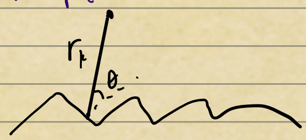
<How does it help?>

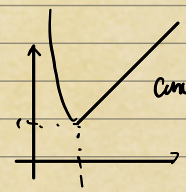



$$I_k = \sum_{j=1}^{m} a_{kj} P_j \qquad a_{kj} = r_{kj}^{-2} \max(\cos\theta_{kj}, 0)$$

problem : $$\min_{0 \le P_j \le P_{max}} \max_{k=1 \cdots n} \left| \log I_k - \log I_{des} \right|$$

$$\Longleftrightarrow \quad \min_{0 \le P_j \le P_{max}} \max_{k=1 \cdots n} h\left(\frac{I_k}{I_{des}}\right) \quad \text{where } h(u) = \max\left(u, \frac{1}{u}\right)$$

↳ not differentiable




$$\left( \operatorname{argmax}_{x>0} |\log x| = \operatorname{argmax}_{x>0} \max\left(x, \frac{1}{x}\right) \right)$$

approximation : ① $P_j = P$

② $|I_k - I_{des}|^2$ ⊕ regularization term

③ $|I_k - I_{des}|$

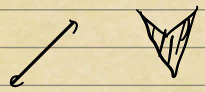
① Recognize problems as convex optimization

② Develop code to the moderate size

③ Characterize optimal solution, limits of performance

Lec 2


line $\theta x_1 + (1-\theta) x_2$ $\theta \in \mathbb{R}$



(o) (x)

affine set

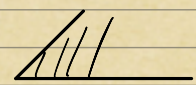
line segment $\theta x_1 + (1-\theta) x_2$ $\theta \in [0,1]$

Convex set

"Convex combination of points": $\sum_{i=1}^{n} \theta_i x_i$ where $\sum_{i=1}^{n} \theta_i = 1$ $\theta_i \geq 0$ $\forall i$

Convex hull : "conv S" Convex Combination of points in $S$

Conic combination $x = \theta_1 x_1 + \theta_2 x_2$ $\theta_1, \theta_2 \geq 0$

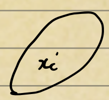


Convex cone of S : all conic combination

hyperplane $\{x : a^T x = b\}$ $a \neq 0$

half space $\{x : a^T x \leq b\}$ $a \neq 0$

ellipsoid $\{x : (x - x_c)^T P^{-1} (x - x_c) \leq 1\}$ $P \in S^n_{++}$ (ex) $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$

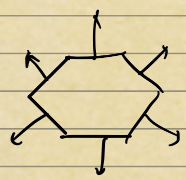



$\{x_c + Au \mid \|u\|_2 \leq 1\}$ A square non singular

polyhedra $\underset{\mathbb{R}^{m \times n}}{A} x \preceq b$ $\underset{\mathbb{R}^{p \times n}}{C} x = d$

component wise

i.e. $\underset{\text{finite}}{\bigcap}$ half space, hyperplane



$(\theta_1, \theta_2 \geq 0)$

$S^n_+$ is a convex cone (check) $A, B \in S^n_+ \Rightarrow \theta_1 A + \theta_2 B \succeq 0$ by definition

$\theta U_1 \Sigma_1 U_1^T + \theta_2 U_2 \Sigma_2 U_2^T$

$S^n_{++}$ = positive definite

## Operations that preserve convexity

- intersection
- affine function $f(x)=Ax+b$ $\quad A\in\mathbb{R}^{m\times n}$ $\quad b\in\mathbb{R}^m$ ; $\mathbb{R}^n\to\mathbb{R}^m$
- perspective function $f(x,t)=\frac{x}{t}$ , $t>0$ ; $\mathbb{R}^{n+1}\to\mathbb{R}^n$
- linear fractional function $f(x)=\frac{Ax+b}{c^Tx+d}$ $\quad c^Tx+d>0$ ; $\mathbb{R}^n\to\mathbb{R}^m$

proper cone : cone that is closed, $\exists$ interior , pointed ($x\in C, -x\in C \Rightarrow x=0$) , convex

(ex) $\mathbb{R}^n_+$ , $K=\{x\in\mathbb{R}^n : \sum_{i=1}^{n} x_i t^{i-1}\ge 0 \quad \forall t\in[0,1]\}$, $S^n_+$ (check)

## Generalized inequality

$$\begin{cases} x\preceq_K y & y-x\in K \\ x\prec_K y & y-x\in \text{int}(K) \end{cases}$$

(properties)

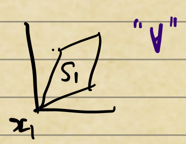
① $x\preceq_K y$, $u\preceq_K v$ $\Rightarrow$ $x+u\preceq_K y+v$

② possible to have $x\not\preceq_K y$ and $y\not\preceq_K x$

$\Updownarrow$ $\qquad$ $\Updownarrow$

$x-y\notin K$ $\qquad$ $y-x\notin K$

minimum vs. minimal





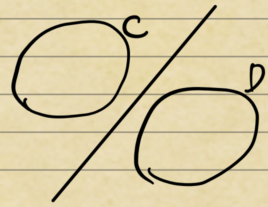
$\forall y\in S$ ; $x\preceq_K y$ $\qquad$ $y\in S$ $\quad y\preceq_K x \Rightarrow y=x$

- minimum is unique

$x\preceq y$ & $y\preceq x \Rightarrow x-y,\ y-x\in K \Rightarrow x-y=0 \Rightarrow x=y$

## · Separating hyperplane theorem

Let C, D be disjoint convex sets. There exists $a\neq 0$, $b$ such that

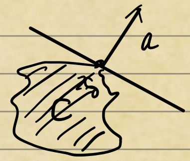



$a^T x \le b \quad \forall x \in C$

$a^T x \ge b \quad \forall x \in D$

- Supporting hyperplane to set C at bdry point $x_0$ is defined by

$\{x \mid a^T x = a^T x_0\}$



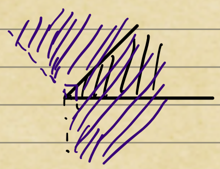

$a \neq 0$ & $a^T x \le a^T x_0 \quad \forall x \in C$

Supporting hyperplane theorem

$\exists$ Supporting hyperplane at every boundary point of convex set

dual cone of a cone $K$

$K^* = \{ y : y^T x \ge 0 \quad \forall x \in K \}$



intuition: "direction"

(Ex) $\mathbb{R}^n_+$, $S^n_+$ (check), $\{(x,t) : \|x\|_2 \le t\}$ self-dual

vs. $\{(x,t) : \|x\|_1 \le t\}$

Dual cone of proper cone is proper (define generalized inequality)

* $\lambda \succ_{K^*} 0$ $x$ minimizes $\lambda^T z$ over $z \in S$ $\longrightarrow$ $x$ minimal

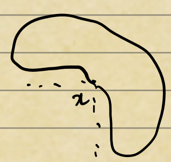
pf) Assume not. $\exists z \in K$ s.t. $z \preceq_K x$ and $z \neq x$

$\Rightarrow \lambda^T (x - z) > 0$

$\Rightarrow \lambda^T z < \lambda^T x \quad z \in S$

$\Rightarrow$ Contradiction

Converse not true




$x$ minimal, $\nexists \lambda$

* $\lambda \succ_{K^*} 0$ $x$ uniquely minimizes $\lambda^T z$ over $z \in S$ $\longleftrightarrow$ $x$ minimum

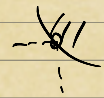
pf) $(\Rightarrow)$ proved

$(\Leftarrow)$ $\forall y \in S$, $y - x \in S$

$\Rightarrow \pi^T(y-x) > 0 \quad \forall y \in S$

$*$ $x$ minimal of convex $S \Rightarrow \exists \pi \succeq_{K^*} 0$ $\pi \neq 0$ s.t. $x$ minimizes $\pi^T z$ over $S$

pf) $x$ minimal $\big((x-K)\setminus\{x\}\big) \cap S = \emptyset$



translation

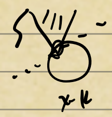
SHT

$\Rightarrow \exists \pi \neq 0$ $\mu$ s.t. $\pi^T(x-y) \leq \mu \quad \forall y \in K \cdots ①$

$\pi^T z \geq \mu \quad \forall z \in S \cdots ②$

convex

$x \in S,\ x \in x-K$ $(0 \in K)$

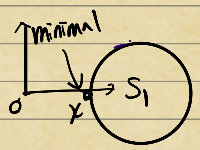



$\rightarrow \pi^T x = \mu$

$\rightsquigarrow$ ① $\Rightarrow$ $\pi \in K^*$

② $\Rightarrow$ $\pi \neq 0$ s.t. $x$ minimizes $\pi^T z$ over $S$

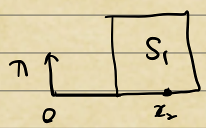
(ex)




$\pi^T z$ over $S$

only when $\pi = (1,0)$ does $\pi^T x$ minimize

But! $\pi \not\succ_{K^*} 0$




$z = x_2$ minimizes $\pi^T z$

Lec 3

Convex function $f$ Strict

① dom $f$ is convex

② $f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y)$ $<$

$\theta \in [0,1]$ $\theta \in (0,1)$

(ex) norm ( spectral norm, affine function, etc )

- dom : $\mathbb{R}$ is easy to verify
- important tool : $f : \text{dom} f \to \mathbb{R}$ is convex iff

$g : \text{dom}\, g \to \mathbb{R}$, $g(t) = f(x + tv)$ dom $g = \{ t \mid x + tv \in \text{dom} f \}$

is convex in $t$ $\forall x \in \text{dom} f$, with $v \in \mathbb{R}^n$

($\Rightarrow$) $g(\theta t_1 + (1-\theta) t_2)$

$$\overset{\parallel}{f(x + \theta t_1 v + (1-\theta) t_2 v)} \le \theta f(x + t_1 v) + (1-\theta) f(x + t_2 v)$$

$$= \theta g(t_1) + (1-\theta) g(t_2)$$

($\Leftarrow$)

$f(x + \theta t_1 v + (1-\theta) t_2 v) \le \theta f(x + t_1 v) + (1-\theta) f(x + t_2 v)$

$\forall x \in \text{dom} f$ $v \in \mathbb{R}^n$ $t$ appropriate

WTS

$\Rightarrow$ $f(\theta x_1 + (1-\theta) x_2) \le \theta f(x_1) + (1-\theta) f(x_2)$

$$\begin{cases} x_1 - x_2 = t_1 v - t_2 v \\ x_2 = x + t_2 v \end{cases}$$

$x_1 = x + t_1 v$

$x_2 = x + t_2 v$

$v = \dfrac{x_1 - x_2}{t_1 - t_2}$ $x = x_1 - \dfrac{t_1(x_1 - x_2)}{t_1 - t_2} = \dfrac{t_1 x_2 - t_2 x_1}{t_1 - t_2} \in \text{dom} f$ since dom $f$ is convex set.

(ex) $f : S^n \to \mathbb{R}$ $f(X) = \log \det X$ dom $f = S^n_{++}$

; STS $f(t) = \log \det(X + tV)$ is convex in $t$ with $X \in S^n_{++}$ $V \in S^n$

$$= \log\left[\det(X)\cdot\det\left(I+tX^{-\frac{1}{2}}VX^{-\frac{1}{2}}\right)\right]$$

$$= \log\det(X) + \log\det\left(I+tX^{-\frac{1}{2}}VX^{-\frac{1}{2}}\right)$$

$$= \log\det(X) + \log\left[\prod_i (1+t\pi_i)\right] \qquad \pi_i = \text{eigenvalue of } X^{-\frac{1}{2}}VX^{-\frac{1}{2}}$$

concave in $t$; $\dfrac{\sum_i \prod_{-i}(1+t\pi_i)}{\prod(1+t\pi_i)} = \sum_i \dfrac{1}{1+t\pi_i}$

Remark. Convenient to define $\tilde{f}$ where outside of domain, $\infty$ (instead of NaN)

1-st order condition

$$f(y) \geq f(x) + \nabla f(x)^T(y-x) \qquad \forall x, y \in \mathrm{dom} f$$

2nd order condition

$\nabla^2 f(x) \succeq 0$

If $\nabla^2 f(x) \succ 0$ then $f$ is strictly convex

(non-example) $f(x) = x^4$ $x \in \mathbb{R}$, So the converse is false

(ex) $\frac{1}{2}x^T P x + q^T x + r$

$\nabla f(x) = Px + q$

$\nabla^2 f(x) = P$

convex if $P \succeq 0$

(ex) $f(x,y) = \frac{x^2}{y}$

$\nabla f = \left(\frac{2x}{y}, -\frac{x^2}{y^2}\right)$

$$\nabla^2 f = \begin{pmatrix} \frac{2}{y} & -\frac{2x}{y^2} \\ -\frac{2x}{y^2} & \frac{2x^2}{y^3} \end{pmatrix} = \frac{2}{y^3}\begin{pmatrix} y^2 & -yx \\ -yx & x^2 \end{pmatrix} = \frac{2}{y^3}\begin{pmatrix} y \\ -x \end{pmatrix}\begin{pmatrix} y & -x \end{pmatrix} \succeq 0 \text{ for } y > 0$$

(ex) log-sum-exp $f(x) = \log\sum_{k=1}^n \exp x_k$ ("soft max" concentrated on $\max(x_i)$)

$\nabla^2 f(x) = \frac{1}{1^T z}\,\mathrm{diag}(z) - \frac{1}{(1^T z)^2} z z^T$ $(z_k = \exp x_k)$ Check

(ex) $|x|^p$, norms, etc

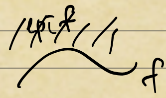
Other connection (convex set $\Leftrightarrow$ convex fn)

- Sublevel set of convex function is convex set

  $\{ x : f(x) \le \alpha \}$

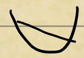
- Converse not true : $-e^{-x}$ not convex but its $\forall$ sublevels convex

