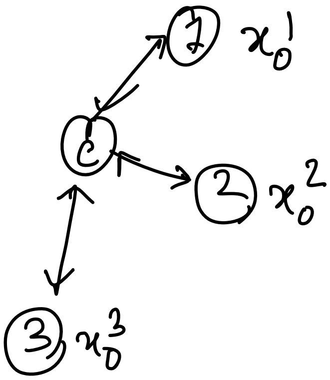
Lecture 30: 27th March

Local SGD

$$\min_x f(x) = \frac{1}{N}\sum_{i=1}^{N} f_i(x),$$

- N agents or servers who are not allowed to share f_i with others.



- Let $\mathcal{I} \subseteq \{0,1,2 \dots\}$ be the synchronization time-steps.

e.g: $\mathcal{I} = \{10, 20, 30 \dots\}$

If $f_i(x) = \sum_{j=1}^{N_i} \ell(x, \xi_i^j)$, then we can replace $\nabla f_i(x)$ with $\underline{\nabla \ell(x, \xi_i^{t_j})}$ where t_j is randomly chosen $\{1, 2 \dots N_i\}$

- at any point of time t,
  - if $t \notin \mathcal{I}$:

$$\underline{x_{t+1}^i} = x_t^i - \eta_t \underline{\nabla f_i(x_t^i)}$$

  - if $\underline{t \in \mathcal{I}}$:
    - $\underline{\bar{x}_{t+1}^i = \frac{1}{N}\sum_{i=1}^{N} x_t^i}$ average calculated by central server.
    - $x_{t+1}^i = \bar{x}_{t+1}^i - \eta_t \nabla f_i(\bar{x}_{t+1}^i)$
  - at this stage,

$\bar{x}_{t+1}^i$ is same for all i

When each f_i is β-smooth & α-strongly convex,

convergence rate $\theta\left(\frac{1}{KT}\right)$,

where K: condition number.

Distributed Optimization

Consider the optimization problem: $\underline{\min_{x} \sum_{i=1}^{N} f_i(x)}$, where f_i is known only to agent i.

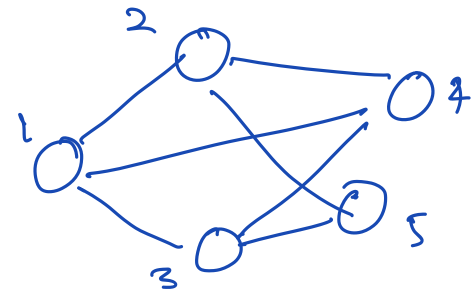
ex: $$\min_{x} \sum_{i=1}^{N} \left(\sum_{j=1}^{D_i} \underline{\left(a_{ij}^T x - b_{ij}\right)^2} \right),$$

D_i: number of data points available with agent i

Agents can communicate over a graph or network $G=(V,E)$

each node corresponds to an agent, $|V|=N$.

$(i,j)\in E$, then agent i communicates with agent j



$\underline{w_{ij}}$: weight of the edge (i,j)

if $w_{ij}=0 \Rightarrow (i,j)\notin E$

$(1,2)\in E$, $(2,3)\notin E$.

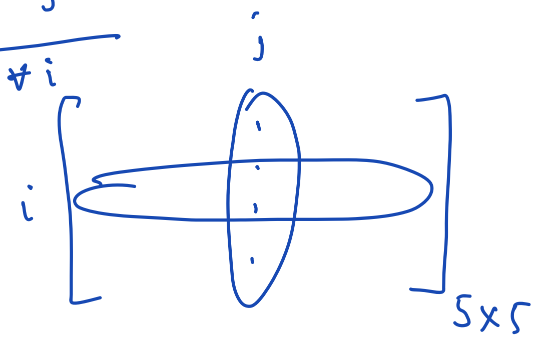
For agent i, $\mathcal{N}_i \subseteq V$ to be its neighbors, $\mathcal{N}_i = \{ j\in V \mid (i,j)\in E\}$

- every agent starts with an initial solution x_0^i at time $=0$.
- at every time t
  - $\rightarrow$ gather: agent i obtains $\underline{x_t^j}$ from its neighbors $j\in \mathcal{N}_i$
  - $\rightarrow$ computes: $\underline{\bar{x}_{t+1}^i = \sum_{j\in \mathcal{N}_i} w_{ij}\, x_t^j}$
  - $\rightarrow$ gradient descent: $x_{t+1}^i = \bar{x}_{t+1}^i - \eta_t \underline{\nabla f_i\left(\bar{x}_{t+1}^i\right)}$.

