Ch2 Convex sets

Line / line segments

$y = \theta x_1 + (1-\theta)x_2$

$x_1 \neq x_2$, $\theta \in R$

affine sets

$C \rightarrow$ affine

$x_1, \dots, x_k \in C$

$\theta_1 + \dots + \theta_k = 1$

$\rightarrow \theta_1 x_1 + \dots + \theta_k x_k \in C$

$\theta_i \in R$

eg.1

$V = C - x_0 = \{x - x_0 \mid x \in C\}$ subspace

$v_1, v_2 \in V$

$v_1 + x_0 \in C$

$v_2 + x_0 \in C$

$\alpha(v_1 + x_0) + \beta(v_2 + x_0) + (1-\alpha-\beta)x_0 \in C$

$\therefore \underbrace{\alpha v_1 + \beta v_2}_{\in V} + x_0 \in C$

$\therefore x_0$ can be anything in C

eg.2

$C = \{x \mid Ax = b\}$

$x_1 \in C$, $x_2 \in C$

$Ax_1 = b$, $Ax_2 = b$

$\Rightarrow A(\theta x_1 + (1-\theta)x_2)$

$= A\theta x_1 + A(1-\theta)x_2$

$= \theta b + (1-\theta) b$

$= b$

$\therefore \theta x_1 + (1-\theta)x_2 \in C$

$C \subseteq R^n$

affine hull

$\text{aff } C = \{\theta_1 x_1 + \dots + \theta_k x_k \mid x_1 \dots x_k \in C, \ \theta_1 + \dots + \theta_k = 1\}$

$C \subseteq S$, $\text{aff } C \subseteq S$

affine dimension & relative interior

relint is particular used when $\text{int}(C) = \emptyset$

e.g.

$C = \{x \in R^3 \mid -1 \le x_1 \le 1, \ -1 \le x_2 \le 1, \ x_3 = 0\}$

$\text{relint } C = \{x \in C \mid B(x, r) \cap \text{aff } C \subseteq C \text{ for } r > 0\}$

relint C

$= \{x \in R^3 \mid -1 < x_1 < 1, \ -1 < x_2 < 1, \ x_3 = 0\}$

- if $\text{int } C \neq \emptyset$, $\text{relint } C = \text{int } C$
- all $\text{int } C \subset \text{relint } C$
- relative boundary of C: $\text{cl } C \setminus \text{relint } C$

polyedra

$P = \{x \mid a_j^T x \le b_j, \ j = 1 \dots m, \ c_j^T x = d_j, \ j = 1 \dots p\}$

- intersection of "FINITE" number of halfspaces & halfplanes

$\Rightarrow P = \{x \mid Ax \preceq b, \ Cx = d\}$

$A = \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix}$ $\quad C = \begin{bmatrix} c_1^T \\ \vdots \\ c_p^T \end{bmatrix}$

eg. simplex

$= C = \text{conv}\{v_0 \dots v_k\}$

$v_0 \dots v_k \rightarrow$ affinely independent.

simplex $\rightarrow$ polyhedron representation

simplex:

$C = \text{conv}\{v_0 \dots v_k\}$

$= \{x \mid \theta_0 v_0 + \dots + \theta_k v_k \mid \theta \succeq 0, \ 1^T\theta = 1\}$

$x = \theta_0 v_0 + \theta_1 v_1 + \dots + \theta_k v_k$

$= (1 - \theta_1 - \theta_2 \dots - \theta_k) v_0 + v_1\theta_1 + \dots + v_k\theta_k$

$= v_0 + (v_1 - v_0)\theta_1 + (v_2 - v_0)\theta_2 + \dots + (v_k - v_0)\theta_k$

let $B = [v_1 - v_0 \ \cdots \ v_k - v_0]$

$y = [\theta_1, \dots, \theta_k]$

$\therefore x = v_0 + By$

$\underbrace{A}_{\text{non-singular matrix exists}} B = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} B = \begin{bmatrix} I \\ 0 \end{bmatrix}$

$\therefore Ax = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}(v_0 + By)$

$\begin{cases} A_1 x = A_1 v_0 + A_1 B y \\ A_2 x = A_2 v_0 + A_2 B y \end{cases}$

$\begin{cases} A_1 x = A_1 v_0 + y \\ A_2 x = A_2 v_0 \end{cases}$

$A_1 x \succeq A_1 v_0$

$A_2 x = A_2 v_0$

convex sets

$C \rightarrow$ convex

$x_1, \dots x_k \in C$

$\theta_1 + \dots + \theta_k = 1$

$\theta_i \ge 0 \in R$

$\rightarrow \theta_1 x_1 + \dots + \theta_k x_k \in C$

$C \subseteq R^n$

convex hull:

$\text{conv } C = \{\theta_1 x_1 + \dots + \theta_k x_k \mid x_1 \dots x_k \in C, \ \theta_1 + \dots + \theta_k = 1, \ \theta_i \ge 0\}$

Convex cone

$\theta_1, \dots \theta_k \ge 0$

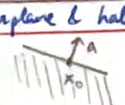
$\theta_1 x_1 + \dots + \theta_k x_k$

$C \subseteq R^n$

conic hull:

$\{\theta_1 x_1 + \dots + \theta_k x_k \mid x_1 \dots x_k \in C, \ \theta_i \ge 0\}$

hyperplane & halfspaces



$\neq$ affine

$=$ convex

Euclidean balls & ellipsoids

$B(x_c, r) = \{x \mid \|x - x_c\|_2 \le r\}$

$= \{x \mid (x - x_c)^T(x - x_c) \le r^2\}$

$= \{x_c + ru \mid \|u\|_2 \le 1\}$

$\mathcal{E} = \{x \mid (x - x_c)^T P^{-1}(x - x_c) \le 1\}$

$= \{x_c + Au \mid \|u\|_2 \le 1\}$

length of $\mathcal{E}$ $\sqrt{\lambda_i}$, $A = P^{1/2}$

$P \in PD$ $\quad P = P^T$

Norm balls, norm cones, second-order cone

norm ball: $C = \{x \mid \|x - x_c\| \le r\}$

norm cone: $C = \{(x, t) \mid \|x\| \le t\} \subseteq R^{n+1}$

second-order cone: $C = \{(x, t) \in R^{n+1} \mid \|x\|_2 \le t\}$

$= \left\{ \begin{bmatrix} x \\ t \end{bmatrix} \middle| \begin{bmatrix} x \\ t \end{bmatrix}^T \begin{bmatrix} I & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ t \end{bmatrix} \le 0, \ t \ge 0 \right\}$

R^3, $\{(x_1, x_2, t) \mid (x_1^2 + x_2^2)^{1/2} \le t\}$

Operation preserve convexity

Intersection

- S_1, S_2 convex

$S_1 \cap S_2$ convex

eg.1

polyhedron is intersection of halfspaces & half-planes.

eg.2

$\bigcap_{z \neq 0} \{X \in S^n \mid \underbrace{z^T X z \ge 0}_{\text{linear}}\}$

$\therefore$ intersection of infinite number of halfspaces $\{z_1^T X z_1 \ge 0, \ \dots, \ z_\infty^T X z_\infty \ge 0$

Affine functions

- $f(x) = Ax + b$

S is convex

$f(S) = \{f(x) \mid x \in S\}$ is convex

$f^{-1}(S) = \{x \mid f(x) \in S\}$ is convex

e.g. • scaling

$\alpha S = \{\alpha x \mid x \in S\}$

• translation

$S + a = \{x + a \mid x \in S\}$

• projection

$T = \{x_1 \in R^m \mid (x_1, x_2) \in S, \ x_2 \in R^n\}$

• sum of two set

$S_1 + S_2 = \{x + y \mid x \in S_1, y \in S_2\}$

• linear function

$f(x_1, x_2) = x_1 + x_2$

$\rightarrow S_1 + S_2$

• partial sum

$S = (x, y_1 + y_2) \mid (x, y_1) \in S_1, (x, y_2) \in S_2\}$

• LMIs (linear matrix inequalities)