- "above"
  epigraph $\quad \text{epi } f = \{ (x,t) \in \mathbb{R}^{n+1} : x \in \text{dom} f \;\; f(x) \le t \}$



★ **f is convex $\Leftrightarrow$ epi f is a convex set**

pf) $(\Rightarrow)$ $\quad f(x_1) \le t_1, f(x_2) \le t_2 \Rightarrow f(\theta x_1 + (1-\theta)x_2) \le \theta t_1 + (1-\theta) t_2$

$(\Leftarrow)$  $\quad t_1 = f(x_1)$

$\quad t_2 = f(x_2)$

- Jensen's inequality

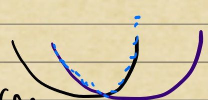
Operations that preserve convexity (usually using hessian is a last resort)

- non-negative multiple
- addition (integrals)
- $f(Ax+b)$

(ex) $f(x) = -\sum_{i=1}^{m} \log (b_i - a_i^T x)$ $\quad$ ($b_i - a_i^T x > 0 \Leftrightarrow$ interior of polyhedron)

(ex) $f(x) = \| Ax+b \|$

- pointwise maximum



pf) $\text{epi} (\max_i f_i) = \bigcap_i \text{epi}(f_i)$

(ex) $f(x) = \max_{i=1 \cdots m} (a_i^T x + b_i)$

(ex) $f(x) = x_{[1]} + x_{[2]} + \cdots + x_{[r]}$ $\quad$ r'th largest component

$= \max \{ x_{i_1} + \cdots + x_{i_r} \mid 1 \le i_1 < i_2 < \cdots < i_r \le n \}$

- pointwise supremum

  $f(x,y)$ is convex in $x$ for each $y \in A$, then

$g(x) = \sup_{y \in A} f(x,y)$ is convex

pf) $\bigcap_{y \in A} \{ (x,t) : f(x,y) \le t \} = \{ (x,t) : \sup_{y \in A} f(x,y) \le t \}$

because $\bigcap_{x \in A} \{ f(x) \le t \} = \{ \sup_{x \in A} f(x) \le t \}$

(ex) $\lambda_{max}(X) = \sup_{\|y\|_2 = 1} y^T X y$

- Composition with scalar functions

$f(x) = h(g(x))$

$g : \mathbb{R}^n \to \mathbb{R}$ $\quad h : \mathbb{R} \to \mathbb{R}$ $\quad f''(x) = \underline{h''(g(x))} (g'(x))^2 + h'(g(x)) g''(x)$

result : $f$ is convex if $g$ convex $h$ convex $\tilde{h}$ non decreasing

extended value fn

( $g$ concave $h$ convex $\tilde{h}$ non increasing )

pf) $h(g(\theta x_1 + (1-\theta) x_2)) \underset{\substack{g \text{ convex} \\ h \text{ nondecreasing}}}{\le} h(\theta g(x_1) + (1-\theta) g(x_2)) \underset{h \text{ convex}}{\le} \theta h(g(x_1)) + (1-\theta) h(g(x_2))$

Left to show $g(\theta x_1 + (1-\theta) x_2) \in \operatorname{dom} h$ ; $\quad g(x_1) \in \operatorname{dom} h \quad g(x_2) \in \operatorname{dom} h$

$\Rightarrow \theta g(x_1) + (1-\theta) g(x_2) \in \operatorname{dom} h$

$\qquad \text{VI}$

$\Rightarrow g(\theta x_1 + (1-\theta) x_2) \in \operatorname{dom} h$

(ex) $\exp g(x)$ , $\dfrac{1}{g(x)}$

- Vector composition

$f$ is convex if $\begin{cases} g \text{ convex} \quad h \text{ convex} \quad \tilde{h} \text{ non decreasing} \\ g \text{ concave} \quad h \text{ convex} \quad \tilde{h} \text{ non increasing} \end{cases}$

$g : \mathbb{R}^n \to \mathbb{R}^m$ $\quad h : \mathbb{R}^m \to \mathbb{R}$ $\quad f''(x) = g'(x)^T \nabla^2 h(g(x)) g'(x) + \nabla h(g(x))^T g''(x)$

(ex) $\sum_{i=1}^m \log g_i(x)$ , $g$ concave & positive

$h(x) = \underbrace{\sum_{i=1}^m \log x_i}_{\text{concave}}$ $\qquad \underline{h(g(x))}$ concave, concave

(ex) $\log \sum_{i=1}^m \exp(g_i(x))$ is convex if $g_i$'s are convex

Lec 4. Use of $\tilde{h}$ in vector composition

Verify $\|Ax+b\|^{1.62}$ is convex

$$h(z) = \begin{cases} z^{1.62} & \text{for } z \geq 0 \\ 0 & z < 0 \end{cases}$$

$f(x,y)$ convex $C$ convex set

$\Rightarrow$ $g(x) = \inf_{y \in C} f(x,y)$ is convex

pf) STS $\{(x,t) : t \geq \inf_{y \in C} f(x,y)\}$ is convex.

By assumption, $\{(x,y,t) : t \geq f(x,y)\}$ is convex and $C$ is convex.

$\bigcup_{y \in C} \{(x,y,t) : t \geq f(x,y)\}$ is convex.

$= \{(x,y,t) : t \geq f(x,y),\ \exists y \in C\}$

$= \{(x,y,t) : t \geq \inf_{y \in C} f(x,y)\}$

$t_1 \geq f(x_1, y_1)$ $y_1 \in C$

$\Rightarrow \theta t_1 + (1-\theta) t_2 \geq f(\theta x_1 + (1-\theta)x_2, \underbrace{\theta y_1 + (1-\theta) y_2}_{\in C})$

$t_2 \geq f(x_2, y_2)$ $y_2 \in C$

$\Rightarrow \bigcup_{y \in C} \{(x,y,t) : t \geq f(x,y)\}$ is convex.

(ex) $f(x,y) = x^T A x + 2x^T B y + y^T C y$

Condition: convex $\Leftrightarrow$ $\begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \succeq 0$ (second derivative test) $A \succ 0$

$g(x) = \inf_y f(x,y) = x^T (A - BC^{-1} B^T) x$

$g$ is convex, so $A - BC^{-1}B^T \succeq 0$

(ex) $\text{dist}(x, S) = \inf_{y \in S} \|x - y\|$ is convex if $S$ is convex

perspective of a function $f: \mathbb{R}^n \to \mathbb{R}$ is the function $g: \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$

$g(x,t) = t f(\frac{x}{t})$ $\text{dom } g = \{(x,t) \mid \frac{x}{t} \in \text{dom } f,\ t > 0\}$

$f$ convex $\Rightarrow$ $g$ convex

(ex) $f(x) = x^T x \qquad g(x,t) = \frac{x^T x}{t}, \ t>0$

(ex) $f(x) = -\log x \qquad g(x,t) = t\log t - t\log x$ on $\mathbb{R}^2_{++}$

(ex) $tf(\frac{x}{t})$ convex $\rightarrow g(x) = (c^T x + d) f((Ax+b)/(c^T x+d))$ convex on $\{x \mid c^T x + d > 0,\ (Ax+b)/(c^T x+d) \in \mathrm{dom} f\}$

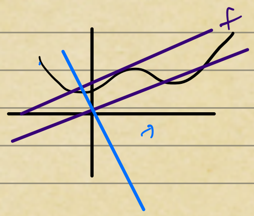
$g(A(\theta x_1 + (1-\theta)x_2) + c,\ c^T(\theta x_1 + (1-\theta)x_2) + d)$

$= g((Ax_1+b)\theta + (Ax_2+b)(1-\theta),\ (c^T x_1 + d)\theta + (c^T x_2 + d)(1-\theta))$

$\leq \theta g(Ax_1+b,\ c^T x_1 + d) + (1-\theta) g(Ax_2+b,\ c^T x_2 + d)$

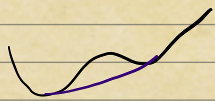
Conjugate function

$$f^*(y) = \sup_{x \in \mathrm{dom} f} (y^T x - f(x))$$

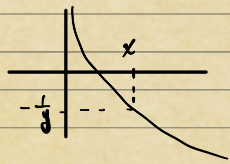



fact. $\mathrm{epi}(f^{env}) = \mathrm{conv}(\mathrm{epi} f)$

where $f^{env} = (f^*)^*$



(ex) $f(x) = -\log x$




$$f^*(y) = \sup_{x>0} (y^T x + \log x)$$

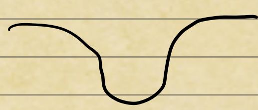
$$= \begin{cases} \infty & y \geq 0 \\ -1 - \log(-y) & y < 0 \end{cases}$$

($y + \frac{1}{x} = 0$ by differentiation)

(unimodal)

Quasiconvex $f: \mathbb{R}^n \rightarrow \mathbb{R}$

dom $f$ is convex and sublevel sets $S_\alpha = \{x \in \mathrm{dom} f : f(x) \leq \alpha\}$ are convex $\forall \alpha$

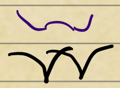


(ex) $\sqrt{|x|}$ 

(ex) $\mathrm{ceil}(x) = \inf_{y \in \mathbb{Z}} \{y \geq x\}$

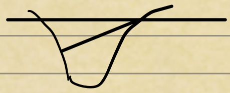
(ex) $f(x) = \frac{a^T x + b}{c^T x + d} \qquad \mathrm{dom} f = \{x \mid c^T x + d > 0\}$ quasilinear (ex) $f(x) = \frac{\|x-a\|_2}{\|x-b\|_2}$ $\quad$ $\mathrm{dom}\, f = \{ x : \|x-a\|_2 \le \|x-b\|_2 \}$ (check)

(non-ex)



**modified Jensen**

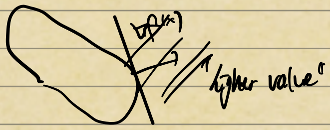
$0 \le \theta \le 1 \Rightarrow f(\theta x + (1-\theta) y) \le \max\{f(x), f(y)\}$



(quasi convex)

(first order) $\mathrm{dom}\, f$ : convex set & $\quad$ ← iff condition

$f(y) \le f(x) \Rightarrow \nabla f(x)^T (y-x) \le 0$

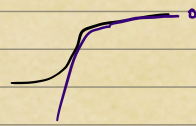



**log concave** $\quad (f \ge 0)$ ← out of domain; $-\infty$

$f(\theta x + (1-\theta) y) \ge f(x)^\theta f(y)^{1-\theta}$ for $0 \le \theta \le 1$

log convex (less common)

(ex) lots of pdf. Gaussian, cdf of Gaussian



(ex) powers, $x^a$ on $\mathbb{R}_+$ $\begin{cases} a \le 0 & \text{log-convex} \\ a \ge 0 & \text{log-concave} \end{cases}$

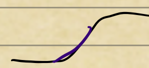
**property of log concave**

- (second derivative)

$$\nabla^2 f(x) \preceq \nabla f(x) \nabla f(x)^T$$

"allow one positive eigenvalue"

"one direction along which the curvature is positive, but not too large"



- product of log concave is log concave
- Sum of log concave is not log-concave (Mixture of Gaussian)
- log concave is quasi concave
- ★ If $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is log-concave, then

$$g(x) = \int f(x,y)\, dy \text{ is log concave}$$

(not easy to show)

- Convolution preserves log concavity (think of random variable)
- If $C \subseteq \mathbb{R}^n$ is convex and $y$ is r.v. with log-concave pdf, then

$$f(x) = \text{prob}(x+y \in C)$$

is log-concave

pf) $\int g(x+y)\, p(y)\, dy$ $\qquad g(u) = \begin{cases} 1 & u \in C \\ 0 & u \notin C \end{cases}$

pdf of $y$

(ex)  $p(x+w)$ is log concave

→ can optimize it

- Convexity w.r.t. generalized inequalities

$f : \mathbb{R}^n \to \mathbb{R}^m$ is $K$-convex if $\text{dom} f$ is convex and

$$f(\theta x + (1-\theta) y) \preceq_K \theta f(x) + (1-\theta) f(y) \qquad \forall x, y \in \text{dom} f \quad 0 \le \theta \le 1$$

(ex) $f : S^m \to S^m$ $\quad f(X) = X^2$ is $S^m_+$-convex

pf) $\|Xz\|^2$ is convex in $X$

$$\|\theta Xz + (1-\theta) Yz\|^2 \le \theta \|Xz\|^2 + (1-\theta) \|Yz\|^2$$

$$z^T (\theta X + (1-\theta) Y)^2 z \le z^T (\theta X^2 + (1-\theta) Y^2) z \qquad \forall X, Y \in S^m \quad 0 \le \theta \le 1$$

$$\to (\theta X + (1-\theta) Y)^2 \preceq_{S^m_+} \theta X^2 + (1-\theta) Y^2$$

Lec 5 minimize $f_0(x)$

subject to $f_i(x) \le 0 \quad i=1\cdots m$

$h_i(x)=0 \quad i=1\cdots p$

convention: if infeasible, min. is $\infty$

called $p^*$

$x$ is feasible if $x \in \mathrm{dom} f_0$ and achieves the constraints (implicitly, should be in the domain of them too)

$x$ is optimal if $f_0(x)=p^*$ $X_{opt}=\{$ set of optimal pts $\}$

$x$ is locally optimal if $\exists R$ such that $x$ is optimal for

minimize $f_0(z)$

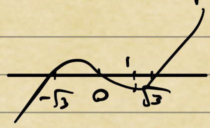
subject to $f_i(z) \le 0 \quad i=1\cdots m$

$h_i(z)=0 \quad i=1\cdots p$

$\|z-x\|_2 \le R$

(ex) $f_0(x)=\frac{1}{x}$, $x>0 \Rightarrow$ no optimal point

(ex) $f(x)=x^3-3x$



local optimum at $x=1$

$p^*=-\infty$

· implicit constraint: $x \in [\bigcap_{i=0}^{m} \mathrm{dom} f_i] \cap [\bigcap_{i=1}^{p} \mathrm{dom}\, h_i]$

explicit constraint $f_i(x) \le 0$, $h_i(x)=0$

"unconstraint" if $m=p=0$

(ex) minimize $f_0(x)=-\sum_{i=1}^{k} \log(b_i-a_i^T x)$ unconstraint, implicit : $a_i^T x < b_i$

Convex optimization problem

min. $f_0(x)$ ← convex

subject to $f_i(x) \le 0 \quad i=1\cdots m$

s.t. $a_i^T x = b_i \quad i=1\cdots p$

$f_0(x)$ quasiconvex: quasiconvex problem (while $f_1 \cdots f_m$ still convex)

prop. feasible set is convex

* This does not characterize convex problem (ex) $\|Ax-b\|=0$

· Any locally optimal point of a convex problem is global optimum

pf) Suppose $x$ locally optimal.