This algorithm is known as distributed gradient descent.

- Assumptions on the network/graph
  - The graph G is connected, undirected,
  - weights satisfy $\sum_{i=1}^{N} w_{ij} = 1 \;\; \forall j$, $\sum_{j=1}^{N} w_{ij} = 1 \;\; \forall i$

    (the entries in each column & row adds to 1)



- Assumptions on step-sizes
  - $\eta_t \geq 0$, $\sum_{t=0}^{\infty} \eta_t = \infty$, $\sum_{t=0}^{\infty} \eta_t^2 < \infty$.
- Assumptions on cost functions
  - each f_i is convex, $\|\nabla f_i(x)\| \leq G_i \quad \forall x \in X$
  - the problem has an optimal solution, denoted x^*

Theorem: The solutions generated by the distributed gradient descent scheme converges to x^* for each agent i.

$$\lim_{t \to \infty} \|x_t^i - x^*\|_2 = 0 \quad \text{for all } i.$$

Here $x^* \in \arg\min \sum_{i=1}^{N} f_i(x)$

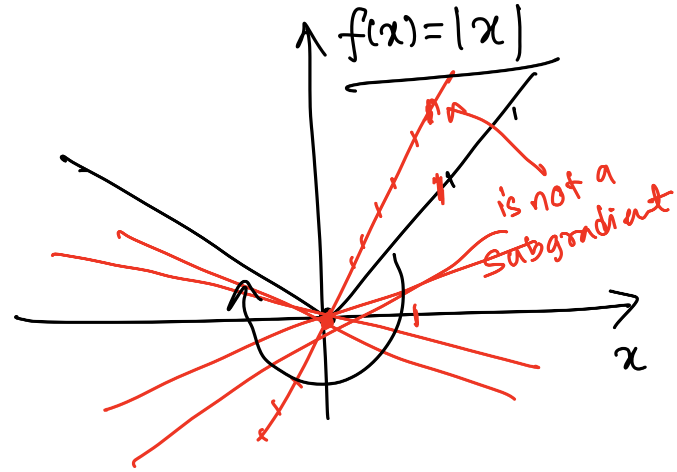
proof involves two steps

1) consensus error goes to 0.

$$\lim_{t \to \infty} \underbrace{\left\| x_t^i - \frac{1}{N} \sum_{j=1}^{N} x_t^j \right\|_2}_{\text{consensus error}} \to 0.$$

2) convergence to the optimal solution: $\lim_{t\to\infty} \left\| \frac{1}{N}\sum_{j=1}^{N} x_t^j - x^* \right\|_2 = 0.$

Subgradients



Definition: For a function $f:\mathbb{R}^n \to \mathbb{R}$, a vector $g \in \mathbb{R}^n$ is a subgradient of f at point x if

$$f(y) \geq f(x) + g^T(y-x)$$

for all $y \in \text{dom} f$.

The set of all subgradients is called subdifferential, denoted by $\partial f(x) = \{ g \in \mathbb{R}^n \mid g \text{ is a subgradient of } f \text{ at } x \}$

If f is a convex function, $\partial f(x) \neq \emptyset$ when $x \in \text{int. dom} f$.

If f is differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$.

For $f(x) = |x|$, $\quad \partial f(x) = \begin{cases} \{1\} & \text{if } x > 0 \\ \{-1\} & \text{if } x < 0 \\ [-1,1] & \text{if } x = 0. \end{cases}$

Functions that are not differentiable everywhere are not smooth, but they can be strongly convex.
Need to be careful while choosing step-sizes for subgradient descent. Usually, square summable stepsizes work.

Subgradient descent: $x_{t+1} = x_t - \eta_t g_t$, where $g_t \in \partial f(x_t)$.

