASSO.

$$\|\vec{x}\|_1 = \sum_{i=1}^{n} |x_i|$$

$$p^* = \min_{x, t_i} \|A\vec{x} - y\|_2^2 + \sum_{i=1}^{n} t_i$$

Quadratic Program.

s.t. $|x_i| \leq t_i \quad \forall i \Leftrightarrow -t_i \leq x_i \leq t_i$

Monotone transformations.

If. $\phi(x)$ is continuous + strictly increasing.

$$g^* = \min_x \ \phi(f_0(\vec{x})).$$
$$\text{s.t. } f_i(\vec{x}) \le 0 \quad i = 1, \dots m.$$
$$A\vec{x} = b$$

$$p^* = \min_x \ f_0(\vec{x})$$
$$\text{s.t. } f_i(\vec{x}) \le 0 \ . \quad i = 1, \dots m.$$
$$A\vec{x} = \vec{b}.$$

Logistic Regression.

$x_1, x_2 \dots x_n. \ \dots$ (Points)

$1, 1, (-1), (-1) \dots$ (Labels).

$y_1, y_2 \dots y_n.$

Want: Predict: Probability that a data point belongs to class $(+1)$, (-1)

$$P(Y = 1 \mid \vec{X} = \vec{x}_i) = p(\vec{x}).$$

Want: A linear function of $\vec{x}$ that gives us $P(Y=1 \mid X = x_i)$.

① $\underset{\in[0,1]}{p(\vec{x})} = \underbrace{\vec{w}^T\vec{x} + \beta}_{\text{unbounded.}}$ $\longrightarrow$ Fails : range issues

no diminishing returns.

② $\log p(x) = \vec{w}^T \vec{x} + \beta$

$\in (-\infty, 0)$

<u>Fails</u> log ~~..~~ does not take $p(x) \to +\infty$. $p(x) \in [0, 1]$.

③ $\log \frac{p(x)}{1-p(x)} = \underbrace{\vec{W}^T \vec{x} + \beta}_{\text{linear}}$.

↳ Rewrite $p(\vec{x})$

$$\frac{p(\vec{x})}{1-p(\vec{x})} = \exp(\vec{w}^T \vec{x} + \beta)$$

$$\Longrightarrow \quad p(\vec{x}) = [1 - p(\vec{x})] \exp(\vec{w}^T \vec{x} + \beta).$$

$$p(\vec{x}) = \frac{\exp(\vec{w}^T \vec{x} + \beta)}{1 + \exp(\vec{w}^T \vec{x} + \beta)} = P(Y = 1 \mid \vec{X} = \vec{x})$$

<u>Max. liklihood</u>: <u>Maximize</u>: $\mathbb{P}(Y_1, Y_2 \ldots Y_n) = \max. \prod_{i=1}^{n} P(y_i)$

$$P(Y=1) = \frac{\exp(\vec{W}^T\vec{x}+\beta)}{1+\exp(\vec{W}^T\vec{x}+\beta)}.$$

$$P(Y=-1) = 1 - \left(\qquad \right) = \frac{1}{1+\exp(\vec{W}^T\vec{x}+\beta)} = \frac{\exp(-(\vec{W}^T\vec{x}+\beta))}{\exp(-(\vec{W}^T x+\beta))+1}$$

$$P(Y=y_i) = \frac{\exp y_i(\vec{W}^T\vec{x}+\beta)}{1+\exp(y_i(\vec{W}^T x+\beta))}$$

Optimization:

$$\underset{\vec{\omega},\beta}{\text{Maximize}} \ \log\left(\prod_{i=1}^{n} \frac{\exp y_i(\vec{\omega}^T\vec{x}_i+\beta)}{1+\exp(y_i(\vec{W}^T\vec{x}_i+\beta))} \right)$$

Maximize: $\log\left(\qquad \right)$