$x$ feasible, $\exists R$ $\forall \|x-z\|^2 \le R$, $z$ feasible, $f_0(x) \le f_0(z)$

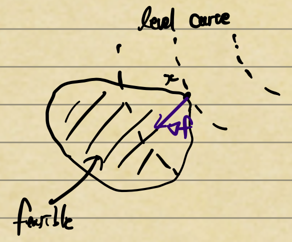
· Say $\exists y$ such that $y$ feasible, $f(y) < f(x)$

$z = \theta x + (1-\theta) y$ is feasible

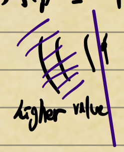
$\theta = \frac{R}{2\|y-x\|_2}$ ; $\|y-x\|_2 > R$ from assumption so $0 < \theta < \frac{1}{2}$

$f_0(z) = f_0(\theta x + (1-\theta) y) < \theta f_0(x) + (1-\theta) f_0(y) < f_0(x)$

· $x$ is optimal if and only if it is feasible and $\nabla f_0(x)^T (y-x) \geq 0$ $\forall$ feasible $y$




pf) $(\Leftarrow)$ $f(y) \geq f(x) + \nabla f(x)^T (y-x)$



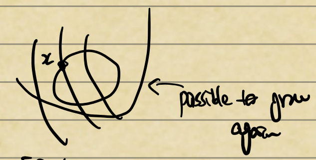

$(\Rightarrow)$ Say $\exists y$ $\nabla f_0(x)^T (y-x) < 0$

$z(t) = ty + (1-t)x$

claim: $f_0(z(t)) - f(x) < 0$ for some $t \in [0,1]$

$\nabla f_0(z(t))^T z'(t)\big|_{t=0} = \nabla f_0(x)^T (y-x) < 0$

Done.




$\nabla f$ defines a supporting hyperplane of a feasible set
$(\neq 0)$

implication · $\nabla f_0(x) = 0$ (optimal) if unconstrained

· equality constrained problem

min. $f_0(x)$ subject to $Ax = b$

$\Updownarrow$

$\exists \nu$ s.t. $x \in \mathrm{dom}\, f_0$, $Ax = b$, $\nabla f_0(x) + A^T \nu = 0$

pf) $\nabla f_0(x)^T (z-x) \geq 0$ $\forall z$, $Az = b$

· Since $Ax = b$, $z - x \in \mathcal{N}(A)$

$\Leftrightarrow \nabla f_0(x) \in \mathcal{N}(A)$

$\Leftrightarrow \nabla f_0(x) \in \mathcal{R}(A^T)$

$\Rightarrow \nabla f_0(x) = A^T u$ $\exists u$

$\Leftrightarrow \nabla f_0(x) + A^T \nu = 0$ $\exists \nu$

· minimization over non negative orthant

$x$ is optimal $\Leftrightarrow$ $x \in \mathrm{dom}\, f_0$, $x \succeq 0$, $\nabla f_0(x)_i \geq 0$ $\quad i = 0$

min $f_0$ s.t. $x \succeq 0$ $\qquad \nabla f_0(x)_i = 0$ $\quad i > 0$

$\nabla f(x)^T (z-x) \geq 0 \quad \forall z \succeq 0$

$\Rightarrow$ $\nabla f(x)^T x \leq 0$

$\nabla f(x) \succeq 0$

$\Rightarrow \nabla f(x)_i \, x_i = 0$ "complementarity"

## equivalent problems

(elimination)

$$\min f_0(x) \quad f_i(x) \leq 0 \quad i=1 \cdots m \quad Ax=b \iff f_0(Fz+x_0) \quad f_i(Fz+x_0) \leq 0 \quad i=1 \cdots m$$

where $Ax=b \iff x = Fz+x_0 \quad \exists z$

(linearization)

$$\min f_0(A_0x+b_0) \quad \text{s.t. } f_i(A_ix+b_i) \leq 0 \quad i=1 \cdots m \iff \min f_0(y_0) \quad \text{s.t. } f_i(y_i) \leq 0 \quad i=1 \cdots m, \quad y_i = A_ix+b_i \quad i=0 \cdots m$$

(slack variables)

$$\min f_0(x) \quad \text{s.t. } a_i^T x \leq b_i \quad i=1 \cdots m \iff \min_{x,s} f_0(x) \quad \text{s.t. } a_i^T x + s_i = b_i \quad i=1 \cdots m, \quad s_i \geq 0 \quad i=1 \cdots m$$

(Epigraph trick) "linear objective is universal"

$$\min_{x,t} \; t$$

s.t. $f_0(x) - t \leq 0 \quad (f_0(x) \leq t)$

$f_i(x) \leq 0 \quad i=1 \cdots m$

$Ax = b$

(minimizing over some variables)

$\min f(x_1, x_2)$

s.t. $f_i(x_1) \leq 0 \quad i=1 \cdots m$

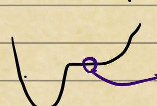
$\iff$

$\min \tilde{f}_0(x_1)$

s.t. $f_i(x_1) \leq 0 \quad i=1 \cdots m$

where $\tilde{f}_0(x_1) = \inf_{x_2} f_0(x_1, x_2)$

Quasi-convex optimization



locally optimal but not global optimal

* important property of quasi convex

$f_0$ quasi convex $\exists\ \phi_t$ such that

- $\phi_t(x)$ convex in $x$ for fixed $t$
- $t$-level set of $f_0$ is 0-level set of $\phi_t$

$$f_0(x) \le t \iff \phi_t(x) \le 0$$

pf) $\mathrm{dist}(x, \{z : f_0(z) \le t\}) \le 0 \iff f_0(x) \le t$

(ex) $f_0(x) = \dfrac{p(x)}{q(x)}$ ← convex (numerator), ← concave (denominator) with $p(x) \ge 0$ $q(x) > 0$ $\forall x \in \mathrm{dom}\, f_0$

pf) $f_0(x) \le t \iff \underbrace{p(x) - t\,q(x)}_{\text{convex}} \le 0$

$(t < 0 : \phi)$

So we can take $\phi_t(x) = p(x) - t\,q(x)$

How do we solve quasi convex problem?

optimal value $\le t$ ?

fixed $t$, convex feasibility problem given by

$$\phi_t(x) \le 0 \quad f_i(x) \le 0 \quad i = 1 \cdots m \quad Ax = b \quad \cdots ①$$

$$\begin{cases} \text{feasible} \Rightarrow t \ge p^* \\ \text{infeasible} \Rightarrow t \le p^* \end{cases}$$

[Algorithm] Given $l \le p^*$, $u \ge p^*$, tolerance $\varepsilon > 0$

repeat

1) $t := (l+u)/2$

2) Solve ①

3) if feasible $u := t$ else $l := t$

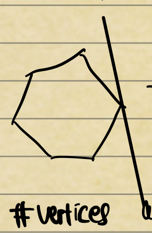
until $u - l \le \varepsilon$

Linear programming (LP) : everything is linear

min. $c^T x + d$

s.t. $Gx \preceq h$

$Ax = b$



feasible set is a polyhedron

# vertices astronomical

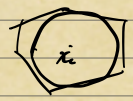
(ex) min. $\max_{i=1\cdots m} (a_i^T x + b_i)$

$\Leftrightarrow$ min. $t$

s.t. $a_i^T x + b_i \le t \quad i = 1 \cdots m$

(ex) Chebyshev center

$a_i x \le b_i \quad \forall x \in B = \{x_c + u \mid \|u\|_2 \le r\}$




$\Updownarrow$

$\sup_i \{a_i^T (x_c + u) \mid \|u\| \le r\} \le b_i$

$= a_i^T x_c + r\|a_i\|_2$

So $\begin{cases} \max \ r \\ \text{s.t.} \quad a_i^T x_c + r\|a_i\|_2 \le b_i \quad i = 1 \cdots m \end{cases}$

$\Rightarrow$ linear programming in $r, x_c$

Lec 6

- linear-fractional program (quasiconvex)

min. $f_0(x) = \dfrac{c^T x + d}{c^T x + f}$ $\quad$ $\mathbf{dom}\, f_0(x) = \{x \mid c^T x + f > 0\}$

subject to $Gx \preceq h$

s.t. $Ax = b$

$\Rightarrow$ can be solved by bisection

but also...

$\Longleftrightarrow$ LP : min. $c^T y + dz$

subject to $Gy \preceq hz$

$Ay = bz$

$\begin{cases} c^T y + fz = 1 \\ z \geq 0 \end{cases}$ (check)

#### 4.3.2 Linear-fractional programming

The problem of minimizing a ratio of affine functions over a polyhedron is called a *linear-fractional program*:

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & Gx \preceq h \\ & Ax = b \end{array} \tag{4.32}$$

where the objective function is given by

$$f_0(x) = \frac{c^T x + d}{e^T x + f}, \qquad \mathbf{dom}\, f_0 = \{x \mid e^T x + f > 0\}.$$

The objective function is quasiconvex (in fact, quasilinear) so linear-fractional programs are quasiconvex optimization problems.

**Transforming to a linear program**

If the feasible set

$$\{x \mid Gx \preceq h,\ Ax = b,\ e^T x + f > 0\}$$

is nonempty, the linear-fractional program (4.32) can be transformed to an equivalent linear program

$$\begin{array}{ll} \text{minimize} & c^T y + dz \\ \text{subject to} & Gy - hz \preceq 0 \\ & Ay - bz = 0 \\ & e^T y + fz = 1 \\ & z \geq 0 \end{array} \tag{4.33}$$

with variables $y$, $z$.

To show the equivalence, we first note that if $x$ is feasible in (4.32) then the pair

$$y = \frac{x}{e^T x + f}, \qquad z = \frac{1}{e^T x + f}$$

is feasible in (4.33), with the same objective value $c^T y + dz = f_0(x)$. It follows that the optimal value of (4.32) is greater than or equal to the optimal value of (4.33).

Conversely, if $(y, z)$ is feasible in (4.33), with $z \neq 0$, then $x = y/z$ is feasible in (4.32), with the same objective value $f_0(x) = c^T y + dz$. If $(y, z)$ is feasible in (4.33) with $z = 0$, and $x_0$ is feasible for (4.32), then $x = x_0 + ty$ is feasible in (4.32) for all $t \geq 0$. Moreover, $\lim_{t\to\infty} f_0(x_0 + ty) = c^T y + dz$, so we can find feasible points in (4.32) with objective values arbitrarily close to the objective value of $(y, z)$. We conclude that the optimal value of (4.32) is less than or equal to the optimal value of (4.33).

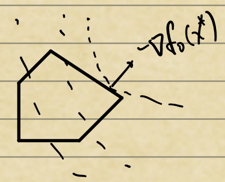
Generalized linear-fractional problem $\xrightarrow{\text{application}}$ Von Neumann

Quadratic Programming (QP) $\quad$ if not, NP hard.

min. $\frac{1}{2} x^T P x + q^T x + r$ $\quad$ $(P \in S_+^n)$

subject to $Gx \preceq h$

$Ax = b$




(ex) min. $\|Ax - b\|^2$

can add linear constraints $l \preceq x \preceq u$ ; still quadratic!

(ex) min. $\bar{c}^T x + \gamma x^T \Sigma x = E\, c^T x + \gamma \operatorname{Var}(c^T x)$

$c$ random vector w. mean $\bar{c}$ and variance $\Sigma$

QCQP (quadratically constrained quadratic program)

min. $\frac{1}{2} x^T P_0 x + q_0^T x + r_0$ $\quad \Leftrightarrow \quad \frac{1}{2} \| P_0^{\frac{1}{2}} x + P_0^{-\frac{1}{2}} q_0 \|^2 + r_0 - \frac{1}{2} \| P_0^{-\frac{1}{2}} q_0 \|^2$

subject to $\frac{1}{2} x^T P_i x + q_i^T x + r_i \leq 0$ $\quad i = 1 \cdots m$

s.t. $Ax = b$

Rmk) $P_0 \in \mathbb{R}^{3\times3}$ rank 2 ;  "cylinder"

## Second-order cone programming (SOCP)

min. $f^T x$

subject to $\|A_i x + b_i\| \le c_i^T x + d_i$, $i = 1 \cdots m$ $\qquad A_i \in \mathbb{R}^{n_i \times n}$ $\qquad \Leftrightarrow (A_i x + b_i,\ c_i^T x + d_i) \in$ second order cone in $\mathbb{R}^{n_i+1}$

$Fx = g$ $\qquad F \in \mathbb{R}^{p \times n}$

[ $n_i = 0$ : LP
$c_i = 0$ : QCQP — so more general

↳ Complete the square

Robust linear programming: two approaches; Conservative
$$\begin{cases} \text{min. } c^T x \\ \text{sub.to } a_i^T x \le b_i \quad \forall a_i \in \mathcal{E}_i \quad i = 1 \cdots m \end{cases} \cdots ①$$

$$\begin{cases} \text{min. } c^T x \\ \text{sub.to } \text{prob}(a_i^T x \le b_i) \ge \eta \quad i = 1 \cdots m \end{cases}$$ ($a_i$ : random variable) $\cdots ②$