Dual Norm

For a given norm $\|\cdot\|$, its dual norm $\|\cdot\|_*$ is defined as

$$\|y\|_* = \sup_{\|x\| \le 1} x^T y$$

For 2-norm

$$\|y\|_* = \sup_{\|x\|_2 \le 1} x^T y \to \le \|x\|_2 \|y\|_2$$

$$= \|y\|_2$$

$$x = \frac{y}{\|y\|_2}, \quad \|x\|_2 = \frac{\|y\|_2}{\|y\|_2} = 1$$

$$x^T y = \frac{\|y\|_2^2}{\|y\|_2} = \|y\|_2$$

$\downarrow$

dual norm of 2-norm is 2-norm.

Holder's inequality: $x^T y \le \|x\|_p \|y\|_q$, when $\frac{1}{p} + \frac{1}{q} = 1$

$$\|x\|_p := \left[\sum_{i=1}^{n} |x_i|^p\right]^{1/p}$$

dual of p-norm is q-norm.

dual norm of ∞-norm is 1-norm

In general, we can show

$$\|x\|_1 = \sup_{\|y\|_\infty \le 1} x^T y = \|x\|_{*-\infty}$$

$$\|x\|_\infty = \sup_{\|y\|_1 \le 1} x^T y.$$

$\|y\|_\infty \le 1 \Rightarrow \max_i |y_i| \le 1$

$$x^T y = \sum_{i=1}^{n} x_i y_i \le \sum_{i=1}^{n} |x_i| = \|x\|_1$$

$x_i y_i$ is maximized when $y_i = sgn(x_i) = \begin{cases} 1 & \text{if } x_i \ge 0 \\ -1 & \text{if } x_i < 0 \end{cases}$

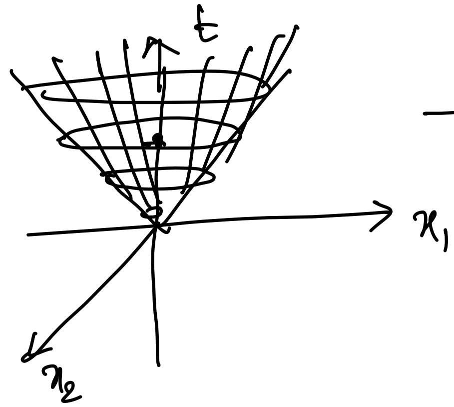
Lecture 31 : 28th March

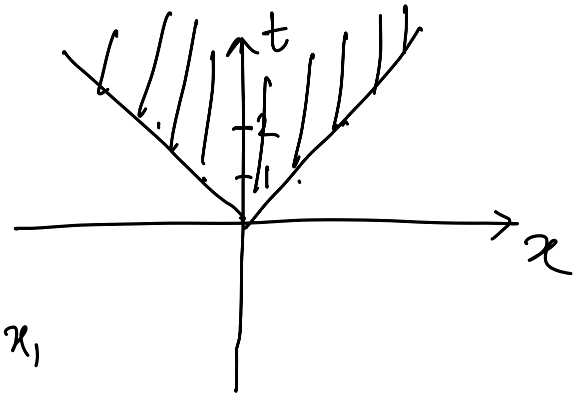
Module D: SOCP and Robust Optimization

Second Order Cone: in $\mathbb{R}^{n+1}$ is the set $\{(x,t),\ x\in\mathbb{R}^n,\ t\in\mathbb{R} \mid \|x\|_2 \le t\}$

Ex: In $\mathbb{R}^2$ $\{(x,t) \mid \|x\|_2 \le t\}$

[a convex set]





Rotated Second Order Cone:

$$K_n^{rc} = \{(x,y,z),\ x\in\mathbb{R}^n,\ y\in\mathbb{R},\ z\in\mathbb{R} \mid x^T x \le 2yz,\ y\ge 0, z\ge 0\}$$

$$\Updownarrow$$

$$\Longleftrightarrow \left\| \underbrace{\begin{bmatrix} x \\ \frac{1}{\sqrt{2}}(y-z) \end{bmatrix}}_{w} \right\|_2 \le \underbrace{\frac{1}{\sqrt{2}}(y+z)}_{q} \quad \cdots (*)$$

we can find a matrix M s.t

$$\begin{bmatrix} w \\ q \end{bmatrix} = M \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