$A(x) = x_1 A_1 + \dots + x_n A_n \preceq B$

$B, A_i \in S^m$

$- \{x \mid A(x) \preceq B\}$

$= \{x \mid f(x) = B - A(x), \ f: R^n - S^m\}$

is convex

• hyperbolic cone

$\{x \mid x^T P x \le (c^T x)^2, \ c^T x \ge 0\}$

$P \in S_+^n$, $c \in R^n$

$\Leftrightarrow \{(z, t) \mid z^T z \le t^2, \ t \ge 0\}$

$\Leftrightarrow \{x \mid f(x) = (P^{1/2}x, c^T x)\}$

• ellipsoid

$\mathcal{E} = \{x \mid (x - x_c)^T P^{-1}(x - x_c) \le\}$

$P \in S_{++}^n$

$P = \{f(u) \mid f(u) = P^{1/2}u + x_c\}$

$\|u\|_2 \le 1$

Positive semidefinite cone

- $S^n = \{X \in R^{n \times n} \mid X = X^T\}$

dim $= n(n+1)/2$

$S_+^n = \{X \in S^n \mid X \succeq 0\}$

$S_{++}^n = \{X \in S^n \mid X \succ 0\}$

- for any $z \in R^n$

$z^T S_+^n z \ge 0$

eg.

$X = \begin{bmatrix} x_1 & x_2 \\ x_2 & x_3 \end{bmatrix} \in S_+^2$

$\Updownarrow$

$x_1 \ge 0$ $x_3 \ge 0$

$x_1 x_3 \ge x_2^2$

proper cones & generalized inequalities

- A cone $K \subseteq R^n$, proper cone:
  - K is convex
  - K is closed
  - K is solid, nonempty interior
  - K is pointed (no line $x \in K$, $x \in -K$, $x = 0$
- proper cone induce generalized inequality.

$x \preceq_K y \Leftrightarrow y - x \in K$

$x \prec_K y \Leftrightarrow y - x \in \text{int } K$

- properties:
  - $x \preceq_K y$, $u \preceq_K v \rightarrow x + u \preceq_K y + v$
  - $x \preceq_K y$, $y \preceq_K z \rightarrow x \preceq_K z$
  - $x \preceq_K y$, $\alpha \ge 0$, $\alpha x \preceq_K \alpha y$
  - $x \preceq_K x$
  - $x \preceq_K y$, $y \preceq_K x$, $x = y$
  - $x_i \preceq_K y_i$ $i = 1, 2, \dots$ $x_i \rightarrow x$, $y_i \rightarrow y$, $i \rightarrow \infty$

$x \preceq_K y$

minimum & minimal elements

- minimum: $x \in S$, every y $x \preceq_K y$; no point is "more" than it
- minimal: $x \in S$, $y \in S$, $y \preceq_K x$ only if $y = x$; no point is "less" than it

- minimum x

$x \in S$, x minimum iff $S \subseteq x + K$

- minimal x

$x \in S$, x minimal iff $(x - K) \cap S = \{x\}$

separating hyperplane theorem

- $C \cap D = \emptyset$

$\exists$ $a^T x \le b$, $x \in C$

$a^T x \ge b$, $x \in D$

$\{x \mid a^T x = b\} \rightarrow$ separating hyperplane

- $C \subseteq R^n$

$x_0 \in \text{bd } C$

$= \text{cl } C \setminus \text{int } C$

$a^T x \le a^T x_0$

$\forall x \in C$

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

$\downarrow$

supporting hyperplane

$\downarrow$

$\forall$ any $\text{int } C \neq \emptyset$,

C is convex,

$x_0 \in \text{bd } C$

$\exists$ hyperplane to C @ x_0

Linear-fractional & perspective functions

① perspective function $P: R^{n+1} \rightarrow R^n$

dom $P = R^n \times R_{++}$

$\Rightarrow$

$P(z, t) = \frac{z}{t}$

- $C \subseteq \text{dom } P$, convex

$\Rightarrow P(C) = \{P(x) \mid x \in C\}$, convex

$\rightarrow P^{-1}(C) = \{(x, t) \in R^{n+1} \mid x/t \in C, \ t > 0\}$ convex

② Linear-fractional function

see review & examples

$g: R^n \rightarrow R^{m+1}$ (affine)

$g(x) = \begin{bmatrix} A \\ c^T \end{bmatrix} x + \begin{bmatrix} b \\ d \end{bmatrix}$

$f: R^n \rightarrow R^m$, $f = P \circ g$

$f(x) = \frac{(Ax + b)}{c^T x + d}$

$\text{dom } f = \{x \mid c^T x + d > 0\}$

image, inverse image are all convex

Show that maximum of a convex function f over the polyhedron P $\text{conv}\{v_1 \dots v_k\}$ is achieved at one of its vertices

$\sup_{x \in P} f(x) = \max f$

Dual generalized inequalities

$\preceq_{K^*}$: dual generalized inequality of $\preceq_K$

- $x \preceq_K y$ iff $\lambda^T x \le \lambda^T y$ $\forall \lambda \succeq_{K^*} 0$
- $x \prec_K y$ iff $\lambda^T x < \lambda^T y$ $\forall \lambda \succeq_{K^*} 0$, $\lambda \neq 0$
- $\lambda \preceq_{K^*} \mu$ iff $\lambda^T x \le \mu^T x$, $\forall x \succeq_K 0$

- Dual cone always convex

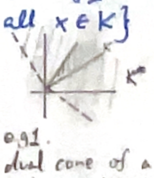
exam 2.31

Dual Cones

K is a cone

K^* is a dual cone

$K^* = \{y \mid x^T y \ge 0 \text{ for all } x \in K\}$



e.g.1

dual cone of a subspace $V \subseteq R^n$:

orthogonal complement

$V^\perp = \{y \mid y^T v = 0, \text{ for all } v \in V\}$

e.g.2 R^2

nonnegative orthant

$y^T x \ge 0$ $\forall$ all $x \succeq 0 \Leftrightarrow y \succeq 0$

self dual

- K^* is closed & convex
- $K_1 \subseteq K_2 \rightarrow K_2^* \subseteq K_1^*$
- $\text{cl } K$ is pointed, $\text{int } K^* \neq \emptyset$
- $\text{int } K \neq \emptyset$, K^* is pointed
- K^{**}: closure of conv K
- K proper cone, K^* also proper

$K^{**} = K$

e.g 3 Positive semidefinite cone

$K = S_+^n = \{X \in S^n \mid z^T X z \ge 0, \ \forall z \in R^n\}$

$K^* = \{Y \mid \text{tr}(X^T Y) \ge 0, \ \forall X \in S_+^n\}$

Show $\begin{cases} Y \notin K \rightarrow Y \notin K^* \\ Y \in K \rightarrow Y \in K \end{cases}$

proof: self-dual

if $Y \notin K$ $(Y \notin S_+^n) \Rightarrow \exists z \mid z^T Y z < 0$, $\text{tr}(z z^T Y) < 0$, $\text{tr}(Y) < 0$

if $Y \in K$ $(Y \in S_+^n) \Rightarrow \text{tr}(XY) \ge 0$? for $X \in S_+^n$

$\Rightarrow \text{tr}(XY) = \text{tr}(X[\sum_{i=1}^n \lambda_i v_i v_i^T]) = \text{tr}(\sum \lambda_i X v_i v_i^T) = \sum \lambda_i \text{tr}(X v_i v_i^T)$

Ch3 Convex functions

(Definition)

$f: \mathbb{R}^n \to \mathbb{R}$

if dom f is convex set

$x, y \in \text{dom } f$, $\theta \in [0,1]$

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

$\to (-f)$ convex, f concave

$\to f$ is convex, i.f.f.