Solved via SOCP ① ⇒ take $\mathcal{E}_i = \{ \bar{a}_i + P_i u : \|u\| \le 1 \}$

② ⇒ $a_i \sim N(\bar{a}_i, \Sigma_i)$

## Geometric programming → not convex problem (comes up all the time!)

min. $f_0(x)$

sub. to. $f_i(x) \le 1$ $\quad i = 1 \cdots m$

$h_i(x) = 1$ $\quad i = 1 \cdots p$

w. $f_i$ posynomial, $h_i$ monomial

positive sum of monomials ↲ ↳ $\text{dom} f_i = \mathbb{R}^n_{++}$, $c x_1^{a_1} \cdots x_n^{a_n}$, $c > 0$

⇓ change of variable $y_i = \log x_i$

Convex Problem
$$\begin{cases} \log \text{sumexp} \\ \log \text{sumexp} \\ Gy + d = 0 \end{cases}$$

(ex) Cantilever beam

Lecture 7

Generalized inequalities

min. $f_0(x)$

sub. to. $f_i(x) \preceq_{K_i} 0 \quad i=1\cdots m$

$Ax=b$

$\downarrow$ special case

Conic form min. $c^T x$ $\Rightarrow$ analogue of LP ($K=\mathbb{R}_+$)

$Fx+g \preceq_K 0$

$Ax=b$

$\downarrow$ special case

Semidefinite programming (SDP)

min. $c^T x$

sub. to. $x_1F_1+x_2F_2+\cdots+x_nF_n+G \preceq 0$ LMI (can be combined into one whole)

$Ax=b$

"SDP embedding"

roughly speaking, (diagonal is positive)

· LP is SDP

· SOCP is SDP : use the fact that for $\begin{pmatrix} A & B \\ B^T & C \end{pmatrix}$, when $A \succ 0$, $S \succeq 0$ iff $X \succeq 0$

SOCP: minimize $f^T x$
subject to $\|A_i x + b_i\|_2 \le c_i^T x + d_i, \quad i=1,\ldots,m$

SDP: minimize $f^T x$
subject to $\begin{bmatrix} (c_i^T x + d_i)I & A_i x + b_i \\ (A_i x + b_i)^T & c_i^T x + d_i \end{bmatrix} \succeq 0, \quad i=1,\ldots,m$

(ex) eigenvalue minimization

min. $\lambda_{max}(A(x))$

where $A(x) = A_0 + x_1A_1 + \cdots + x_nA_n$ ($A_i \in S^k$)

$\Updownarrow$

SDP min. $t$

sub. to. $A(x) \preceq tI$ (affine in $x$ and $t$)

b.c. $\lambda_{max}(A) \le t \iff A \preceq tI$

(ex) Matrix norm minimization

min. $\|A(x)\|_2$

$\Updownarrow$

min. $t$

sub. to. $\begin{bmatrix} tI & A(x) \\ A(x)^T & tI \end{bmatrix} \succeq 0$

< Cone Solver >

$\min \; c^T x$

sub. to. $a_0^j + \sum x_i a_i^j \preceq_{K_i} 0$ $\quad K_i : \mathbb{R}^+,\ SOC(k),\ LMI(k),\ \vdots$

$F(x) = g$

Vector optimization

$$\min.\ (\text{w.r.t. } K) \quad f_0(x)$$
$$\text{sub. to.} \quad f_i(x) \le 0 \quad i = 1 \cdots m$$
$$h_i(x) \le 0 \quad i = 1 \cdots p$$

What does minimizing mean here? In what sense?

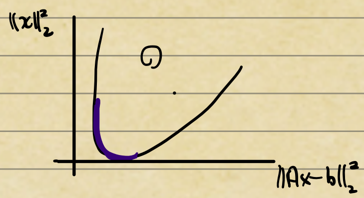
minimal vs. minimum

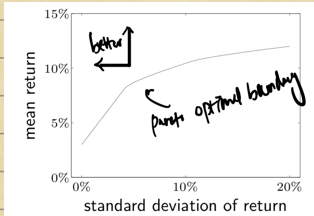
(pareto optimal)

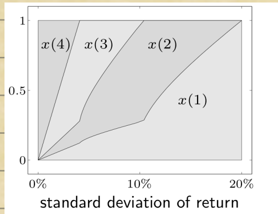
↓ special case

Multi criterion problem : Vector opt. problem with $K = \mathbb{R}^q_+$

(ex) $\min.\ (\text{w.r.t. } \mathbb{R}^2_+)\ \left( \|Ax-b\|_2^2,\ \|x\|_2^2 \right)$




⟶ diversifying portfolio

Scalarization

Fact: $x \succeq_K y$ iff $\lambda^T x \ge_{K^*} \lambda^T y \quad \forall \lambda \succeq_{K^*} 0$

Choose $\lambda \succ_{K^*} 0$. "weight vector"

$$\begin{cases} \min \; \lambda^T f_0(x) \\ \text{sub. to. } f_i(x) \le 0 \quad i = 1 \cdots m \\ h_i(x) \le 0 \quad i = 1 \cdots p \end{cases}$$

pf) Say $x$ is not pareto optimal.

$\exists y$ s.t. $f_0(y) \preceq_K f_0(x)$ & $f_0(x) \ne f_0(y)$

$f_0(x) - f_0(y) \succeq_K 0$
$\ne 0$

$\Rightarrow \quad \pi^T (f_0(x) - f_0(y)) \succ 0 \qquad \dot{x}$

(ex) $\quad \max \quad p^T x - \gamma x^T \Sigma x$

$\qquad \mathbf{1}^T x = 1 \qquad x \succeq 0$

"risk adjusted return"

Lec 8

min. $f_0(x)$

sub. to. $f_i(x) \leq 0 \quad i = 1 \cdots m$

$h_i(x) = 0 \quad i = 1 \cdots p$

dual variables

Lagrange multipliers

Lagrangian : $L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum \nu_i h_i(x)$

Lagrange dual function: $g(\lambda, \nu) = \inf_{x \in D} \left\{ f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum \nu_i h_i(x) \right\}$

concave since pointwise infimum of affine fn

$\lambda \succeq 0 \rightarrow g(\lambda, \nu) \leq p^*$

(ex) min. $x^T x$

sub. to. $Ax = b$

$L(x, \nu) = x^T x + \nu^T (Ax - b)$

Convex quadratic $\Rightarrow \nabla_x L(x, \nu) = 2x - A^T \nu = 0$

$\Rightarrow x = -\frac{1}{2} A^T \nu$

$g(\nu) = \frac{1}{4} \nu^T A A^T \nu + \nu^T \left( -\frac{1}{2} A A^T \nu - b \right)$

$= -\frac{1}{4} \nu^T A A^T \nu - b^T \nu$ ; concave fn of $\nu$

So... $-\frac{1}{4} \nu^T A A^T \nu - b^T \nu \leq p^*$

(ex) min $\|x\|$

sub. to. $Ax = b$

$\inf_x (\|x\| - y^T x)$

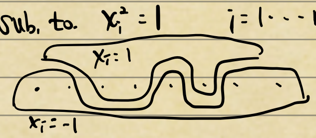
$= \begin{cases} -\infty & \|y\|_* > 1 \\ 0 & \|y\|_* \leq 1 \end{cases}$

$\|x\| + \nu^T (b - Ax) = \begin{cases} -\infty & \|A^T \nu\|_* > 1 \\ +\nu^T b & \|A^T \nu\|_* \leq 1 \end{cases}$

$\Rightarrow \quad \|A^T \nu\|_* \leq 1 \Rightarrow b^T \nu \leq p^*$

(ex) min. $x^T W x = \sum_{i,j} x_i x_j W_{ij}$ measure of how much i hates j

sub. to. $x_i^2 = 1 \quad i = 1 \cdots m$




"two way partition"

$g(\nu) = \inf_x \left( x^T W x + \sum \nu_i (x_i^2 - 1) \right)$

$= \inf_x \; x^T (W + \text{diag}(\nu)) x - 1^T \nu$

$= \begin{cases} -1^T \nu & \text{if } W + \text{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise} \quad (\text{at least one negative eigenvalue}) \end{cases}$

So if $W + \text{diag}(\nu)$, $-1^T \nu \leq p^*$ not obvious at all!

Lagrange dual and conjugate function

min. $f_0(x)$

sub. to. $Ax \preceq b$, $Cx = d$

$g(\lambda, \nu) = \inf_{x \in D} [f_0(x) + \lambda^T (Ax - b) + \nu^T (Cx - d)]$

$= -\sup_x [-f_0(x) - \lambda^T (Ax - b) - \nu^T (Cx - d)]$ recall: $f^*(\lambda) = \sup_x (\lambda^T x - f(x))$

$= -\sup_x [-f_0(x) + (-\lambda^T A - \nu^T C) x] - \lambda^T b - \nu^T d$

$= -f_0^*(-A^T \lambda - C^T \nu) - b^T \lambda - d^T \nu$

(ex) maximizing entropy problem: can be very sophisticated!

Lagrange dual problem

often. pull the implicit constraint

max $g(\lambda, \nu)$ → convex optimization problem

sub. to. $\lambda \succeq 0$

(ex) min. $c^T x$ $\xrightarrow{\text{dual}}$ max $g(\lambda, \nu) = \begin{cases} -b^T \nu & A^T \nu + c - \lambda = 0 \\ -\infty & \end{cases}$

sub. to. $Ax = b$

$x \succeq 0$

sub. to. $\lambda \succeq 0$

$\Updownarrow$

max $-b^T \nu$

sub. to. $A^T \nu + c \succeq 0$

Weak duality $d^* \leq p^*$

Strong duality $d^* = p^*$

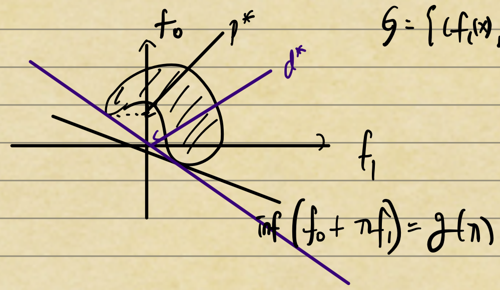
Conditions that guarantee for convex problems are called constraint qualifications

Some non-convex problem satisfies strong duality (in appendix)

Slater's Constraint qualification (covers most engineering problems)

- $\exists x \in \text{int } D$ $(=)$ $f_i(x) < 0$ $i = 1 \cdots m$ $Ax = b$ : feasible set has non empty interior

geometric interpretation




$\mathcal{G} = \{(f_1(x), f_0(x)) \mid x \in D\}$

$p^* - d^* =$ optimal duality gap

roughly speaking, $\mathcal{G}$ is convex so $p^* = d^*$

(lower left corner)

$\mathcal{A} = \{(u, t) \mid f_1(x) \le u, f_0(x) \le t$ for some $x \in D\}$ is interesting for strong-duality


Complementary Slackness

(connecting $x^*$ and $(\lambda^*, \nu^*)$)

$$f(x^*) = g(\lambda^*, \nu^*) = \inf_x \left( f_0(x) + \sum_i \lambda_i^* f_i(x) + \sum_i \nu_i^* h_i(x) \right)$$

$$\le f_0(x^*) + \sum_i \lambda_i^* f_i(x^*) + \underline{\sum_i \nu_i^* h_i(x^*)} = 0$$

$$\le f_0(x^*)$$

① $x^*$ minimizes $f_0(x) + \sum_i \lambda_i^* f_i(x) + \sum_i \nu_i^* h_i(x)$

② $\lambda_i^* f_i(x^*) = 0$ $\forall i$

$(=)$ $\begin{cases} \lambda_i^* > 0 \Rightarrow f_i(x^*) = 0 \\ f_i(x^*) < 0 \Rightarrow \lambda_i^* = 0 \end{cases}$ $\forall i$

Lec 9

① Strong duality $\longrightarrow$ KKT

(for any problem)

$$\begin{cases} f_i(x) \le 0 \quad i=1\cdots m \qquad h_i(x)=0 \quad i=1\cdots p \\ \lambda \succeq 0 \\ \lambda_i f_i(x) = 0 \quad \forall i=1\cdots m \\ \nabla_x L = \nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0 \end{cases}$$

② Convex problem & KKT $(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) \longrightarrow \tilde{x} = x^*, (\tilde{\lambda}, \tilde{\nu}) = (\lambda^*, \nu^*)$

from complementary slackness, $L(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) = f_0(\tilde{x})$ $\cdots$ ③

Since $L$ is convex and $\nabla_x L(x) = 0$, $g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$ $\cdots$ ②, ④

①: "qualified"

$\Rightarrow f_0(\tilde{x}) = g(\tilde{\lambda}, \tilde{\nu})$

If Slater condition is satisfied, (for convex problem)

$x$ is optimal iff $\exists \lambda, \nu$ that, together with $x$, satisfy KKT conditions.

(ex) min. $-\sum_{i=1}^{n} \log(x_i + \alpha_i)$

sub. to. $x \succeq 0 \quad 1^T x = 1$ (intuition: allocate onto small $\alpha_i$'s)

Slater holds by $x_i = \frac{1}{n}$, so

$x$ is optimal iff $\exists \lambda, \nu$ s.t.