By taking square on both sides of $(*)$, we obtain

$$\begin{bmatrix} x & \frac{1}{\sqrt{2}}(y-z) \end{bmatrix} \begin{bmatrix} x \\ \frac{1}{\sqrt{2}}(y-z) \end{bmatrix} \le \frac{1}{2}(y+z)^2$$

$$\Longleftrightarrow x^T x + \frac{1}{2}(y-z)^2 \le \frac{1}{2}(y+z)^2$$

$$\Longleftrightarrow x^T x - yz \le yz \Longleftrightarrow x^T x \le 2yz$$

Quadratic Constraint as Second Order Cone Constraint

Consider a quadratic constraint $x^T Q x + \underline{c^T x + d} \leq 0$,

with $\underline{Q = Q^T \succeq 0}$.

Let $W = \sqrt{Q}$.

Then $\underbrace{x^T W^T \underbrace{W x}_{\bar{x}}} \leq \underline{(-c^T x - d)}$

$$\Leftrightarrow (\bar{x})^T \bar{x} \leq 2 \underbrace{\left(\tfrac{1}{2}\right)}_{\bar{z}} \underbrace{(-c^T x - d)}_{\bar{y}}$$

$$\Updownarrow$$

$$\left\| \begin{bmatrix} \bar{x} \\ \frac{1}{\sqrt{2}}(\bar{y} - \bar{z}) \end{bmatrix} \right\|_2 \leq \frac{1}{\sqrt{2}}(\bar{y} + \bar{z})$$

$$\Leftrightarrow \left\| \begin{bmatrix} Q^{1/2} x \\ \frac{1}{\sqrt{2}}(-c^T x - d - 1/2) \end{bmatrix} \right\|_2 \leq \frac{1}{\sqrt{2}}\left(-c^T x - d + \frac{1}{2}\right)$$

$$\Leftrightarrow \left\| \begin{bmatrix} \sqrt{2}\, Q^{1/2} x \\ -c^T x - d - 1/2 \end{bmatrix} \right\|_2 \leq \left(-c^T x - d + \frac{1}{2}\right)$$

which is a second order cone constraint.

Second Order Cone Programming (SOCP) in Standard Form

$$\min_{x \in \mathbb{R}^n} \quad c^T x$$
$$\text{s.t.} \quad \|A_i x + b_i\|_2 \leq c_i^T x + d_i, \qquad i = 1, 2, \dots m$$

LPs are SOCPs:

$$\min_{x \in \mathbb{R}^n} \quad c^T x$$
$$\text{s.t.} \quad a_i^T x + b_i \leq 0 \Rightarrow 0 \leq -a_i^T x - b_i \Leftrightarrow \left[\begin{array}{l} A_i = 0, b_i = 0 \\ c_i = -a_i \\ d_i = -b_i \end{array}\right.$$

QPs are SOCPs:

$$\left[\begin{array}{ll} \min_{x \in \mathbb{R}^n} & x^T Q x + r^T x \\ \text{s.t.} & a_i^T x + b_i \leq 0 \Rightarrow \text{SOCP constraints} \end{array}\right.$$

$$\Updownarrow$$

$$\min_{x, t} \quad t$$
$$\text{s.t.} \quad x^T Q x + r^T x \leq t$$
$$a_i^T x + b_i \leq 0$$

as seen above, this quadratic constraint can be written as a second order cone constraint.

QCQPs are SOCPs:

$$\min_{x \in \mathbb{R}^n} \quad x^T Q x + r^T x$$
$$\text{s.t.} \quad x^T Q_i x + r_i^T x \leq z_i, \quad i = 1, 2 \dots m$$

can also be framed as a second order cone constraint as shown above.

Min-max Inequality

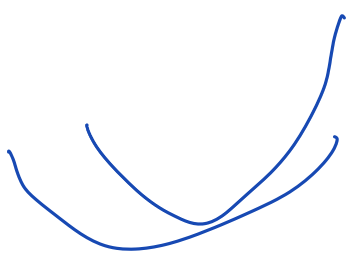
Consider a function $\phi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and let X and Y are two non-empty subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively.