$g(t) = f(x+tv)$ is convex

$\{t \mid x+tv \in \text{dom } f\}$ (on a line)

(1st-order condition)

f differentiable

f convex i.f.f.

dom f convex

$\to f(y) \ge f(x) + \nabla f(x)^T (y-x)$ (convex)

(2nd-order condition)

f differentiable twice

f convex i.f.f.

dom f convex

$\to \nabla^2 f(x) \succeq 0$ (positive semidefinite) (concave $\preceq 0$)

(Examples)

- → Exponential e^{ax} on $\mathbb{R}$ convex
- → Power x^a on $\mathbb{R}_{++}$, $a \ge 1$, $a \le 0$ convex; x^a -- $0 \le a \le 1$ concave
- → Power of absolute value $|x|^p$, $p \ge 1$
- → logarithm $\log x$ concave
- → negative entropy $x \log x$ convex

on $\mathbb{R}^n$

- → Norms convex
- → Max function convex: $f(x) = \max\{x_1, \dots, x_n\}$
- → Quadratic-over-linear function: $f(x) = x_1^2/x_2$, $x \in \mathbb{R}^2$, $x_2 > 0$ convex
- → log-sum-exp: $f(x) = \log(e^{x_1} + \dots + e^{x_n})$ convex
- → geometric mean: $f(x) = (\prod_{i=1}^n x_i)^{1/n}$ concave
- → log-determinant: $f(x) = \log \det(X)$ on dom $f = S^n_{++}$ concave

★ Method in Sum:

1. Check basic inequality
2. 2nd-order: Hessian Matrix
3. restrict to an abritrary line & verify convexity on $\mathbb{R}$, e.g. $g(t) = \log\det(Z + tV)$
4. ★ operations

(sublevel sets)

→ α-sublevel set: $f: \mathbb{R}^n \to \mathbb{R}$

$C_\alpha = \{x \in \text{dom } f \mid f(x) \le \alpha\}$ — C_α of convex func is convex

→ α-superlevel set:

$C_\alpha = \{x \in \text{dom } f \mid f(x) \ge \alpha\}$ — C_α of concave func is convex

(Epigraph) epi f

→ graph of $f: \mathbb{R}^n \to \mathbb{R}$

$= \{(x, f(x)) \mid x \in \text{dom } f\} \subseteq \mathbb{R}^{n+1}$

→ epigraph of $f: \mathbb{R}^n \to \mathbb{R}$ convex

$= \{(x, t) \mid x \in \text{dom } f, f(x) \le t\} \subseteq \mathbb{R}^{n+1}$

→ a function is convex i.f.f. epigraph is convex set.

→ hypograph of $f: \mathbb{R}^n \to \mathbb{R}$ concave

$= \{(x, t) \mid x \in \text{dom } f, t \le f(x)\} \subseteq \mathbb{R}^{n+1}$

(Jensen's inequality & extensions)

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

$\Rightarrow f(\theta_1 x_1 + \dots + \theta_k x_k) \le \theta_1 f(x_1) + \dots + \theta_k f(x_k)$, $\sum \theta = 1$

$\Rightarrow f(\int_S p(x) x\, dx) \le \int_S f(x) p(x)\, dx$

$\Rightarrow f(Ex) \le Ef(x)$

⇒ convex inequality:

prob$(x = x_1) = \theta$, prob$(x = x_2) = (1-\theta)$

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

(Cauchy-Schwarz inequality)

$(a^T a)(b^T b) \ge (a^T b)^2$

★ Operations that preserve convexity

(Nonnegative weighted sum)

→ $f = w_1 f_1 + \dots + w_m f_m$ is convex, given $f_1, \dots, f_m$ are convex

(Composition w/ affine mapping)

→ $f: \mathbb{R}^n \to \mathbb{R}$

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

$g: \mathbb{R}^m \to \mathbb{R}$

$g(x) = f(Ax + b)$, dom $g = \{x \mid Ax + b \in \text{dom } f\}$

$\mathbb{R}^{n\times m}\mathbb{R}^m + \mathbb{R}^n$

→ if f convex, g is convex

(Pointwise maximum)

→ $f(x) = \max\{f_1(x), \dots, f_m(x)\}$

e.g. $f(x) = \sum_{i=1}^r x_{[i]}$

$x_{[1]} \ge x_{[2]} \ge \dots \ge x_{[n]}$

$\Rightarrow f(x) = \sum_{i=1}^r x_{[i]}$

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

(pointwise maximum of $\frac{n!}{r!(n-r)!}$ linear functions)

(Pointwise supremum)

→ if $f(x,y)$ convex in x for each $y \in A$

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

convex

(Composition)

(w/ scalar function)

$h: \mathbb{R}^k \to \mathbb{R}$

$g: \mathbb{R}^n \to \mathbb{R}^k$

$f = h \circ g: \mathbb{R}^n \to \mathbb{R}$

→ $f(x) = h(g(x))$

dom $f = \{x \in \text{dom } g \mid g(x) \in \text{dom } h\}$

to determine convexity use $f''(x) = h''(g(x)) g'(x)^2 + h'(g(x)) g''(x)$

(w/ vector function)

→ $f(x) = h(g(x)) = h(g_1(x), \dots, g_k(x))$

$h: \mathbb{R}^k \to \mathbb{R}$

$g_i: \mathbb{R}^n \to \mathbb{R}$

dom $g = \mathbb{R}$

dom $h = \mathbb{R}^k$

to determine convexity use

$f''(x) = g'(x)^T \nabla^2 h(g(x)) g'(x) + \nabla h(g(x))^T g''(x)$.

- f convex: h convex, h non-decreasing, g convex
- f convex: h convex, h non-increasing, g concave
- f concave: h concave, h non-decreasing, g concave

(Minimization)

→ if $f(x,y)$ convex

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

(Perspective of a function)

$f: \mathbb{R}^n \to \mathbb{R}$.

$g: \mathbb{R}^{n+1} \to \mathbb{R}$, $g(x,t) = t f(x/t)$

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

→ f convex, g convex

→ $(x, t, s) \in \text{epi } g \iff t f(x/t) \le s$

$\iff f(x/t) \le s/t$

$\iff (x/t, s/t) \in \text{epi } f$

∴ epi $g \longleftrightarrow$ epi f is perspective mapping (operation preserve convex set)

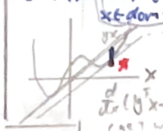
(Conjugate function)

$f: \mathbb{R}^n \to \mathbb{R}$

$f^*: \mathbb{R}^n \to \mathbb{R}$

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

what kind of x makes y^Tx sup.



$\frac{d}{dx}(y^T x - f(x)) = 0$

$y - f'(x) = 0$

$f'(x) = y$

occurs @ A

$f^*(y)$ = difference

difference of $y^T x$ & $f(x)$ $= f^*(y)$

f^* convex

pointwise supremum of affine of y

(log-concave / -convex)

$f: \mathbb{R}^n \to \mathbb{R}$

$f(x) > 0$, $x \in \text{dom } f$

$\log f$ convex

→ f is log-convex

→ $f(\theta x + (1-\theta) y) \ge f(x)^\theta f(y)^{1-\theta}$

$\log f(\theta x + (1-\theta) y) \ge \theta \log f(x) + (1-\theta) \log f(y)$

(properties)

f is twice differentiable

dom f convex

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

f is log-convex, $x \in \text{dom } f$

→ $f(x) \nabla^2 f(x) \succeq \nabla f(x) \nabla f(x)^T$

→ multiplication, addition, integration } log-convex preserve

(Quasiconvex definition)

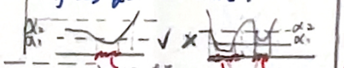
$f: \mathbb{R}^n \to \mathbb{R}$

→ $S_\alpha = \{x \in \text{dom } f \mid f(x) \le \alpha\}$ convex

$f \to$ quasiconvex

→ $S_\alpha = \{x \in \text{dom } f \mid f(x) \ge \alpha\}$ convex

$f \to$ quasiconcave.