① $-x \preceq 0$, $1^T x = 1$

② $\lambda \succeq 0$

③ $\lambda_i x_i = 0 \quad \forall i$

④ $\frac{-1}{x_i + \alpha_i} - \lambda_i + \nu = 0 \quad \forall i$

$\frac{1}{x_i + \alpha_i} + \lambda_i = \nu \quad \forall i$

if $\nu < \frac{1}{\alpha_i} \Rightarrow$ if $\lambda_i \ne 0 \Rightarrow x_i = 0 \Rightarrow \nu = \frac{1}{\alpha_i} + \overset{\oplus}{\lambda_i} < \frac{1}{\alpha_i}$ contradiction

So $\lambda_i = 0$ and $x_i = \frac{1}{\nu} - \alpha_i$

if $\nu \ge \frac{1}{\alpha_i} \Rightarrow$ if $x_i \ne 0 \Rightarrow \lambda_i = 0 \Rightarrow \frac{1}{x_i + \alpha_i} = \nu \ge \frac{1}{\alpha_i}$ contradiction

so $x_i = 0$ and $\lambda_i = \nu - \frac{1}{\alpha_i}$

In summary, $1 = 1^T x = \underline{\sum_{i=1}^{n} \max\{0, \frac{1}{\nu} - \alpha_i\}}$

$\rightsquigarrow$ classical but not very practical

decreasing in $\nu$, so we can determine $\nu$

perturbation and sensitivity analysis (should always be done)

min. $f_0(x)$

sub. to. $f_i(x) \le 0 \quad i=1 \cdots m$

$h_i(x) = 0 \quad i = 1 \cdots p$

max $g(\pi, \nu)$

s.t. $\pi \succeq 0$

View as a function of $u_i$ and $v_i$

w. $\begin{cases} f_i(x) \le u_i \\ h_i(x) \le v_i \end{cases}$

max. $g(\pi, \nu) - u^T \pi - v^T \pi$

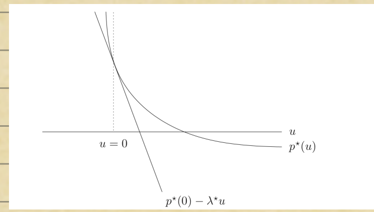
s.t. $\pi \succeq 0$

We are interested in $p^*(u, v)$ caveat: changing the resource, we reoptimize each time.

Say $\pi^*, \nu^*$ dual optimal for unperturbed:

global: $p^*(u,v) \ge g(\pi^*, \nu^*) - u^T \pi^* - v^T \nu^*$

$= p^*(0,0) - u^T \pi^* - v^T \nu^*$




local sensitivity: $\pi_i^* = - \dfrac{\partial p^*(0,0)}{\partial u_i}$ $\quad \nu_i^* = - \dfrac{\partial p^*(0,0)}{\partial v_i}$

(require differentiability)

$\pi^*$

(ex) $f_1(x^*) = 0$ 0.001 (can wiggle)

$f_2(x^*) = 0$ 10 (very sensitive)

$f_3(x^*) = 0$ 0.02

- Reformulations (when the dual is uninteresting)

$P \to D$

$\updownarrow$ not trivial connection

$\tilde{P} \to \tilde{D}$

(ex) min. $f_0(Ax+b)$

strong duality holds ; but useless

min. $f_0(y)$

s.t. $Ax + b - y = 0$

$\Longleftrightarrow$

max $-f_0^*(\nu) + b^T \nu$

s.t. $A^T \nu = 0$

$g(\nu) = \inf_{x,y} f_0(y) + \nu^T (Ax + b - y)$

$= \begin{cases} -f_0^*(\nu) + b^T \nu & A^T \nu = 0 \\ -\infty & \text{o.w.} \end{cases}$

(ex) min $\|y\|$

s.t. $y = Ax - b$

$\Longleftrightarrow$

max. $b^T \nu$

s.t. $A^T \nu = 0 \quad \|\nu\|_* \le 1$

$$g(\nu) = \inf_{x,y} \|y\| + \nu^T(y - Ax + b)$$

$$= \begin{cases} b^T y + \inf_y(\|y\| + \nu^T y) & A^T\nu = 0 \\ -\infty & \text{o.w.} \end{cases}$$

$$= \begin{cases} b^T y & \|\nu\|_* \le 1, \ A^T\nu = 0 \\ -\infty & \text{o.w.} \end{cases} \quad \left( \|\nu\|_* = \sup_{\|y\|=1} |\nu^T y| \le 1 \right)$$

impl: duality for feasibility problems, theorems of the alternative (skipped)

Generalized inequalities : same

Lec 10

Application

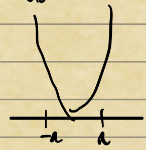
min. $\phi(r_1) + \cdots + \phi(r_m)$

sub. to. $r = Ax - b \qquad A \in \mathbb{R}^{m \times n}$

$\phi(u) = u^2$ quadratic

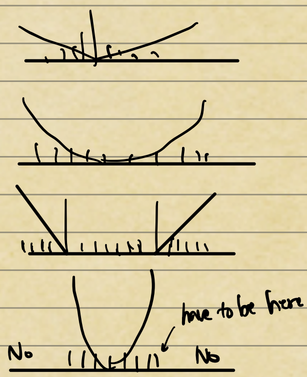
$\phi(u) = \min(0, |u| - a)$ deadzone-linear width $a$

$$\phi(u) = \begin{cases} -a^2 \log\left(1 - \left(\frac{u}{a}\right)^2\right) & |u| < a \\ \infty & \text{o.w.} \end{cases}$$

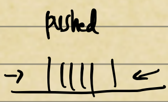


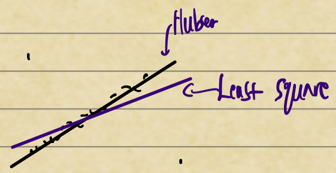

for small $u$, very close to quadratic

* shape of penalty has huge effect on the distribution of residuals




$L^\infty$ norm:




$$\phi_{hub}(u) = \begin{cases} u^2 & |u| \le M \\ M(2|u| - M) & |u| > M \end{cases}$$

$\Rightarrow$ Robust

min. $\|x\|$

sub. to. $Ax = b$

$L_2$ solution: solved via duality $\quad \|x\|^2 + \nu^T(Ax - b)$

$$\Rightarrow 2x + A^T \nu = 0$$

$$x = -\frac{1}{2} A^T \nu$$

$L_1$ solution: LP $\quad$ min. $1^T y$

sub. to. $-y \le x \le y$

$Ax = b$

Regularized approximation ($\|Ax-b\|, \|x\|$)

Scalarization – regularization path ( Scalarization, duality picture )

min. $\|Ax-b\| + \tau\|x\|$

other method : min. $\|Ax-b\|^2 + \tau\|x\|^2$

linear dynamical system problem

$$y(t) = \sum_{\tau=0}^{t} h(\tau)\, u(t-\tau)$$

$$J_{track} = \sum_{t=0}^{N} \left(y(t) - y_{des}(t)\right)^2$$

$$J_{mag} = \sum_{t=0}^{N} u(t)^2$$

$$J_{der} = \sum_{t=0}^{N-1} \left(u(t+1) - u(t)\right)^2$$

$$\min_{u(0),\dots,u(N)} \; J_{track} + \delta J_{der} + \eta J_{mag}$$

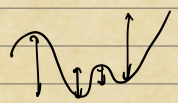
idea of regularization: give up little while still getting desirable result

Signal reconstruction

min. $\|\hat{x} - x_{cor}\|_2 , \phi(\hat{x})$

$$\phi_{quad}(\hat{x}) = \sum_{i=1}^{n-1} \left(\hat{x}_{i+1} - \hat{x}_i\right)^2 \quad \text{"smoothness"}$$

$$\phi_{TV} = \sum_{i=1}^{n-1} |\hat{x}_{i+1} - \hat{x}_i|$$



Robust approximation

Stochastic regularization term ⇒ variance

↕

worst-case example of non convex problem with strong duality

Lec 11 Statistical estimation

max. $l(x) = \sum_{i=1}^{m} \log p(y_i - a_i^T x)$ ← "parameter" of the model $y_i = a_i^T x + v_i$

better be log concave

Gaussian noise – LS estimation

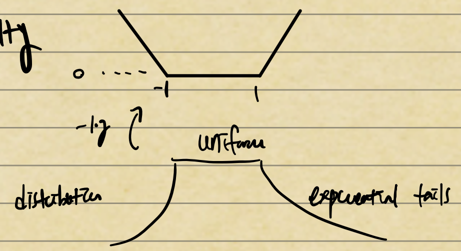
Laplacian – $l_1$ estimation → robust

heavy tails, allow more outliers

uniform $\quad -m \log(2a) \quad |a_i^T x - y_i| \le a \quad i = 1 \cdots m$

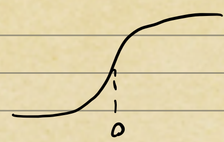
$\qquad\qquad -\infty \quad$ o.w.

(ex) penalty




Logistic regression

$p = \text{prob}(y=1) = \dfrac{\exp(a^T u + b)}{1 + \exp(a^T u + b)}$




(Rmk) $a^T u + b = 0$ ; $\frac{1}{2}$ , $> 0$ likely to be 1 , $< 0$ likely to be 0

a, b parameters to be estimated, $u_i$ & $y_i$ given as data

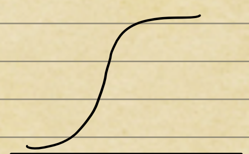
For $y_1 = \cdots = y_k = 1 \quad y_{k+1} = \cdots = y_m = 0$,

$$l(a,b) = \sum_{i=1}^{k} (a^T u_i + b) - \sum_{i=1}^{m} \log \underbrace{(1 + \exp(a^T u_i + b))}_{\text{log sum exp}}$$

Concave in (a, b)

$-l(a,b)$ as a function of $a^T u_i + b$ is called logistic loss function.

p(y=1)



b determines neutral pt, a determines stretch

→

$$l(a,b) = \sum_{i=1}^{k} (a^T u_i + b) - \cancel{\sum_{i=1}^{m} \log(1 + \exp(a^T u_i + b))}$$

↳ enlarged to $\infty$ $\qquad$ check!

hypothesis testing $\quad \hat{\theta} | X \otimes X | \theta = \hat{\theta} | \theta$

$\qquad\qquad\qquad\quad$ T $\qquad$ P $\qquad$ D

(ex) $\begin{array}{c|c} & X_1 \cdots X_n \\ \hat{\theta}_1 & \\ \hat{\theta}_2 & \end{array} \left( \quad \right) \quad \begin{array}{c|c} & \theta_1 \; \theta_2 \\ X_1 & \\ \vdots & \\ X_n & \end{array} \left( \quad \right)$

$D = (T_p \quad T_q) = \begin{pmatrix} 1-P_{fp} & P_{fn} \\ \underline{P_{fp}} & 1-P_{fn} \end{pmatrix}$

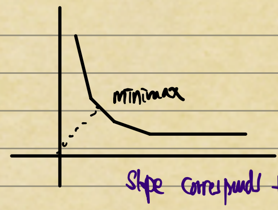
True 1

Select 2 ("positive": discovery)

Goal: minimize $P_{fn}$ & $P_{fp}$ : bicriterion problem

min. $P_{fn} + \pi P_{fp}$

min $\max(P_{fn}, P_{fp})$




Slope corresponds to different $\pi$

piecewise-linear since constraint is linear!

Experiment design

$w_i \sim$ IID $N(0,1)$

$y_i = a_i^T x + w_i$

$a_i \in \{v_1 \cdots v_p\}$

Goal: "min." $E = \left(\sum_{k=1}^{p} m_k v_k v_k^T\right)^{-1}$

Sub. to. $m_k \geq 0 \quad m_1 + \cdots + m_p = m$

$m_k \in \mathbb{Z}$

$\Big\downarrow$ relaxed

"min." $\frac{1}{m}\left(\Sigma\, \pi_k v_k v_k^T\right)^{-1}$

sub. to. $\pi \succeq 0 \quad 1^T\pi = 1$

Scalarize: D-optimal (volume)

⋮

Lec 12

"min." $\frac{1}{m}\left(\sum \pi_k v_k v_k^T\right)^{-1}$

sub. to. $\pi \succeq 0 \quad 1^T\pi = 1$

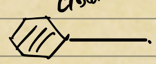
$\Big\downarrow$ dual (in the loose sense)

max. $\log\det W + n\log n$

sub. to. $v_k^T W v_k \leq 1 \quad k = 1 \cdots p$

} interpretation: maximum volume

* Often times, dual has nice interpretation

Geometric problems  closed convex set — iff exis unique (ex) $\{\mathbb{R}^{m\times n} : \text{rank } k\}$

Löwner-John Ellipsoid of a set $C$ : minimum volume ellipsoid $C \subseteq \mathcal{E}$

Need to choose right parameterization

$\{v : \|U\Sigma V^T v + b\|_2 \leq 1\}$

$= \{v : \|\underbrace{V\Sigma V^T}_{\tilde{A}} v + \underbrace{VU^T b}_{\tilde{b}}\|_2 \leq 1\}$

so WLOG $A \in S^n_{++}$

$\{v \mid \|Av + b\|_2 \leq 1\}$ (could be $\{x : (x-x_c)^T P^{-1}(x-x_c) \leq 1\}$) $\leftarrow A^TA$

$\leftarrow$ log det is concave for positive definite

$\min_{A,b} \quad \log\det A^{-1}$

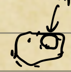
sub. to. $\|Av+b\|_2 \leq 1 \quad \forall v \in C$

} ⇒ convex problem

"$C \subseteq \mathcal{E}$"

$\downarrow$

intractable: Choose $C = \{x_1 \cdots x_m\}$ ⇒ become convex

not obvious (ex)  np hard (maximum ellipsoid inside)

application

$v_1 \cdots v_{1000} \in \mathbb{R}^{10}$ Q. outliers?

smallest ellipsoid that covers

; points on the surface are candidates of outliers

(remove and see if our statistical method gets better)

"ellipsoidal peeling" ← iteratively

$\mathcal{E} = \{Bu + d \mid \|u\|_2 \leq 1\}$ ← forward image WLOG $B \in S^n_{++}$

$\text{Vol } \mathcal{E} \propto \det B$

So...

$$\max_{B,d} \ \log\det B$$

sub. to. $\sup_{\|u\|\le 1} I_C(Bu+d) \le 0 \iff Bu+d \in \overset{\text{convex}}{C} \ \ \|u\|\le 1$ ; convex constraint

$I_C(x)=0 \ \ x\in C, \ \ \infty \ \ x\notin C$

(ex) $C$ is polyhedron $\{x \mid a_i^T x \le b_i \ \ i=1\cdots m\}$

$\mathcal{H}=\{x \mid f^T x \le g\}$

$Bu+d \in \mathcal{H} \ \ \forall \|u\|_2 \le 1$

$\iff f^T(Bu+d) \le g \ \ \forall \|u\|_2 \le 1$

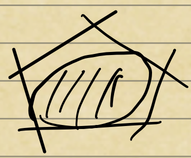
$\iff (B^T f)^T u + f^T d \le g \ \ \forall \|u\| \le 1$

$\iff \underbrace{\|B^T f\|}_{\text{cvx in } B} + \underbrace{f^T d}_{\text{cvx in } d} \le g$

So $\max. \ \log\det B$

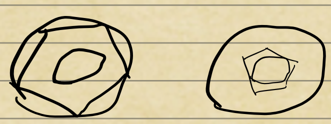
sub. to. $\|B a_i\|_2 + a_i^T d \le b_i \quad i=1\cdots m$ } cvx problem

($B$: symmetric)



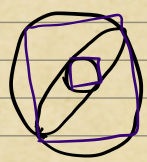
: cvx problem

Ellipsoidal approximation



Convexity is imp. here

(ex)



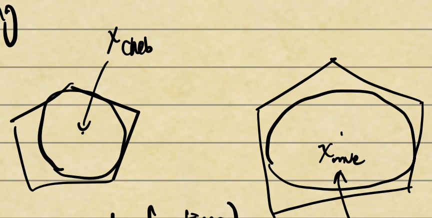
ball approximation X

box — " — X either

ellipsoid gives enough "degrees of freedom"

sense of "first order" approximation of convex set

Centering




(variant under transformation) maximum volume ellipsoid (affine invariant)

→ should ask "how trustworthy is your coordinate?"