Then

$$\underline{\sup_{y \in Y} \inf_{x \in X}} \phi(x,y) \leq \underline{\inf_{x \in X} \sup_{y \in Y}} \phi(x,y).$$

<u>proof</u>: Let us fix $\tilde{y}$. Then

$$\underbrace{\phi(x,\tilde{y})}_{= f_1(x)} \leq \underbrace{\sup_{y \in Y} \phi(x,y)}_{= f_2(x)}$$



$$\Rightarrow f_1(x) \leq f_2(x) \quad \forall x$$

$$\Rightarrow \inf_{x \in X} f_1(x) \leq \inf_{x \in X} f_2(x)$$

$$\Rightarrow \inf_{x \in X} \phi(x,\tilde{y}) \leq \underbrace{\inf_{x \in X} \sup_{y \in Y} \phi(x,y)}_{\text{does not depend on } \tilde{y}}.$$

$$\Rightarrow \sup_{\tilde{y} \in Y} \inf_{x \in X} \phi(x,\tilde{y}) \leq \inf_{x \in X} \sup_{y \in Y} \phi(x,y).$$

<u>When does equality hold?</u>

Sion's Theorem: Let $X \subseteq \mathbb{R}^n$ be convex and compact. Let $Y \subseteq \mathbb{R}^m$ be a convex set.

For each fixed $y \in Y$, $\phi(\cdot, y)$ is a convex function (of x)

For each fixed $x \in X$, $\phi(x, \cdot)$ is a concave function (of y).

Then $\sup_{y \in Y} \min_{x \in X} \phi(x,y) = \min_{x \in X} \sup_{y \in Y} \phi(x,y)$.

Lagrangian and Saddle Points

We have already encountered min-max problems in the context of Lagrangian duality.

$$\mathcal{L}(x,\lambda,\mu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)$$

When the optimization problem is convex, and $\lambda \geq 0$, then

$\mathcal{L}(x,\lambda,\mu)$ is convex in x for fixed $\lambda \geq 0, \mu$.

Further, $\mathcal{L}(x,\lambda,\mu)$ is affine, & hence concave in (λ,μ) for fixed x.

$$d^* = \sup_{\substack{\lambda \geq 0 \\ \mu}} \inf_{x \in \mathbb{R}^n} \mathcal{L}(x,\lambda,\mu)$$

Claim: $p^* = \inf_x \left[\sup_{\substack{\lambda \geq 0 \\ \mu}} \mathcal{L}(x,\lambda,\mu) \right] = \left[\begin{cases} f_0(x), & \text{when } h_j(x) = 0 \ \forall j, \ f_i(x) \leq 0 \ \forall i \\ \infty & \end{cases} \right]$

$$= \inf_x f_0(x) \quad \text{s.t. } f_i(x) \leq 0,\ h_j(x) = 0$$

Therefore, min-max inequality established above implies $d^* \leq p^*$, which is nothing but weak-duality.

Recall: $\|z\|_2 = \sup_{\|y\|_2 \le 1} y^T z$

SOCP Duality

$\lambda \|z\|_2 = \sup_{\|y\|_2 \le \lambda} y^T z$

Recall that a SOCP in standard form is

$$\min_{x \in \mathbb{R}^n} c^T x$$
$$\text{s.t. } \|A_i x + b_i\|_2 \le e_i^T x + f_i \quad : \lambda_i, \quad i = 1, 2 \ldots m$$

$$\mathcal{L}(x, \lambda) = c^T x + \sum_{i=1}^m \lambda_i \left[\|A_i x + b_i\|_2 - e_i^T x - f_i \right]$$

$$= c^T x + \sum_{i=1}^m \left[\sup_{\|y_i\|_2 \le \lambda_i} y_i^T (A_i x + b_i) - \lambda_i e_i^T x - \lambda_i f_i \right]$$

$$= \sup_{\substack{\|y_i\|_2 \le \lambda_i, \\ i = 1, 2 \ldots m}} \left[c^T x + \sum_{i=1}^m \left[y_i^T (A_i x + b_i) - \lambda_i e_i^T x - \lambda_i f_i \right] \right]$$

$$p^* = \inf_{x \in \mathbb{R}^n} \max_{\lambda \ge 0} \left[\mathcal{L}(x, \lambda) \right]$$

$$= \inf_{x \in \mathbb{R}^n} \sup_{\|y_i\|_2 \le \lambda_i} \left[\sum_{i=1}^m (y_i^T b_i - \lambda_i f_i) + \underbrace{\left[c^T + \sum_{i=1}^m (y_i^T A_i - \lambda_i e_i^T) \right]}_{\alpha^T} x \right]$$

$$\ge \sup_{\|y_i\|_2 \le \lambda_i} \inf_{x \in \mathbb{R}^n} \left[\quad " \quad \right]$$

is finite only when $\alpha^T = c^T + \sum_{i=1}^m (y_i^T A_i - \lambda_i e_i^T) = 0$