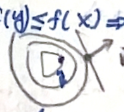
− $f(\theta x + (1-\theta) y) \le \max\{f(x), f(y)\}$

$(x, f(x))$ $(y, f(y))$

(1st order condition for quasiconvex func.)

$f: \mathbb{R}^n \to \mathbb{R}$, dom f convex, $x, y \in \text{dom } f$

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



(2nd-order condition for quasiconvex func.)

if $y^T \nabla f(x) = 0$

then $y^T \nabla^2 f(x) y \ge 0$

(operation preserve quasiconvexity)

→ nonnegative weighted maximum

$f(x) = \max\{w_1 f_1, \dots, w_m f_m\}$

$w_i \ge 0$, f_i quasiconvex

→ pointwise supremum

$f(x) = \sup_{y \in C} (w(y) g(x,y))$, $w(y) \ge 0$, g quasiconvex

→ composition

$g: \mathbb{R}^n \to \mathbb{R}$ quasiconvex

$h: \mathbb{R} \to \mathbb{R}$ nondecreasing

$f = h \circ g$ is quasiconvex

→ minimization

$f(x,y)$ is quasiconvex jointly

C convex

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

(Convexity w.r.t. GI)

$K \subseteq \mathbb{R}^m$ proper cone induces $\preceq_K$

$f: \mathbb{R}^n \to \mathbb{R}^m$ K-convex

$f(\theta x + (1-\theta) y) \preceq_K \theta f(x) + (1-\theta) f(y)$

(Dual characteristic of K-convexity)

f is K-convex i.f.f. for every $w \succeq_{K^*} 0$

$w^T f$ is convex

(Differentiable K-convex function)

$f(y) \succeq_K f(x) + Df(x)(y - x)$

Some exercise:

① conjugate functions: $f^*(y) = \sup(y^T x - f(y))$

a. $f(x) = -\log x$, dom $f = \mathbb{R}_{++}$

$f^*(y) = \sup(y^T x + \log x)$

if $y \ge 0$, $f^*(y) \to \infty$

if $y < 0$, $\frac{d}{dx}(y^T x + \log x) = 0$

hv sup when $x = \frac{-1}{y}$

$f^*(y) = \begin{cases} \infty & y \ge 0 \\ -1 + \log(\frac{-1}{y}) & \text{o.w.} \end{cases}$

b. $f(x) = e^x$

$f^*(y) = \sup(y^T x - e^x)$

if $y < 0$, then $f^*(y) \to \infty$

if $y \ge 0$, $\frac{d}{dx}(y^T x - e^x) = 0$

hv sup when $x = \log y$

$f^*(y) = \begin{cases} \infty & y < 0 \\ y \log y - y & \text{o.w.} \end{cases}$

c. $f(X) = \log \det X^{-1}$ on S^n_{++}

$f^*(Y) = \sup_{X \succ 0} (\text{tr}(YX) + \log \det X)$ (inner product of matrix S^n)

$\text{tr}(YX) + \log\det X$ unbounded above unless $Y \prec 0$:

→ if $Y \nprec 0$, Y has v, $\|v\|_2 = 1$, eigenvalue $\lambda \ge 0$

$X = I + tvv^T$,

$\text{tr}(YX) + \log\det X$

$= \text{tr}\, Y + t\lambda + \log\det(I + tvv^T)$

$= \text{tr}\, Y + t\lambda + \log(1+t)$

$t \to \infty$

→ if $Y \prec 0$,

$\nabla_X(\text{tr}(YX) + \log\det X)$

$= Y + X^{-1} = 0$

$X = -Y^{-1}$

$f^*(Y) = \log\det(-Y)^{-1} - n$

dom $f^* = -S^n_{++}$

② convexity on arbitrary line

a. $f(X) = \log\det X$, $g(t) = \log\det(Z + tV)$

$= \log\det(Z^{1/2}(I + tZ^{-1/2} V Z^{-1/2}) Z^{1/2})$

$= \sum_{i=1}^n \log(1 + t\lambda_i) + \log\det Z$, $\lambda_i = \text{eig}(Z^{-1/2} V Z^{-1/2})$

$g'(t) = \sum_{i=1}^n \frac{\lambda_i}{1 + t\lambda_i}$, $g''(t) = -\sum_{i=1}^n \frac{\lambda_i^2}{(1 + t\lambda_i)^2} \le 0$

b. $f(X) = \text{tr}(X^{-1})$

$g(t) = \text{tr}((Z + tV)^{-1})$

let $Z^{-1/2} V Z^{-1/2}$ eigenvalue decomp.

$Z^{-1/2} V Z^{-1/2} = Q \Lambda Q^T$

$= \text{tr}[Z(I + tZ^{-1/2} V Z^{-1/2})]^{-1}$

$= \text{tr}(Z^{-1}(I + tQ\Lambda Q^T)^{-1})$

$= \text{tr}(Z^{-1}(QQ^T + tQ\Lambda Q^T)^{-1})$

$= \text{tr}(Z^{-1} Q (I + t\Lambda)^{-1} Q^T)$

$= \text{tr}(Q^T Z^{-1} Q (I + t\Lambda)^{-1})$

$= \sum_{i=1}^n [(Q^T Z^{-1} Q)]_{ii} (1 + t\lambda_i)^{-1}$

③ matrix convexity $f(X) = X^{-1}$

$g(X) = y^T X^{-1} y$ (if g is convex, $f(x)$ is convex)

epi $g = \{(X, t) \mid X \succ 0, y^T X^{-1} y \le t, \forall y \in \mathbb{R}^n\}$

$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0 \iff C - B^T A^{-1} B \succeq 0$

$t - y^T X^{-1} y \ge 0$ $\therefore \begin{bmatrix} X & y \\ y^T & t \end{bmatrix} \succeq 0$ LMI

epi g convex ↓ $g(x)$ convex → $f(X)$ convex

Ch4 Convex problem

(Basic terminology)

minimize $f_0(x)$
subject to $f_i(x) \le 0,\ i=1,\dots,m$
$h_i(x)=0,\ i=1,\dots,p$

(optimal value)

$p^* = \inf\{f_0(x) \mid f_i(x)\le 0,\ i=1\dots m,\ h_i(x)=0,\ i=1,\dots,p\}$ (no x satisfies)

$p^*=\infty$ if problem infeasible

$p^*=-\infty$ if problem unbounded below

D.T.W $D=\bigcap_{i=0}^{m} \operatorname{dom} f_i \cap \bigcap_{i=1}^{p} \operatorname{dom} h_i$

(optimal & locally optimal points)

$\rightarrow X_{opt}=\{x \mid f_i(x)\le 0,\ h_i(x)=0,\ f_0(x)=p^*\}$ $i=1,\dots,m,\ i=1,\dots,p$

$\rightarrow$ x @ $f_0(x)\le p^*+\epsilon$, ϵ-suboptimal

$\rightarrow f_0(x)=\inf\{f_0(z) \mid f_i(z)\le 0,\ h_i(z)=0$
if $\exists R>0$ $\quad \|z-x\|_2\le R\}$ (locally problem)

$\rightarrow$ optimal value attained / achieved $\rightarrow$ solvable

$\rightarrow$ optimal value infeasible ∞, unbounded $-\infty$

find x subject to $f_i(x)\le 0$ $h_i(x)=0$
$\rightarrow$ feasibility problem

implicit constraints $\quad x\in D=\bigcap_{i=0}^{m}\operatorname{dom} f_i \cap \bigcap_{i=1}^{p}\operatorname{dom} h_i$
explicit constraints $\quad h_i(x)=0 \quad f_i(x)\le 0$

(Convex Optimization)

minimize $f_0(x)$ convex
subject to $f_i(x)\le 0,\ i=1,\dots,m$ convex
$a_i^T x = b_i,\ i=1,\dots,p$ affine

feasible set $D=\bigcap_{i=0}^{m}\operatorname{dom} f_i$

$\rightarrow$ minimize a convex objective function over a convex set

(Local & global optima)

locally optima = global optima

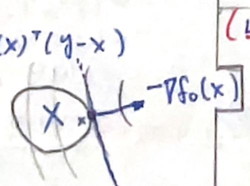
$\rightarrow$ x is locally optimal if x feasible
$f_0(x)=\inf\{f_0(z) \mid z \text{ feasible},\ \|z-x\|_2\le R\}$
$R>0$