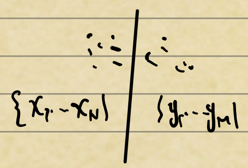
Analytic center of a set of inequalities

$f_i(x) \le 0 \qquad F(x) = g$

$$\max. \sum_{i=1}^{m} \log(-f_i(x)) \qquad \text{"product"}$$

sub. to. $F(x) = g$

Linear discrimination



$a^T x_i + b_i > 0 \qquad a^T x_i + b_i < 0$

$\Rightarrow$ strict inequalities. ( non strict ; $a = b = 0$ will do, problematic! )

$\Leftrightarrow$ $\exists a, b$ s.t.

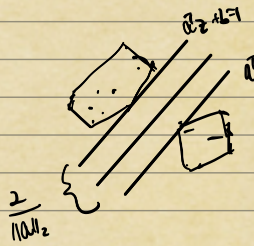
$\sim \ge 1, \sim \le -1$

Lec13

Linear discrimination

$\Rightarrow$ just linear programming

Set of $(a, b)$ $\begin{cases} \text{open polyhedral cone for } >0, <0 \\ \text{convex set } \geq 1, \leq -1 \end{cases}$




distance btwn convex hull → duality

$$\min. \quad \frac{1}{2}\|a\|_2$$

sub. to. $a^T x_i + b \geq 1 \quad i = 1 \cdots N$

$a^T y_i + b \leq -1 \quad i = 1 \cdots M$

$\Rightarrow$ QP in $a, b$

$\Big\downarrow$ dual

$$\max. \quad 1^T \lambda + 1^T \mu$$

sub. to. $2\|\sum_{i=1}^{N} \lambda_i x_i - \sum_{i=1}^{M} \mu_i y_i\|_2 \leq 1$

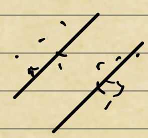
$1^T \lambda = 1^T \mu \quad \lambda \succeq 0 \quad \mu \succeq 0$

$\Big\downarrow$ Change of variable (use homogeneity)

$$\min. \quad t$$

sub. to. $\|\underbrace{\sum_{i=1}^{N} \theta_i x_i}_{\text{convex hull}} - \sum_{i=1}^{M} \gamma_i y_i\|_2 \leq t$

$\underbrace{\theta \succeq 0 \quad 1^T \theta = 1}$ $\quad \gamma \succeq 0 \quad 1^T \gamma = 1$



- What if not perfectly separable?

minimize the # of misclassification : NP hard

Heuristic:

$$\min \quad 1^T u + 1^T v$$

sub to. $a^T x_i + b \geq 1 - u_i \quad i = 1 \sim N$

$a^T y_i + b \leq -1 + v_i \quad i = 1 \sim M$

$u \succeq 0 \quad v \succeq 0$

combine

$\Rightarrow$ SVM

$$\min. \quad \|a\|_2 + \gamma(1^T u + 1^T v)$$

Sub. to. $a^T x_i + b \geq 1 - u_i$ $\quad i = 1 \cdots N$

$a^T x_i + b \leq 1 - v_i$ $\quad i = 1 \cdots M$

$u \succeq 0 \quad v \succeq 0$

Non linear discrimination

$f(z) = \theta^T F(z)$

$f(x_i) > 0 \quad i = 1 \cdots N \qquad f(x_i) < 0 \quad i = 1 \cdots M$

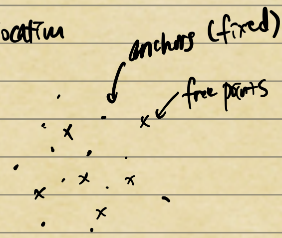
$\Leftrightarrow \quad \theta^T F(x_i) \geq 1 \quad i = 1 \cdots N \qquad \theta^T F(y_i) \leq -1 \quad i = 1 \cdots M$ linear program

(ex) $f(z) = z^T P z + q^T z + r$ (parameterized by $P, Q, R$)

Can make $P \preceq -I$ (ellipsoid)

placement and facility location

$x \in \mathbb{R}^2$ (or $\mathbb{R}^3$)




cost

min. $\sum_{i \neq j} f_{ij}(x_i, x_j)$

Algorithm

Usually $Ax = b$ takes $O(n^3)$

$A \in \mathbb{R}^{n \times n}$

low level

LAPACK

BLAS I II III

$A = A_1 A_2 \cdots A_k$

$A_1(A_2 \cdots (A_k x) \cdots) = b$

$A \backslash b$ $\quad n^3 + n^2$

$A \backslash [b_1 \, b_2 \cdots b_k]$ $\quad n^3 + kn^2$

identical!!

$(k \ll n)$

$A = PLU$ can see $O(n^3)$

non unique

backsolve: $n^2$

$\Rightarrow O(n^3)$

Lec 14

Sparse LU factorization

$A = P_1 \underline{LU} P_2$

row permu — column

↳ made to be sparse

Cholesky factorization

$A = LL^T$ ($\exists$ LU factorization and $L = U$)

↳ unique

Sparse Cholesky factorization

$A = PLL^TP^T$

↳ sparse

$P^TAP = LL^T$

{ dense : time is predictable
{ sparse : unpredictable
↳ depends on solver to get $P$ computer

$LDL^T$ (for symmetric matrix)

$P^TAP = LDL^T$

↳ block diagonal $\begin{pmatrix} D & \\ & \ddots \end{pmatrix}$ → $1\times1$ or $2\times2$

* $A\backslash b$ vs. $L = \text{chol}(A)$ $L'\backslash(L\backslash b)$
(faster)

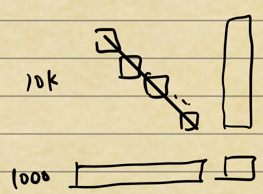
**Algorithm C.4** *Solving linear equations by block elimination.*

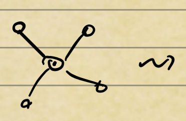
**given** a nonsingular set of linear equations (C.3), with $A_{11}$ nonsingular.

1. Form $A_{11}^{-1}A_{12}$ and $A_{11}^{-1}b_1$.
2. Form $S = A_{22} - A_{21}A_{11}^{-1}A_{12}$ and $\tilde{b} = b_2 - A_{21}A_{11}^{-1}b_1$.
3. Determine $x_2$ by solving $Sx_2 = \tilde{b}$.
4. Determine $x_1$ by solving $A_{11}x_1 = b_1 - A_{12}x_2$.

· flop count same. but in computer, there's a way to make it faster (cache, etc)

· when $A_{11}$ is special, it makes it faster like crazy



⇝ block elimination

(ex) DFT matrix    inverse takes $N\log N$

(ex) block diagonal

⋮

(ex) Toeplitz, etc.

Structured matrix plus low rank term

Q. Solve $(A+\underline{BC})x=b$ ?

$||$
low rank

Assume $\underset{\text{structured}}{\tilde{A}}x=b$ is easy

(un-eliminate)

$$\begin{pmatrix} A & B \\ C & -I \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} b \\ 0 \end{pmatrix}$$

$\therefore$

$x = A^{-1}(b-By)$    (should use $A^{+}$)

$CA^{-1}(b-By)-y=0$

$(I+CA^{-1}B)y = CA^{-1}b$    $Ax = b-By = b-B(I+CA^{-1}B)^{-1}CA^{-1}b$

$\underset{\text{non singular}}{(A+BC)^{-1}} = A^{-1}(I-B(I+CA^{-1}B)^{-1}CA^{-1})$

assumed to be invertible?

$(D+pq^T)x=b$ ;    $dI+gg^Tp$ ; EVD takes $O(n^3)$ again!!

In general, if a matrix is of form Diagonal + rank $k$ is impossible.

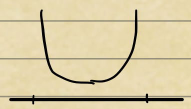
Lec 15

min. $f(x)$ unconstrained

$x^{(0)} \in \mathrm{dom}\, f$ ↪ not always easy to verify

sublevel set $S = \{x \mid f(x) \le f(x^{(0)})\}$ is closed (assumption)

⇐ epi $f$ is closed (check)



· true if dom $f = \mathbb{R}^n$

(check)

if $f(x) \to \infty$ as $x \to$ bd dom $f$

no equality

ex) $-\sum_{i=1}^{m} \log(b_i - a_i^T x)$

↳ counted as unconstrained problem even if $b_i - a_i^T x > 0$ has to hold.

theoretically,

All sorts of assumptions here that's not even verifiable

So it is only useful in so far as it helps us with understanding

$f$ is strongly convex if $\exists m$ s.t. $\nabla^2 f \succeq mI \quad \forall x \in S$ (minimum curvature)

(non-ex) $\frac{1}{x}$

implication

$(m>0)$

$\forall x, y \in S, \quad f(y) \ge f(x) + \nabla f^T(y)(y-x) + \frac{m}{2}\|y-x\|^2$ (minimum curvature away from linear approximation)

( + sublevel set $\{y \in \mathbb{R}^n : f(y) \le c\}$ is bdd

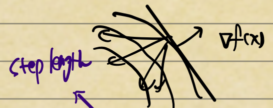
$p^* > -\infty$ and for $x \in S$

$f(x) - p^* \le \frac{1}{2m}\|\nabla f(x)\|_2^2$ useful as stopping criterion (if you know $m$)

Stopping criterion: "if $\nabla f$ is small, we stop"

Do we know $m$ in general? No we do not. ^^

Descent methods




$x^+ = x + t\Delta x$ → search direction

$f(x^+) < f(x) \longrightarrow \nabla f(x)^T \Delta x < 0$ because $f(x^+) \ge f(x) + t\Delta x^T \nabla f(x)$

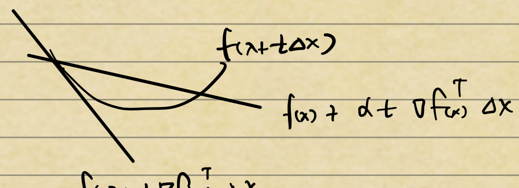
One choice: $\Delta x = -\nabla f(x)$ (very poor direction though)

Choose $t$.

① exact line search

$$t = \underset{t>0}{\operatorname{argmin}}\ f(x + t\Delta x)$$

② backtracking line search (very crude)




$$f(x + t\Delta x) < f(x) + \underbrace{\alpha t \nabla f(x)^T \Delta x}_{<0}$$

$\alpha \in (0, \frac{1}{2})$ $\beta \in (0,1)$

Fact: it doesn't have a whole lot difference

Should be careful with domain. For ODD, convenient to put $\infty$

. Gradient descent method: Take $\Delta x = -\nabla f(x)$ (this is the greedy algorithm)

cvg: $f(x^{(k)}) - p^* \le c^k (f(x^{(0)}) - p^*)$ $\Rightarrow$ exponentially improved!