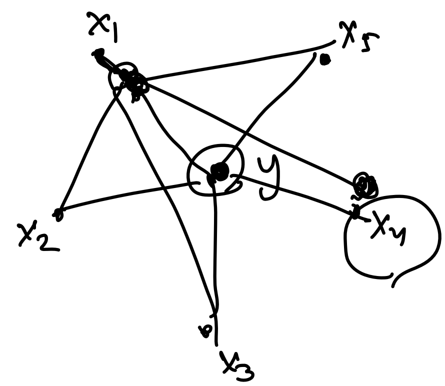
$$= \begin{bmatrix} \sup_{y, \lambda} & \sum_{i=1}^m (y_i^T b_i - \lambda_i f_i) \\ \text{s.t.} & \|y_i\|_2 \le \lambda_i, \ \forall i, \text{ second order cone constraint} \\ & \sum_{i=1}^m (A_i^T y_i - \lambda_i e_i) + c = 0 \end{bmatrix}$$

Example: Facilty Location Problem

Given a set of locations : $\mathcal{L} = (x_1, x_2 \dots x_m)$, $x_i \in \mathbb{R}^n$

For a location y,

its distance from points in $\mathcal{L}$



$$\left.\begin{array}{l} \|y - x_1\|_2 \\ \|y - x_2\|_2 \\ \vdots \\ \|y - x_n\|_2 \end{array}\right\}$$

worst-case : $\min\limits_{y} \max\limits_{1 \le i \le m} \|y - x_i\|_2 \iff \begin{array}{ll} \min\limits_{y,t} & t \\ \text{s.t} & \max\limits_{1 \le i \le m} \|y - x_i\|_2 \le t \end{array}$

$$\Updownarrow$$

$$\begin{array}{ll} \min\limits_{y,t} & t \\ \text{s.t.} & \|y - x_i\|_2 \le t, \\ & \text{for } i = 1, 2 \dots m \end{array}$$

average : $\min\limits_{y} \dfrac{1}{m} \sum\limits_{i=1}^{m} \|y - x_i\|_2$

$$\Updownarrow$$

$$\begin{array}{ll} \min\limits_{y,\, t \in \mathbb{R}^m} & \dfrac{1}{m} \sum\limits_{i=1}^{m} t_i \\ \text{s.t.} & \|y - x_i\|_2 \le t_i \;\; \forall i \end{array}$$

– Both are instances of SOCP problems.

Square-root Lasso : $\min\limits_{x} \underline{\|Ax - b\|_2} + \underline{\alpha \|x\|_1}$

Let us find its dual.

$$\min_{x} \|Ax - b\|_2 + \alpha \|x\|_1 = \min_{x} \max_{\substack{\|u\|_2 \le 1 \\ \|v\|_\infty \le \alpha}} u^T(Ax - b) + v^T x$$