$\rightarrow$ proof:
if x not globally optimal,
$\exists y\ \ f_0(y)<f_0(x),\ \|y-x\|_2>R$
also
$z=(1-\theta)x+\theta y \quad \theta=\frac{R}{2\|y-x\|_2} \rightarrow \|z-x\|_2 = R/2 < R$
$\therefore f_0(z)\le(1-\theta)f_0(x)+\theta f_0(y) < \theta f_0(x)$
$\Updownarrow$ contradicts
$f_0(x)=\inf\{f_0(z) \mid z \text{ feasible},\ \|z-x\|_2\le R\}$

(optimality criterion)

recall $f_0(y)\ge f_0(x)+\nabla f_0(x)^T(y-x)$

x is optimal i.f.f.
$\nabla f_0(x)^T(y-x)\ge 0$



(more on vector optimization)

Scalarization $\in R^n$

for any $\lambda \succeq_{K^*} 0$, if $\tilde{x}$ is an optimal point for the scalar optimization problem below

minimize $\lambda^T f_0(x)$
subject to $f_i(x)\le 0,\ i=1,\dots,m$
$h_i(x)=0,\ i=1,\dots,p$

$\rightarrow$ then $\tilde{x}$ is pareto optimal for the vector optimization problem

$\rightarrow$ for every pareto optimal point x^{po}, $\exists \lambda \succeq_{K^*} 0,\ \lambda\neq 0$, such that $\tilde{x}$ is an optimal point of scalarized problem

(equivalent convex problems)

(Eliminating equality constraints)

minimize$_x$ $f_0(x)$
subject to $f_i(x)\le 0 \quad i=1\dots m$
$h_i(x)=0 \quad i=1,\dots,p$
$\Updownarrow$
minimize $f_0(Fz+x_0)$
subject to $f_i(Fz+x_0)\le 0,\ i=1\dots m$

(introducing equality constraints)

minimize $f_0(A_0x+b_0)$
subject to $f_i(A_ix+b_i)\le 0,\ i=1\dots m$

minimize$_{x,y}$ $f_0(y_0)$
subject to $f_i(y_i)\le 0, \quad i=1\dots m$
$y_i=A_ix+b_i, \quad i=0,1\dots m$

(introducing slack variables for linear inequalities)

minimize $f_0(x)$
subject to $a_i^Tx\le b_i, \quad i=1,\dots,m$
$\Updownarrow$
minimize$_{x,s}$ $f_0(x)$
subject to $a_i^Tx+s_i=b_i,\ i=1\dots m$
$s_i\ge 0, \quad i=1\dots m$

(epigraph problem form)

minimize$_{x,t}$ t
subject to $f_0(x)-t\le 0$
$f_i(x)\le 0 \quad i=1,\dots,m$
$a_i^Tx=b_i \quad i=1,\dots,p$

(minimizing over some variables)

minimize $f_0(x_1,x_2)$
subject to $f_i(x_1)\le 0, \quad i=1\dots m$
$\Updownarrow$
minimize $\tilde{f}_0(x_1)$
subject to $f_i(x_1)\le 0, \quad i=1,\dots m$
where $\tilde{f}_0(x_1)=\inf_{x_2} f_0(x_1,x_2)$

(Quasiconvex function)

minimize $f_0(x)$
subject to $f_i(x)\le 0 \quad i=1\dots m$
$Ax=b$
w/ $f_0: R^n\rightarrow R$ quasiconvex
$f_1,\dots f_m$ convex

$f_0(x)\le t \iff \phi_t(x)\le 0$
t-sublevel set

formulate as feasibility problem

find x
subject to $\phi_t(x)\le 0$
$f_i(x)\le 0 \quad i=1,\dots,m$
$Ax=b$

$\rightarrow$ suppose f_0 is differentiable
Let X be the feasible set.
if $x\in X$ & $\nabla f_0(x)^T(y-x)>0$
$\forall$ all $y\in X\setminus\{x\}$:
x is optimal

(Linear Optimization Problem)

minimize c^Tx+d affine
subject to $Gx\preceq h$ affine
$Ax=b$ affine
(feasible set polyhedron)

linear-fractional program

minimize $f_0(x)$
subject to $Gx\preceq h$
$Ax=b$

$f_0(x)=\frac{c^Tx+d}{e^Tx+f} \quad \operatorname{dom} f_0=\{x \mid e^Tx+f>0\}$

is equivalent

$\rightarrow$ minimize c^Ty+dz
subject to $Gy\preceq hz$
$Ay=bz$
$e^Ty+dz=1$
$z\ge 0$

(Quadratic Program)

minimize $\frac{1}{2}x^TPx+q^Tx+r$
subject to $Gx\preceq h$
$Ax=b$

$P\in S_+^n$

(Quadratically constrained quadratic program)

minimize $\frac{1}{2}x^TP_0x+q_0^Tx+r_0$
subject to $\frac{1}{2}x^TP_ix+q_i^Tx+r_i\le 0$
$i=1,\dots,m$
$Ax=b \quad P_i\in S_+^n\ i=0\sim m$

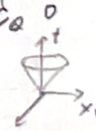
over a feasible region that's intersection of ellipsoids

(Second-order cone programming)

minimize f^Tx
subject to $\quad i=1\dots m$
$\|A_ix+b_i\|_2\le c_i^Tx+d_i$
$Fx=g$

SOCP could be interpreted as vector
$\begin{bmatrix} Ax+b \\ c^Tx+d\end{bmatrix}\in R^{k+1}$ lies in SOC
$\iff (Ax+b,\ c^Tx+d)\succeq_Q 0$

SOC: $k=2$
$\begin{bmatrix} x_1 \\ x_2 \\ t\end{bmatrix} \iff \|x\|_2\le t$



(Geometric Programming)

$f: R^n\rightarrow R \quad \operatorname{dom} f=R_{++}^n$
$f(x)=cx_1^{a_1}x_2^{a_2}\cdots x_n^{a_n}$
$c>0,\ a_i\in R$
$\rightarrow$ monomial function

sum of monomials
$f(x)=\sum_{k=1}^{K} c_k x_1^{a_{1k}} x_2^{a_{2k}}\cdots x_n^{a_{nk}}$
$c_k>0$
$\rightarrow$ posynomial function

closed under (posynomial) { addition, multiplication, nonnegative scaling
(monomial) { multiplication, division

posynomial × monomial → posynomial
posynomial / monomial → posynomial

minimize $f_0(x)$ posy.
subject to $f_i(x)\le 1,\ i=1,\dots,m$ mono.
$h_i(x)=1,\ i=1,\dots,p$ mono.