$c \in (0,1)$ depends on $m$, $x^{(0)}$, line search type

(ex) $f(x) = \frac{1}{2}(x_1^2 + \gamma x_2^2)$ $\gamma > 0$

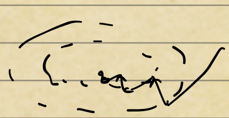
$\gamma > 1$: $\gamma$ condition number

$\gamma < 1$: $\frac{1}{\gamma}$ —— " ——

Start at $(\gamma, 1)$

$x_1^{(k)} = \gamma \left(\frac{\gamma - 1}{\gamma + 1}\right)^k$ & $x_2^{(k)} = \left(-\frac{\gamma - 1}{\gamma + 1}\right)^k$

optimal: $(0, 0)$



Steepest descent method

$\Delta x_{nsd} = \operatorname{argmin} \{ \nabla f(x)^T v : \|v\| = 1 \}$ — might not necessarily be $\|\cdot\|_2$

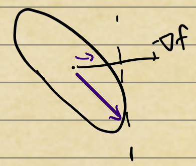
$\Delta x_{sd} = \|\nabla f(x)\|_* \Delta x_{nsd}$

$\nabla f(x)^T \Delta x_{sd} = -\|\nabla f(x)\|_*^2$

Q. why does it even depend on the norm?

It's like using different unit in different direction

Intuition

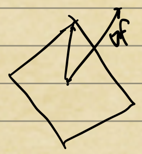



$\|x\|_P = (x^T P x)^{\frac{1}{2}}$ $\quad P \in S_{++}^n$

$\Delta x_{sd} = -\underset{\sim}{P^{-1}} \nabla f(x)$

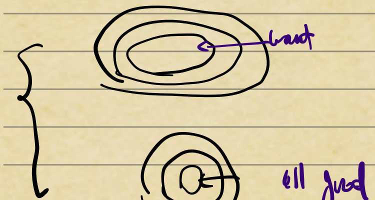
rotate (but less than 90 degrees)

$L^1$



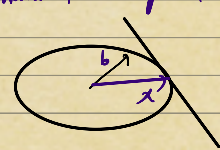

for $L^1$: on the vertex

$L^1$: optimize over one component in the variable




$\hat{f}(x) = f(x^*) + \nabla f\cancel{(x^*)^T}(x - x^*) + \frac{1}{2}\underbrace{(x-x^*)^T \nabla^2 f(x^*)(x-x^*)}$

at least near $x^*$, use the norm induced by Hessian. "Newton's method"



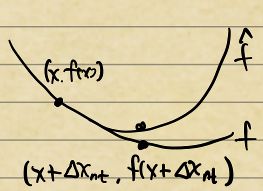

$\rightsquigarrow Ax \propto b$ ? yes, $x^T A x = 1 \xrightarrow{d} Ax = 0$

$\Delta x_{nt} = -\nabla^2 f(x)^{-1} \nabla f(x)$

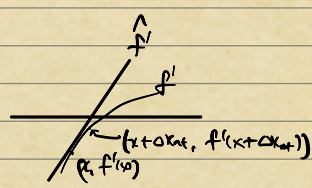
$x + \Delta x_{nt}$ minimizes second order approximation

or... make linear approximation of $\nabla f(x+v)$ 0. $\quad (\hat{\nabla f}(x+v) = \nabla f(x) + \nabla^2 f(x)\, v = 0)$

1)



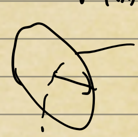

2)




Something like this is called "variable metric method"

$\Delta x_{nt}$ is steepest descent at $x$ in local Hessian norm

$\|u\|_{\nabla^2 f(x)} = (u^T \nabla^2 f(x)\, u)^{\frac{1}{2}}$



Newton decrement

$\lambda(x) = \left(\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)\right)^{\frac{1}{2}}$

a measure of the proximity of $x$ to $x^*$

* $f(x) - \inf_y \hat{f}(y) = \frac{1}{2}\lambda^2$ (the smaller, the closer)
* $\lambda(x) = \left(\Delta x_{nt} \nabla^2 f(x) \Delta x_{nt}\right)^{\frac{1}{2}}$
* $\Delta f(x)^T \Delta x_{nt} = -\lambda(x)^2$ used in line search anyway

* affine invariant (check)

Newton's method

① get $\Delta x_{nt}$ and $\lambda$
② $\frac{\lambda^2}{2} < \varepsilon$ ; quit
③ Do backtracking line search
④ $x := x + \Delta x_{nt}$

why it works? affine invariant

$Ty = x$

$\tilde{f}(y) = f(Ty)$

$\nabla \hat{f}(y) = T' f(Ty)$ change coordinate → gradient method changes in general

We said the best one is "steepest gradient descent"

For Newton's method, they commute!

- Convergence

Assumption: $f$ strongly convex on $S$ w. constant $m$

$\nabla^2 f$ is Lipschitz continuous on $S$ with constant $L > 0$.

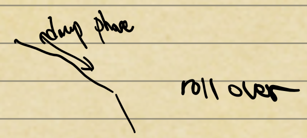
$\|\nabla^2 f(x) - \nabla^2 f(y)\|_2 \le L\|x-y\|_2$ (since Newton's method uses second order TE, this makes sense)

$\exists \eta \in (0, \frac{m^2}{L})$ $\gamma > 0$ s.t.

. If $\|\nabla f(x)\|_2 \ge \eta$ $f(x^{(k+1)}) - f(x^{(k)}) \le -\gamma$

$\|\nabla f(x)\|_2 < \eta$

(damp phase) $\frac{L}{2m^2}\|\nabla f(x^{(k+1)})\|_2 \le \left(\frac{L}{2m^2}\|\nabla f(x^{(k)})\|_2\right)^2$



Lec 16

Newton's method $(\nabla^2 f)^{-1}$ takes $O(n^3)$ ; Not a big deal anymore

L small → Newton's method works well.

implementation Newton's method ⇒ invariable under affine transformation

BUT... convergence analysis depends (for example, L, M, etc)

Need $\dfrac{``|f'''|}{f''} \le L''$

* Self-concordance

- Convergence analysis that does not depend on coordinate
  - does not depend on $L, m$
  - can be generalized

Def

(Convex) $f: \mathbb{R} \to \mathbb{R}$ is self-concordant if $|f'''(x)| \le 2 f''(x)^{\frac{3}{2}}$ for all $x \in \mathrm{dom} f$

(ex) $-\log x$

(ex) linear/quadratic

(ex) $x \log x - \log x$

- Affine invariance

$$|f'''(x)| \le 2 f''(x)^{\frac{3}{2}}$$

$$\rightarrow |a^3 f'''(ay+b)| \le 2[a^2 f''(ay+b)]^{\frac{3}{2}}$$

- preserved under $\alpha \ge 1$ scaling and sum

- $g$ is convex w. dom $g = \mathbb{R}_{++}$ $|g'''(x)| \le 3 g''(x)/x$

$\Rightarrow$ $f(x) = \log(-g(x)) - \log x$ is self-concordant

(ex) $-\sum_i \log(b_i - a_i^T x)$

ex) $-\log \det X$

(ex) $-\log(y^2 - x^T x)$

Convergence of self-concordant function (similar)

- $\eta, \gamma$ depend on $\alpha, \beta$ (back tracking parameters) so can get actual numbers!

(ex) $\alpha = 0.1$ $\beta = 0.8$ $\varepsilon = 10^{-10}$ then # iterations $\le 375(f(x^{(0)}) - p^*) + 6$

$\pi(x) > \eta \qquad f(x^{(l+1)}) - f(x^{(l)}) \leq -\gamma$

$\pi(x) \leq \eta \qquad 2\pi(x^{(l+1)}) \leq (2\pi\, x^{(l)})^2$

- Solving $\underbrace{\nabla^2 f(x)}_{\text{positive definite}} \Delta x = -\nabla f(x)$

$\Rightarrow$ Use cholesky ( $O(n^3)$ )

If structured. much faster

(ex) sparse Hessian : each variables couple non-linearly only into handful of others

$$\sum_{i=1}^{3} \exp(x_i - x_{i+1}) = e^{x_1 - x_2} + e^{x_2 - x_3} + e^{x_3 - x_4}$$

$$\begin{pmatrix} \times & \times & 0 & 0 \\ \times & \times & \times & 0 \\ 0 & \times & \times & \times \\ 0 & 0 & \times & \times \end{pmatrix} \quad \text{tri diagonal} \quad O(n)$$

(ex) $f(x) = \sum_i \psi_i(x_i) + \psi_0(Ax+b)$ $\quad H = D + A^T H_0 A$ $\quad$ diagonal + low rank

$H_0 = \nabla^2 \psi_0(Ax+b)$, $\quad D$: $\psi_i''(x_i)$

assume $A$ dense $\in \mathbb{R}^{p \times n}$, $p \ll n$

$\big(\qquad\big)$

$H_0 = L_0 L_0^T \qquad \begin{cases} D\Delta x + A^T L_0 w = -g \\ L_0^T A \Delta x - w = 0 \end{cases}$

Eliminate $\Delta x$ ; $\quad (I + L_0^T A D^{-1} A^T L_0) w = -L_0^T A D^{-1} g$

$D \Delta x = -g - A^T L_0 w$

can compute $w, \Delta x$ and it will take $2p^2 n$ (dominated by $L_0^T A D^{-1} A^T L_0$)

Forming $H$ directly is highly inefficient. (cannot even store it)

* Equality constraint minimization

min. $f(x)$

sub. to. $Ax = b$

$\Leftrightarrow \exists \nu^* \text{ s.t. } \underbrace{\nabla f(x^*) + A^T \nu^* = 0}_{\text{dual residual} \,=\, r_d} \qquad \underbrace{Ax^* = b}_{\text{primal residual } Ax-b \,=\, r_p}$

- Assume $A = \big(\qquad\big)$ (solvable)
- $p^* > -\infty$

( $n+p$ equations)

linear in $\nu$ $\qquad$ linear

(ex) min. $\frac{1}{2} x^T P x + q^T x + r$

sub. to. $Ax = b$

$$\begin{pmatrix} P & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x^* \\ \nu^* \end{pmatrix} = \begin{pmatrix} -q \\ b \end{pmatrix} \quad \begin{matrix} \text{dual feasibility} \\ \text{primal feasibility} \end{matrix}$$

KKT matrix

KKT matrix non-singular $\Leftrightarrow$ $Ax=0$ $x\neq 0$ $\Rightarrow x^TPx>0$

$\Leftrightarrow$ $P+A^TA \succ 0$

$\Leftrightarrow$ $N(P)\cap N(A) = \{0\}$

prof) ($\Leftarrow$) Assume Singular. $\exists x, w$ s.t.

$\begin{pmatrix} P & A^T \\ A & 0 \end{pmatrix}\begin{pmatrix} x \\ w \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

$Px + A^Tw$ $\rightarrow x^TPx = 0 \rightarrow Px=0$

$= Ax = 0$ $\rightarrow x \in null(A) \cap null(P)$

$\rightarrow x=0$

$\rightarrow w\neq 0 \rightarrow null(A^T) \neq \{0\}$

$\Leftrightarrow$ rank $A = P$

($\Rightarrow$) Assume non singular

$Px = Ax = 0$ but $x\neq 0$

$\begin{pmatrix} P & A^T \\ A & 0 \end{pmatrix}\begin{pmatrix} x \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ ※

$\{x \mid Ax=b\} = \{Fz+\hat{x} \mid z\in\mathbb{R}^{n-p}\} = \hat{x}+N(A)$

range of $F$ = Null $(A)$, rank $F = n-p$, $AF=0$

$p\times n$ $n\times(n-p)$

$\rightsquigarrow$ reduced to unconstrained problem

min. $f(Fz+\hat{x})$

Then, $x^* = Fz^* + \hat{x}$ and $v^* = -(AA^T)^{-1}A\nabla f(x^*)$ Since $\nabla f(x^*) + A^Tv^* = 0$

(ex) min. $f_1(x_1) + \cdots + f_n(x_n)$

sub to. $x_1 + \cdots + x_n = b$

elimination: $\min_{x_1\cdots x_{n-1}} f(x_1) + \cdots + f_{n-1}(x_{n-1}) + f_n(b - x_1 - \cdots - x_{n-1})$

$\left[\begin{array}{l} F = \begin{pmatrix} I \\ -\mathbf{1}^T \end{pmatrix} \in \mathbb{R}^{n\times(n-1)} \qquad Fz+\hat{x} = (z_1, \cdots, z_{n-1}, b - z_1 - \cdots - z_{n-1}) \\ \hat{x} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ b \end{pmatrix} \end{array}\right]$

Hessian: $D$ + rank 1 $\Rightarrow$ take order $n$ ("so" newton's step, each taking $O(n)$)

Newton step $\begin{pmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{pmatrix}\begin{pmatrix} \Delta x_{nt} \\ w \end{pmatrix} = \begin{pmatrix} -\nabla f(x) \\ 0 \end{pmatrix}$

① $\Delta x_{nt}$ Solves

$\min_v \hat{f}(x+v) = f(x) + \nabla f(x)^T v + \frac{1}{2}v^T \nabla^2 f(x) v$

Sub. to. $A(x+v) = b$

② $\underbrace{\nabla f(x+v)}_{\text{linearize}} + A^Tw = 0$ $A(x+v)=b$ optimality condition

$\rightsquigarrow \nabla f(x) + \nabla^2 f(x) v$

Newton Decrement (there are many)

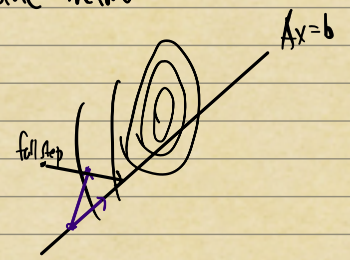
$\lambda(x) = (\Delta x_{nt}^T \nabla^2 f(x) \Delta x_{nt})^{\frac{1}{2}} = (-\nabla f(x)^T \Delta x_{nt})^{\frac{1}{2}}$

Algorithm (feasible descent method, affine invariant)

many choices of F: irrelevant

Convergence analysis: same (can verify $\lambda$ and newton step same. well-defined same)

infeasible method




$$\begin{pmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x_{nt} \\ w \end{pmatrix} = -\begin{pmatrix} \nabla f(x) \\ Ax-b \end{pmatrix} \quad r_p$$

Lec 17

min. $f(x)$

sub. to. $Ax = b$

1) $\begin{pmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \Delta x_{nt} \\ w \end{pmatrix} = \begin{pmatrix} -\nabla f(x) \\ 0 \end{pmatrix}$ Solve Newton System

2) $\frac{\lambda^2}{2} < \varepsilon$ : quit

3) backtracking line search

4) $x := x + t\Delta x_{nt}$

- $\Delta x_{nt} \in \text{null}(A)$ so $x + t\Delta x_{nt}$ is feasible
- $f(x^{(k+1)}) < f(x^{(k)})$ b.c. line search.
- Affine invariant

So this method vs. elimination of variable. which is better?

faster if $\begin{pmatrix} \nabla^2 f(x) & A^{-1} \\ A & 0 \end{pmatrix}$ has a structure

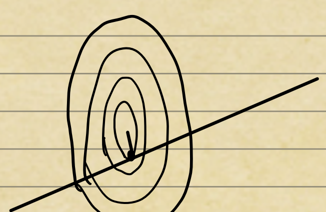
- Infeasible Newton method

$\begin{pmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \Delta x_{nt} \\ w \end{pmatrix} = \begin{pmatrix} -\nabla f(x) \\ -(Ax-b) \end{pmatrix}$

generalization of feasible Newton method

$A(x + \Delta x_{nt}) = b$

Full Newton step: go into feasible set immediately $(t=1)$



backtracking : $\left\| \begin{bmatrix} r_d \\ r_p \end{bmatrix} \right\|$ goes down

$r(y) = (\nabla f(x) + A^T v,\ Ax - b)$

$r_d = \nabla f(x) + A^T v$

$r_p = Ax - b$

- directional derivative

· entire method works well

· KKT systems

$$\begin{array}{c} n \\ m \end{array}\begin{pmatrix} H & A^T \\ A & 0 \end{pmatrix}\begin{pmatrix} v \\ w \end{pmatrix} = -\begin{pmatrix} g \\ h \end{pmatrix}$$

LDL$^T$ factorization  symmetric, not positive definite

elimination  $\underbrace{AH^{-1}A^T}_{\text{Schur complement}} w = h - AH^{-1}g \qquad Hv = -(g + A^T w)$ ; $H^{-1}$ is easy

elimination w. singular H

$$\begin{pmatrix} H + A^T Q A & A^T \\ A & 0 \end{pmatrix}\begin{pmatrix} v \\ w \end{pmatrix} = -\begin{pmatrix} g + A^T Q h \\ h \end{pmatrix}$$

$Q \succeq 0$ for which $H + A^T Q A \succ 0$

(ex) $\min. \ -\sum_i \log x_i$

sub. to. $Ax = b$

$\Updownarrow$ in numerical algebra pov, same difficulty!