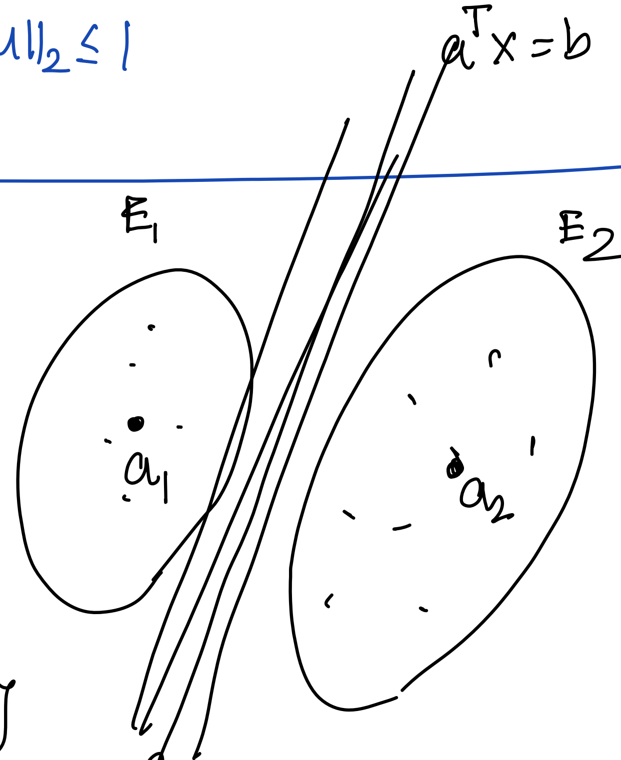
$$\ge \max_{\substack{\|u\|_2 \le 1 \\ \|v\|_\infty \le \alpha}} \min_{x} (u^T A + v^T) x - u^T b$$

(equality will hold due to Sion's theorem)

$$= \max_{\substack{\|u\|_2 \le 1 \\ \|v\|_\infty \le \alpha}} -u^T b \quad \text{s.t.} \quad u^T A + v^T = 0$$

Example: Separation of Ellipsoids

$$= \left[\begin{array}{ll} \max_u & -u^T b \\ \text{s.t.} & \|A^T u\|_\infty \le \alpha, \ \|u\|_2 \le 1 \end{array} \right.$$



Given two ellipse E_1 and E_2, find $(a \in \mathbb{R}^n, b \in \mathbb{R})$ s.t. $\{x \mid a^T x = b\}$ separates E_1 and E_2.

$$E_1 = \{x \in \mathbb{R}^n \mid x = a_1 + R_1 \rho, \ \|\rho\|_2 \le 1\}$$

$$E_2 = \{x \in \mathbb{R}^n \mid x = a_2 + R_2 \rho, \ \|\rho\|_2 \le 1\}.$$

we want (a, b) to satisfy:

if $x \in E_1$, then $a^T x \le b \Rightarrow a^T [a_1 + R_1 \rho] \le b \quad \forall \|\rho\|_2 \le 1$

if $x \in E_2$, then $a^T x \ge b$

$$\Downarrow$$

$$a^T (a_2 + R_2 \rho) \ge b \quad \forall \|\rho\|_2 \le 1$$

$$\sup_{\|\rho\|_2 \le 1} a^T a_1 + a^T R_1 \rho \le b \iff a^T a_1 + \|R_1^T a\|_2 \le b.$$

Likewise, we have $\quad a^T(a_2 + R_2 p) \geq b \quad \forall \|p\|_2 \leq 1$

$$\iff a^T a_2 + \inf_{\|p\|_2 \leq 1} a^T R_2 p \geq b$$

$$\iff a^T a_2 - \sup_{\|p\|_2 \leq 1} p^T(-a^T R_2) \geq b$$

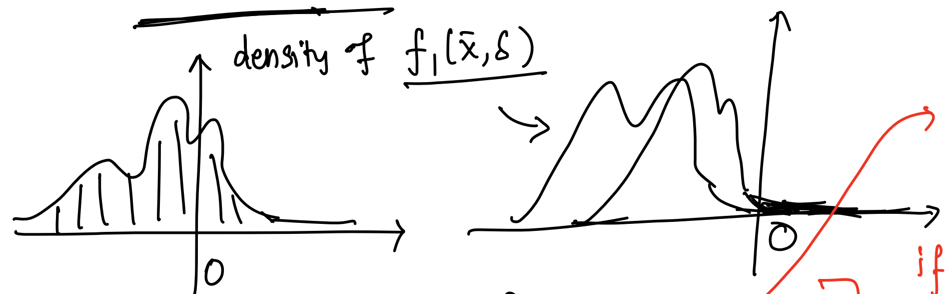
$$\iff a^T a_2 - \|R_2^T a\|_2 \geq b$$

The goal is to find (a,b) that satisfy:

$$a^T a_1 + \|R_1^T a\|_2 \leq b$$
$$a^T a_2 - \|R_2^T a\|_2 \geq b$$

Optimization under uncertainty

$$\min_x \; f_0(x)$$
$$\text{s.t. } \; f_1(x,\delta) \leq 0, \qquad \text{where } \delta \text{ is a random variable.}$$