$\operatorname{dom} f=R_{++}^n \quad x\succ 0$ (implicity)

transform to convex

let $y_i=\log x_i \quad x_i=e^{y_i}$
$\therefore f(x)=f(e^{y_1},\dots e^{y_n})$
$=c(e^{y_1})^{a_1}\cdots(e^{y_n})^{a_n}$
$=e^{a^Ty+b} \quad b=\log c$

$f(x)=\sum_{k=1}^{K} c_k x_1^{a_{1k}}x_2^{a_{2k}}\cdots x_n^{a_{nk}}$
$\Rightarrow f(x)=\sum_{k=1}^{K} e^{a_k^Ty+b_k}$ $\quad a_k=a_{1k},\dots a_{nk}$, $b_k=\log c_k$

minimize $\sum_{k=1}^{K_0} e^{a_{0k}^Ty+b_{0k}}$
subject to $\sum_{k=1}^{K_i} e^{a_{ik}^Ty+b_{ik}}\le 1,\ i=1,\dots,m$
$e^{g_i^Ty+h_i}=1,\ i=1,\dots,p$

minimize $\tilde{f}_0(y)=\log\left(\sum_{k=1}^{K_0} e^{a_{0k}^Ty+b_{0k}}\right)$
subject to $\tilde{f}_i(y)=\log\left(\sum_{k=1}^{K_i} e^{a_{ik}^Ty+b_{ik}}\right)\le 0 \quad i=1,\dots,m$
$\tilde{h}_i(y)=g_i^Ty+h_i=0 \quad i=1,\dots,p$

(Generalized inequality constraints)

minimize $f_0(x)$ $R^n\rightarrow R$
subject to $f_i(x)\preceq_{K_i} 0,\ i=1,\dots,m$
$Ax=b$

$R^n\rightarrow R^{k_i}$, K_i-convex; K_i → proper cone $\subseteq R^{k_i}$

(conic form problem)

minimize c^Tx
subject to $Fx+g\preceq_K 0$
$Ax=b$

- convex feasible set
- local optima = global
- differentiable, criterion holds

(semidefinite programming)

$K\in S_+^k$

minimize c^Tx
subject to $x_1F_1+\cdots+x_nF_n+G\preceq 0$ (LMI)
$Ax=b$

$G,F_1,\dots,F_n\in S^k$
$A\in R^{p\times n}$

(if they are all diagonal, SDP → LP)

LP→SDP

minimize c^Tx+d
subject to $Gx\preceq h$
$Ax=b$
$\Updownarrow$
minimize c^Tx+d
subject to $Gx-h\preceq 0$
$Ax=b$
$\Updownarrow$
minimize c^Tx+d
subject to $\operatorname{diag}(Gx-h)\preceq 0$
$Ax=b$

QP→SDP

minimize $\frac{1}{2}x^TPx+q^Tx+r$
subject to $Gx\preceq h$
$Ax=b$
$\Updownarrow$
minimize$_{t,x}$ $t+q^Tx+r$
subject to $\frac{1}{2}x^TPx\le t$
$Gx\preceq h$
$Ax=b$
$\Updownarrow$
minimize$_{t,x}$ $t+q^Tx+r$
subject to $\begin{bmatrix} 1 & \frac{1}{\sqrt{2}}(P^{\frac{1}{2}})^Tx \\ \frac{1}{\sqrt{2}}x^TP^{\frac{1}{2}} & tI\end{bmatrix}\succeq 0$
$\operatorname{diag}(Gx-h)\preceq 0$
$Ax=b$

QCQP→SDP

minimize $\frac{1}{2}x^TP_0x+q_0^Tx+r_0$
subject to $\frac{1}{2}x^TP_ix+q_i^Tx+r_i \quad i=1\sim m$
$Ax=b$
$\Updownarrow$
minimize$_{u,t,x}$ $t+q_0^Tx+r_0$
subject to $\frac{1}{2}x^TP_0x\le t$
$\frac{1}{2}x^TP_ix\le u_i \quad i=1\sim m$
$u_i+q_i^Tx+r_i\le 0 \quad i=1\sim m$
$Ax=b$
$\Updownarrow$
minimize$_{u,t,x}$ $t+q_0^Tx+r_0$
subject to $\begin{bmatrix} I & \frac{1}{\sqrt{2}}(P^{\frac{1}{2}})^Tx \\ \frac{1}{\sqrt{2}}x^TP^{\frac{1}{2}} & tI\end{bmatrix}\succeq 0$
$\begin{bmatrix} I & \frac{1}{\sqrt{2}}(P_i^{\frac{1}{2}})^Tx \\ \frac{1}{\sqrt{2}}x^TP_i^{\frac{1}{2}} & u_iI\end{bmatrix}\succeq 0$
$u_i+q_i^Tx+r_i \quad i=1\sim m$
$Ax=b$

SOCP→SDP

minimize f^Tx
subject to $\|A_ix+b_i\|_2\le c_i^Tx+d_i$
$Fx=g \quad i=1\sim m$
$\Updownarrow$
minimize f^Tx
subject to $\begin{bmatrix} (c_i^Tx+d_i)I & A_ix+b_i \\ A_ix+b_i & c_i^Tx+d_i\end{bmatrix}\succeq 0 \quad i=1\sim m$
$Fx=g$

(Vector optimization)

minimize w.r.t. K $f_0(x)$ $R^n\rightarrow R^q$
subject to $f_i(x)\le 0 \quad i=1\sim m$
$h_i(x)=0 \quad i=1\sim p$
$R^n\rightarrow R$ $\quad R^n\rightarrow R$

① $f_0(x)$ K-convex
② $f_1\sim f_m$ convex
③ h_i affine

$O=\{f_0(x) \mid \exists x\in D\}$ $\quad f_i(x)\le 0,\ h_i(x)=0$
$\subseteq R^q$ achievable objective values

$O\subseteq f_0(x^*)+K$ optimal value

$(f_0(x)-K)\cap O=\{f_0(x)\}$

$P\subseteq O\cap \operatorname{bd} O$ Pareto

Duality Ch5 Ch5! $g(\lambda,\nu)=\inf_{x\in D} L(x,\lambda,\nu)\le L(\tilde x,\lambda,\nu)\le f_0(\tilde x)$ (any $\tilde x$ feasible, (λ,ν) feasible)

Algorithms Ch9, 10, 11

The Lagrangian

minimize $f_0(x)$
subject to $f_i(x)\le 0$, $i=1,\dots,m$
$h_i(x)=0$, $i=1,\dots,p$

$L(x,\lambda,\nu)=f_0(x)+\sum_{i=1}^m \lambda_i f_i(x)+\sum_{i=1}^p \nu_i h_i(x)$
$\mathrm{dom}\,L = D\times R^m\times R^p$

The Lagrange dual function
$g(\lambda,\nu)=\inf_{x\in D} L(x,\lambda,\nu)$ (concave, affine)

Lower bounds
$g(\lambda,\nu)\le p^*\le f_0(\tilde x)\ \forall\lambda\succeq 0$
Dual feasible $\lambda\succeq 0$, $(\lambda,\nu)\in\mathrm{dom}\,g$

The Lagrange dual problem
maximize $g(\lambda,\nu)$
subject to $\lambda\succeq 0$

Weak duality
$d^*\le p^*$; p^*-d^*: optimal duality gap
best bound

Strong duality occurs when
- $d^*=p^*$
- SLATER'S condition
$x\in\mathrm{relint}\,D$
such that
$f_i(x)<0$, $Ax=b$ (strictly feasible)
(but affine inequalities could just be feasible, i.e. $f_i(x)\le 0$)

Optimality conditions
- Complementary slackness

$f_0(x^*)=g(\lambda^*,\nu^*)$ (strong duality)
$=\inf_x\left(f_0(x)+\sum_{i=1}^m\lambda_i^* f_i(x)+\sum_{i=1}^p \nu_i^* h_i(x)\right)$
① x^* is one of the minimizer of $L(x,\lambda^*,\nu^*)$
$\le f_0(x^*)+\sum_{i=1}^m\lambda_i^* f_i(x^*)+\sum_{i=1}^p\nu_i^* h_i(x^*)$ (≥ 0, ≤ 0, $=0$)
$\le f_0(x^*)$

② $\sum_{i=1}^m \lambda_i^* f_i(x^*) \Rightarrow \lambda_i^* f_i(x^*)=0$
③ $\lambda_i^*>0 \Rightarrow f_i(x^*)=0$
$f_i(x^*)<0 \Rightarrow \lambda_i^*=0$

- KKT optimality conditions
$\rightarrow x^*$ is one of the minimizer of $L(x,\lambda^*,\nu^*)$
$\Rightarrow \nabla f_0(x^*)+\sum_{i=1}^m\lambda_i^*\nabla f_i(x^*)+\sum_{i=1}^p\nu_i^*\nabla h_i(x^*)=0$
$\Rightarrow$ KKT conditions:
$f_i(x^*)\le 0$
$h_i(x^*)=0$
$\lambda_i^*\ge 0$
$\lambda_i^* f_i(x^*)=0$
$\nabla f_0(x^*)+\sum_{i=1}^m\lambda_i^*\nabla f_i(x^*)+\sum_{i=1}^p\nu_i^*\nabla h_i(x^*)=0$

- strong duality holds
- x^* primal, (λ^*,ν^*) dual, must satisfy KKT!