$\max \ -b^T v + \sum_i \log (A^T v)_i + n$

infeasible method

perhaps primal feasibility achieved

$A \operatorname{diag}(x)^2 A^T w = b$  elimination KKT system (feasibility)

$A \operatorname{diag}(A^T v)^{-2} A^T \Delta v = -b + A \operatorname{diag}(A^T v)^{-1} \mathbf{1}$  dual

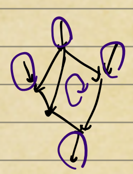
$A \operatorname{diag}(x)^2 A^T w = 2Ax - b$  elimination KKT system (infeasibility)

complexity is same!

ex) KCL

$Ax = b$

edges



null space: circulation

nodes $\begin{pmatrix} 1 & \cdots \\ -1 & \end{pmatrix}$

min. $\sum_{i=1}^{n} \phi_i(x_i)$ $\Rightarrow$ Hessian diagonal $\begin{pmatrix} \ddots & \square \\ \square & \circ \end{pmatrix}$

sub. to. $Ax=b$

check! sparsity pattern of $\underbrace{AH^{-1}A^T}_{\text{(need to inverse)}}$ $\equiv$ pattern of $AA^T$ $\Leftrightarrow$ $(i,j)$ entry : $i$ and $j$ are connected

(ex) Analytic center of linear matrix inequality

min. $-\log\det \overset{n}{\widehat{X}}\}n$

sub. to. $tr(A_i X) = b_i$ $i=1\cdots p$ $X \in S^n$

(ex) $\sim \begin{cases} Z \sim N(0, X) \\ E(G^T Z)^2 = G^T X G = Tr(X G G^T) \end{cases}$

$\left(\frac{n(n+1)}{2} + p\right)^3$ too big.

# of variables

$\Rightarrow pn^3 + p^2n^2 + \frac{1}{3}p^3$ (check)

* Interior point method

non diff constrained

transformation

inequality constraint

smooth unconstrained

quadratic unconstrained

(i.e. linear equations)

min. $f_0(x)$

sub. to $f_i(x) \le 0$ $i=1\cdots m$

$Ax=b$ $A \in \mathbb{R}^{p\times n}$ rank $A = p$

- Slater's condition
- smooth
- finite $p^*$

* discussion about smoothing $|x|$ to $\sqrt{x^2+(10^{-5})^2} - 10^{-5}$

third derivative is not bounded!

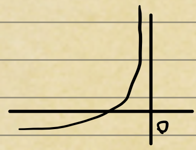
Lec 18

Rewrite to $f_0(x) + \sum_{i=1}^{m} I_-(f_i(x))$ $\quad I_-(x) = 0$ if $x \le 0$ else $\infty$

Sub. to. $Ax = b$

min. $f_0(x) - \frac{1}{t}\sum_{i=1}^{m} \log(-f_i(x))$

Sub. to. $Ax = b$



logarithmic barrier

smooth approximation

tradeoff: harder with Newton's method

log barrier function

$$\phi(x) = -\sum_{i=1}^{m} \log(-f_i(x))$$

· Convex

$f(x) = -\log \prod_i (-f_i(x))$ "slacks"

$$\nabla f(x) = \sum_i \frac{1}{-f_i(x)} \nabla f_i(x)$$

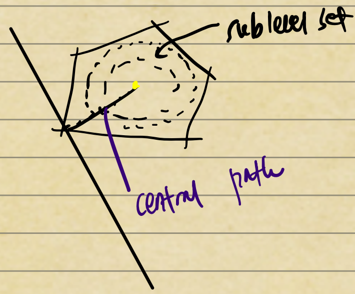
$$\underbrace{\sum_i \frac{1}{f_i^2(x)} \nabla f_i(x) \nabla f_i(x)^T}_{\text{rank } 1} + \sum_i \frac{1}{-f_i(x)} \nabla^2 f_i(x)$$

Take $x^*(t)$ be the solution of

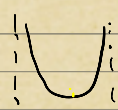
min. $t f_0(x) + \phi(x)$ ... $\circledast$

Sub. to. $Ax = b$

· $\{x(t) : t > 0\}$ : central path




$\phi$ : keep you away from bdry



$t \nabla f_0(x) + \nabla\phi(x) + A^T w = 0$ $\quad Ax = b$

$$= t\nabla f_0(x) + \sum_i \frac{1}{-f_i(x)} \nabla f_i(x) + A^T w = 0$$

$$\nabla f_0(x) + \sum_i \underbrace{\frac{1}{-t f_i(x)}}_{:= \lambda_i} \nabla f_i(x) + A^T \underbrace{\frac{w}{t}}_{:= \nu_i} = 0$$

$L(x,\lambda,\nu) = f_0(x) + \sum_i \lambda_i f_i(x) + \nu^T(Ax-b)$

$f_0(x^*(t)) \to p^*$ if $t \to \infty$

$p^* \ge g(\lambda^*(t), \nu^*(t))$

$= L(x^*(t), \lambda^*(t), \nu^*(t)) = f_0(x) + \sum_i \frac{1}{-t\cancel{f_i(x)}} \cancel{f_i(x)} + \cancel{\nu^T(Ax-b)}$

$= f_0(x^*(t)) - \frac{m}{t}$

We by solving ⊗ using Newton's method, we get a lower bound of $p^*$

$f_0(x) - \frac{m}{t} \le p^* \le f_0(x)$

(S)UMT Sequential unconstrained minimization technique

$x = x^*(t)$ $\lambda = \lambda^*(t)$ $\nu = \nu^*(t)$

1. $f_i(x) \succ 0$ $i = 1 \dots m$ $Ax = b$
2. $\lambda \succ 0$
3. $-\lambda_i f_i(x) = \frac{1}{t}$ ~> only difference in terms of KKT conditions
4. $\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + A^T \nu = 0$

$\underbrace{t f_0(x)}_{\text{potential of force field}} - \underbrace{\sum_i \log(-f_i(x))}$ ↗ potential of force field $\frac{1}{f_i(x)} \nabla f_i(x) = F_i(x)$

↳ potential of force field $-t\nabla f_0(x) = F_0(x)$

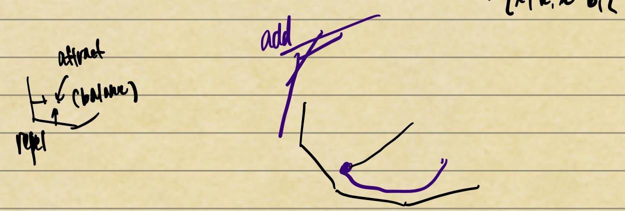
$F_0(x^*(t)) + \sum_{i=1}^{m} F_i(x^*(t)) = 0$ force balanced at $x^*(t)$

(ex) $f_0(x) = c^T x$

force field $F_0(x) = -tc$ ( like tilting the plane)

$F_i(x) = \frac{-a_i}{b_i - a_i^T x}$ $\|F_i(x)\| = \frac{1}{\mathrm{dist}(x, \mathcal{H}_i)}$

↳ $\{x \mid a_i^T x = b_i\}$

⇒ 


Barrier method (SUMT)

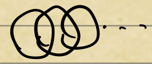
Strictly feasible, $t = t^{(0)} > 0$, $\mu > 1$ and tolerance $\varepsilon > 0$.

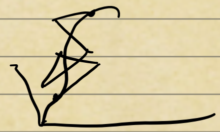
$$\begin{cases} \text{Compute } x^*(t) \text{ by min. } tf_0+\phi \text{ sub. to. } Ax=b \quad (\text{initial guess: } x^* \text{ in the previous step}) \\ \text{Update } x = x^*(t) \\ \text{quit if } m/t < \varepsilon \\ t := \mu t \end{cases}$$

↳ centering steps

outer iterations

Several heuristics for $t^{(0)}$. $\mu = 10-20$.





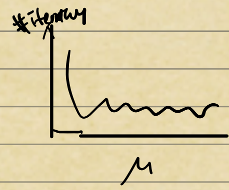
Convergence $\left\lceil \dfrac{\log\left(\frac{m}{\varepsilon t^{(0)}}\right)}{\log \mu} \right\rceil$ iterations + initial centering step to compute $x^*(t^{(0)})$

Each centering problem min. $\underset{\sim\sim\sim}{tf_0(x)+\phi(x)}$

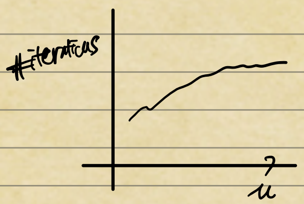
$t \geq 1 \Rightarrow$ self-concordance, few more assumptions needed

Q.

← no error for each Newton step?




$\Rightarrow$ usually need not be tuned




analysis: $\sqrt{m}$ newton steps $\times$ $\underline{m^3 \text{ each newton}}$ $\propto m^{3.5}$

\# constraint

assump

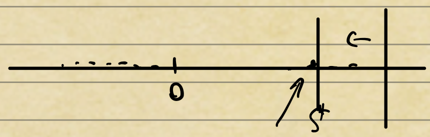
empirically $O(1)$ 20~50 steps

$\underline{2m \text{ variables, } m \text{ constrains}}$ # iterations doesn't grow!

Feasibility problem: $f_i(x) < 0 \quad i=1\cdots m \quad Ax=b$

$$\left\{ \begin{array}{ll} \min & s \\ \text{sub. to.} & f_i(x) \leq s \\ & Ax = b \end{array} \right.$$

as soon as $s<0$ ⇝ feasible




bunch will be concentrated here

$$\min. \ \mathbf{1}^T s$$

$$\text{sub. to. } \ s \succeq 0 \quad f_i(x) \leq s_i \quad i = 1 \cdots m$$

$$Ax = b$$

$\Rightarrow$ heuristic for violating just a few

Lec 19

find strictly feasible point

① basic method — introduce 1 variable

② SVM type method — introduce m variables (doesn't matter)

$\min \|S\|_1$

sub. to $f_i(x) \le S_i$ → small # of constraint violation (heuristic)

$Ax = b$

Convergence analysis

constant ⇐ how infeasible the problem is

In principle this shouldn't be allowed

In practice, ~80 steps at most

$\mu t f_0 + \phi$

how do we get this? Duality

$$\le \frac{\mu t f_0(x) + \phi(x) - \mu t f_0(x^+) - \phi(x^+)}{\gamma} + 5$$

$\gamma$ ← constant

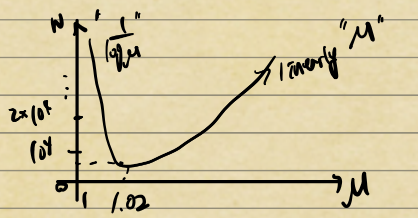
$x - \frac{1}{2}\mu x^2$

$$\le \left[ \frac{m(\mu - 1 - \log \mu)}{\gamma} + c \right]$$

$t \to \infty$ doesn't grow!

$$\otimes \left[ \frac{\log\left(\frac{m}{t^{(0)}\varepsilon}\right)}{\log \mu} \right]$$

"decrease in duality gap per each Newton step"




{ sharp x
{ far off


for $\mu = 1 + \frac{1}{\sqrt{m}}$

$$N = O\left(\sqrt{m} \log\left(\frac{m/t^{(0)}}{\varepsilon}\right)\right)$$

# of bts steps — duality gap = ignorance reduction ratio

this is where we can improve a lot!

n is used for computing Hessian ... (# flops ...) ~ $(n+m)^3$ — SOCP etc... $O(m^{3.5})$!

Generalized inequalities (SOCP, SDP)

min. $f_0(x)$

sub. to. $f_i(x) \preceq_{K_i} 0$ $i = 1 \cdots m$ ← proper cone

$Ax = b$

SDP can be solved in a classical way, but harder

min. $c^T x$

→ $A(x) \succeq 0$

$L_{ii}(A(x)) \geq 0$

$(i,i)$ of Cholesky factor of $A$, can be shown to be concave

$-L_{ii}(A(x)) \leq 0$

Generalized logarithm

1) $\text{dom}\,\psi = \text{int}\,K$

2) $\nabla^2 \psi(y) \prec 0$

3) $\psi(sy) = \psi(y) + \theta \log s$ for $y \succ_K 0$ $s > 0$ ($\theta$ is degree of $\psi$)

"ray"

(ex) $K = S^n_+$ $\psi(Y) = \log\det Y$

$\nabla \log\det Y = Y^{-1}$

$\forall y \succ_K 0$, $\nabla\psi(y) \succeq_{K^*} 0$ $y^T \nabla\psi(y) = \theta$ (make sense since $\text{Tr}\, Y Y^{-1} = n$)

(ex) Second order cone $\theta = 2$

$$\underbrace{\phi(x) = -\sum_{i=1}^{m} \psi_i(-f_i(x))}_{\text{By composition rule, convex}}$$

Identical argument: duality gap $\sum \theta_i / t$ ($\frac{m}{t}$ in scalar case)

Advanced method