$\bigcap_{\delta} \text{Sublev.}(f_1)(0)$

Ⓐ Robust optimization: $\quad \min_x f_0(x) \quad \text{s.t. } f_1(x,\delta) \leq 0 \quad \forall \delta$

only meaningful if δ belongs to a known bounded set

$\Rightarrow$ if $f_1(\cdot,\delta)$ is convex in (x) for fixed δ; then problem is convex

Ⓑ Chance contrained optimization:

$$\min_x \; f_0(x)$$
$$\text{s.t. } \; \mathbb{P}\left[f_1(x,\delta) \leq 0\right] \geq 1-\alpha,$$

α is a small +ve constant

In general, nonconvex and NP-Hard.
except for few special cases.

Example: Chance Constrained Optimization

$$\min_{x \in \mathbb{R}^n} \quad c^T x$$

$$\text{s.t} \quad \mathbb{P}\left[\delta_i^T x \leq b_i\right] \geq p_i, \quad i = 1, 2 \ldots m$$

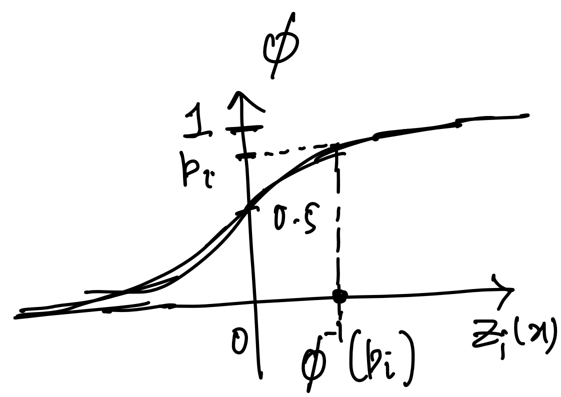
Assumption (1) $p_i > 0.5 \ \forall i$, (2) δ_i are Gaussian random vectors with mean $\bar{\delta}_i$, and covariance matrix $\Sigma_i \succ 0$

Let us examine $\mathbb{P}\left[\delta_i^T x \leq b_i\right] \geq p_i$

$\delta_i^T x$: scalar, Gaussian random variable,

with mean: $\bar{\delta}_i^T x$

variance: $x^T \Sigma_i x$



Then $$\frac{\delta_i^T x - \bar{\delta}_i^T x}{\sqrt{x^T \Sigma_i x}} \sim \mathcal{N}(0, 1)$$

$$\delta_i^T x \leq b_i \iff \underbrace{\frac{\delta_i^T x - \bar{\delta}_i^T x}{\sqrt{x^T \Sigma_i x}}}_{z_i(x)} \leq \underbrace{\frac{b_i - \bar{\delta}_i^T x}{\sqrt{x^T \Sigma_i x}}}_{\tau_i(x)}$$

$$\Rightarrow \mathbb{P}\left[\delta_i^T x \leq b_i\right] \geq p_i \iff \mathbb{P}\left[z_i(x) \leq \tau_i(x)\right] \geq p_i$$

$$\mathrm{CDF}_{\mathcal{N}(0,1)}\left(\tau_i(x)\right) \geq p_i$$

Let $\phi(x)$ be the CDF of $\mathcal{N}(0,1)$

we can equivalently write $\quad z_i(x) \geq \phi^{-1}(p_i)$.

$$\Rightarrow \quad \frac{b_i - \bar{\delta}_i^T x}{\sqrt{x^T \Sigma_i x}} \geq \phi^{-1}(p_i)$$

$$\Rightarrow \quad b_i \geq \bar{\delta}_i^T x + \phi^{-1}(p_i) \underbrace{\sqrt{x^T \Sigma_i x}}_{\|\Sigma_i^{1/2} x\|_2}$$

$$\Rightarrow \quad \bar{\delta}_i^T x \leq b_i - \phi^{-1}(p_i) \|\Sigma_i^{1/2} x\|_2$$: SOC constraint.

Robust Optimization

Robust Optimization with Box Uncertainty Set

Robust Optimization with Ellipsoidal Uncertainty Set

Robust Least Squares

Robust Classification