Perturbation & Sensitivity analysis

minimize $f_0(x)$
subject to $f_i(x)\le u_i$
$h_i(x)=v_i$

$p^*(u,v)=\inf\{f_0(x)\mid f_i(x)\le u_i,\ h_i(x)=v_i\}$
$u\in R^m$, $v\in R^p$

△ assume strong duality holds
(λ^*,ν^*) dual optimal of unperturbed problem
$\Rightarrow p^*(0,0)=g(\lambda^*,\nu^*)\le f_0(x)+\sum_{i=1}^m\lambda_i^* f_i(x)+\sum_{i=1}^p \nu_i^* h_i(x)$
$\le f_0(x)+\lambda^{*T}u+\nu^{*T}v$
$\Rightarrow p^*(0,0)\le f_0(x)+\lambda^{*T}u+\nu^{*T}v$
$\Rightarrow f_0(x)\ge p^*(0,0)-\lambda^{*T}u-\nu^{*T}v$
$\Rightarrow p^*(u,v)\ge p^*(0,0)-\lambda^{*T}u-\nu^{*T}v$

Sensitivity (global)
$p^*(u,v)\ge p^*(0,0)-\lambda^{*T}u-\nu^{*T}v$
- λ_i^* large, $u_i<0$, $p^*(u,v)\uparrow$
- ν_i^* large$^+$, $v_i<0$; large$^-$, $v_i>0$ } $p^*(u,v)\uparrow$

Sensitivity (local)
△ $p^*(u,v)$ is a function of u,v (perturbed values)
△ λ^*,ν^* (optimal dual variables): $\nabla p^*(u,v)$ @ $(u,v)=(0,0)$
$\lambda_i^*=-\frac{\partial p^*(0,0)}{\partial u_i}$ $\left(\partial p^*(0,0)=-\lambda_i^*\partial u_i\right)$
$\nu_i^*=-\frac{\partial p^*(0,0)}{\partial v_i}$ $\left(\partial p^*(0,0)=-\nu_i^*\partial v_i\right)$

Duality w/ Generalized Inequalities

minimize $f_0(x)$
subject to $f_i(x)\preceq_{K_i} 0$ (proper cone)
$h_i(x)=0$
$K_i\subseteq R^{k_i}\rightarrow$ proper cone

$L(x,\lambda,\nu)=f_0(x)+\sum_{i=1}^m\lambda_i^T f_i(x)+\sum_{i=1}^p \nu_i h_i(x)$
$\lambda_i\in R^{k_i}$ vector
$\lambda_i\succeq_{K_i^*} 0$ (dual cone, nonnegative)
$g(\lambda,\nu)=\inf_{x\in D} L(x,\lambda,\nu)$

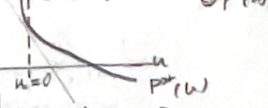
dual problem w/ G.I.
maximize $g(\lambda,\nu)$
subject to $\lambda_i\succeq_{K_i^*} 0$
if $\lambda_i\succeq_{K_i^*} 0$
$f_i(\tilde x)\preceq_{K_i} 0$
via duality
$\lambda_i^T f_i(\tilde x)\le 0$
$\therefore f_0(\tilde x)+\sum_{i=1}^m\lambda_i^T f_i(\tilde x)+\sum_{i=1}^p\nu_i h_i(\tilde x)\le f_0(\tilde x)$
$\tilde x$ feasible, $\lambda_i\succeq_{K_i^*}0$ ↓ inf
$g(\lambda,\nu)\le p^*$
above shows weak duality
strong duality $d^*=p^*$
Slater's condition
$x\in\mathrm{relint}\,D$
such that $f_i(x)\prec_{K_i} 0$, $Ax=b$

Optimality conditions
- complementary slackness
$f_0(x^*)=g(\lambda^*,\nu^*)$ ((x^*,λ^*,ν^*) optimal pts)
$\le f_0(x^*)+\sum_{i=1}^m\lambda_i^{*T}f_i(x^*)+\sum_{i=1}^p\nu_i^* h_i(x^*)$
$\le f_0(x^*)$
① x^* is one of the minimizer of $L(x,\lambda^*,\nu^*)$
② $\sum\lambda_i^{*T}f_i(x^*)=0$
③ $\lambda_i^*\succ 0 \Rightarrow f_i(x^*)=0$
$f_i(x^*)\prec_{K_i} 0\Rightarrow\lambda_i^*=0$
K dual cone, K proper cone (it is possible to have $\lambda_i^*\ne 0$, $f_i(x^*)\ne 0$ at the same time)

- KKT optimality conditions
$\rightarrow x^*$ is one of the minimizer of $L(x,\lambda^*,\nu^*)$
$\therefore \nabla f_0(x^*)+\sum_{i=1}^m Df_i(x^*)^T\lambda_i^*+\sum_{i=1}^p\nu_i^*\nabla h_i(x^*)=0$
$Df_i(x^*)\in R^{k_i\times n}$ = Jacobian
$\therefore$ KKT conditions:
$f_i(x^*)\preceq_{K_i}0$, $\lambda_i^*\succeq_{K_i^*}0$
$h_i(x^*)=0$, $\lambda_i^{*T}f_i(x^*)=0$
$\nabla f_0(x^*)+\sum_{i=1}^m Df_i(x^*)^T\lambda_i^*+\sum_{i=1}^p\nu_i^*\nabla h_i(x^*)=0$
similarly: strong duality holds, x^*,λ^*,ν^* optimal pts satisfy KKT

Perturbation & sensitivity analysis
minimize $f_0(x)$
subject to $f_i(x)\preceq_{K_i} u_i$
$h_i(x)=v_i$
$u_i\in R^{k_i}$, $v\in R^p$
△ assume strong duality holds
(λ^*,ν^*) dual optimal of unperturbed problem
△ $p^*(u,v)\ge p^*(0,0)-\sum_i\lambda_i^{*T}u_i-\nu^{*T}v$
global sensitivity
$\lambda_i^*=-\nabla_{u_i}p^*(0,0)$ (local sensitivity)
vector here

Sensitivity graphic presentation
① $p^*(u)\ge p^*(0)-\lambda^{*T}u$



Convergences for Newton's method
- Assumptions:
1) $\forall x$: $mI\preceq\nabla^2 f(x)\preceq MI$, $0<m\le M$
2) $\nabla^2 f$ is Lipschitz continuous w/ constant $L>0$
i.e., $\|\nabla^2 f(x)-\nabla^2 f(y)\|_2\le L\|x-y\|_2$

- convergence results
1) Damped Newton phase
△ if k in this phase, we have $f(x^{(k+1)})-f(x^{(k)})\le-\gamma$
$\gamma>0$, depends on α,β,m,M,L
△ $K\le\frac{f(x^{(0)})-p^*}{\gamma}$
2) Quadratically converge rate
△ enter @ k
△ $l\ge k$, $f(x^{(l)})-p^*\le\frac{2m^3}{L^2}\left(\frac12\right)^{2^{l-k+1}}$

usually ~ 6
$\therefore$ total k: $\frac{f(x^{(0)})-p^*}{\gamma}+6$

Unconstrained minimization

Descent method (General)
① (given) $x^{(0)}\in\mathrm{dom}\,f$
② (repeat) 1) determine a descent direction Δx
2) line search: choose stepsize $t\ge 0$
3) update $x^{(k+1)}=x^{(k)}+t^{(k)}\Delta x^{(k)}$ (search length, search direction)
③ (until) $\|\nabla f(x)\|_2\le\eta$ (often e.g.)

Exact line search
$t=\arg\min_{s\ge 0} f(x+s\Delta x)$

Backtracking line search
① (given) $\Delta x^{(k)}$ @ $f(x^{(k)})$, $x^{(k)}\in\mathrm{dom}\,f$
- $\alpha\in(0,0.5)$
- $\beta\in(0,1)$
② (set) $t:=1$
③ (while) $f(x+t\Delta x)>f(x)+\alpha t\nabla f(x)^T\Delta x$
(do) $t:=\beta t$

Gradient descent
① (given) $x^{(0)}\in\mathrm{dom}\,f$
② (repeat) a. $\Delta x:=-\nabla f(x)$
b. backtracking line search get $t^{(k)}$
c. $x^{(k+1)}:=x^{(k)}+t\Delta x^{(k)}$
③ (until) $\|\nabla f(x)\|_2\le\eta$

Newton's Method
① (given) $x^{(0)}\in\mathrm{dom}\,f$
② (repeat)
a. $\Delta x_{nt}^{(k)}:=-\nabla^2 f(x^{(k)})^{-1}\nabla f(x^{(k)})$
$\lambda^{(k)2}:=\nabla f(x^{(k)})^T\nabla^2 f(x^{(k)})^{-1}\nabla f(x^{(k)})$
b. (until) $\lambda^{(k)2}/2\le\epsilon$
c. backtracking line search get $t^{(k)}$
d. $x^{(k+1)}=x^{(k)}+t^{(k)}\Delta x^{(k)}$

Strong convexity
- $\forall m>0$, $f:R^n\to R$ & $\nabla^2 f(x)\succeq mI$, $\forall x$
$\hookrightarrow$ strong convexity
- reason for assuming strong convexity is for "Analysis"

FACT I
$f(y)\ge f(x)+\nabla f(x)^T(y-x)+\frac m2\|y-x\|_2^2$
proof
△ Taylor: $f(y)=f(x)+\nabla f(x)^T(y-x)+\frac12(y-x)^T\nabla^2 f(\theta x+(1-\theta)y)(y-x)$ ($\succeq mI$)
$\ge f(x)+\nabla f(x)^T(y-x)+\frac m2\|y-x\|_2^2$

FACT II
$f(x)\le p^*+\frac{1}{2m}\|\nabla f(x)\|_2^2$ (smaller $\nabla f(x)$ (closer to optimality))
proof
$p^*=f(x^*)\ge f(x)+\nabla f(x)^T(x^*-x)+\frac m2\|x^*-x\|_2^2$
$=f(x)+\frac m2\left(\|x^*-x\|_2^2+\frac2m(x^*-x)^T\nabla f(x)\right)$
$=f(x)+\frac m2\left(\|x^*-x+\frac1m\nabla f(x)\|_2^2-\frac{1}{m^2}\|\nabla f(x)\|_2^2\right)$
$\ge f(x)-\frac{1}{2m}\|\nabla f(x)\|_2^2$

FACT III
$\forall x$: $\|x^*-x\|_2\le\frac2m\|\nabla f(x)\|_2$
proof (from Cauchy)
$f(x^*)\ge f(x)+\nabla f(x)^T(x^*-x)+\frac m2\|x^*-x\|_2^2$
$\ge p^*-\|\nabla f(x)\|_2\|x^*-x\|_2+\frac m2\|x^*-x\|_2^2$
$\Rightarrow\frac m2\|x^*-x\|_2^2\le\|\nabla f(x)\|_2\|x^*-x\|_2$

FACT IV upper bound: (i) $\forall x,y$: $f(y)\le f(x)+\nabla f(x)^T(y-x)+\frac M2\|x-y\|_2^2$
(ii) $\forall x$: $p^*\le f(x)-\frac{1}{2M}\|\nabla f(x)\|_2^2$
proof (i) $f(y)=f(x)+\nabla f(x)^T(y-x)+\frac12(y-x)^T\nabla^2 f(\theta x+(1-\theta)y)(y-x)$ ($\preceq MI$)
$\le f(x)+\nabla f(x)^T(y-x)+\frac M2\|y-x\|_2^2$
(ii) $y=x-\frac1M\nabla f(x)$
$f(x-\frac1M\nabla f(x))\le f(x)+\nabla f(x)^T(-\frac1M\nabla f(x))+\frac M2\cdot\frac{1}{M^2}\|\nabla f(x)\|_2^2$
$=f(x)-\frac{1}{2M}\|\nabla f(x)\|_2^2$ (similar to FACT II)

Convergence Analysis
- Convergence of Gradient Descent w/ exact line search
△ assume $mI\preceq\nabla^2 f(x)\preceq MI$, $0<m\le M$
△ $f(x^{(k)})-p^*\le(1-\frac mM)^k(f(x^{(0)})-p^*)$
△ linear convergence rate
proof
△ $f(x^{(k)}-t\nabla f(x^{(k)}))\le f(x^{(k)})+\nabla f(x^{(k)})^T(-t\nabla f(x^{(k)}))+\frac M2 t^2\|\nabla f(x^{(k)})\|_2^2$
let $t=\frac1M$ we have $=f(x^{(k)})+(-t+\frac{Mt^2}{2})\|\nabla f(x^{(k)})\|_2^2$
$f(x^{(k)}-\frac1M\nabla f(x^{(k)}))\le f(x^{(k)})-\frac{1}{2M}\|\nabla f(x^{(k)})\|_2^2$
△ since exact line search
$t^{(k)}=\arg\min_{s\ge 0} f(x^{(k)}-s\nabla f(x^{(k)}))$
$\therefore f(x^{(k)}-t^{(k)}\nabla f(x^{(k)}))\le f(x^{(k)}-s\nabla f(x^{(k)}))$
$\therefore f(x^{(k+1)})\le f(x^{(k)})-\frac{1}{2M}\|\nabla f(x^{(k)})\|_2^2$ — ①
$\rightarrow f(x^{(k)})\le p^*+\frac{1}{2m}\|\nabla f(x)\|_2^2$ (from previous)
$\hookrightarrow\|\nabla f(x^{(k)})\|_2^2\ge 2m(f(x^{(k)})-p^*)$ — ②
from ①&② $f(x^{(k+1)})\le f(x^{(k)})-\frac mM(f(x^{(k)})-p^*)$
$\rightarrow f(x^{(k+1)})-p^*\le(1-\frac mM)f(x^{(k)})+(1-\frac mM)(-p^*)$
$\rightarrow f(x^{(k+1)})-p^*\le(1-\frac mM)(f(x^{(k)})-p^*)$

- Convergence of GD w/ back-tracking line search
$f(x^{(k)})-p^*\le C^k(f(x^{(0)})-p^*)$, $C=1-\min\{2m\alpha,\frac{2\beta\alpha m}{M}\}$
remark
△ from $mI\preceq\nabla^2 f(x)\preceq MI$ △ $1\le K(\nabla^2 f(x))=\frac{\lambda_{max}(\nabla^2 f(x))}{\lambda_{min}(\nabla^2 f(x))}\le\frac Mm$
$\hookrightarrow\lambda_{min}(\nabla^2 f(x))\ge m$
$\hookrightarrow\lambda_{max}(\nabla^2 f(x))\le M$ $\therefore$ if $M>>m$, converge slow

on Newton's method

- $\Delta x_{nt}=-\nabla^2 f(x)^{-1}\nabla f(x)$
- $\nabla f(x)^T\Delta x_{nt}=-\nabla f(x)^T\nabla^2 f(x)^{-1}\nabla f(x)<0$
minimizer of 2nd order approximation
- Taylor expansion (2nd order)
$\hat f(x+v)=f(x)+\nabla f(x)^Tv+\frac12 v^T\nabla^2 f(x)v$
$\Rightarrow$ minimized @ $\nabla_v\hat f(x+v)=0$
$\Rightarrow\nabla f(x)+\nabla^2 f(x)v=0\Rightarrow v=-\nabla^2 f(x)^{-1}\nabla f(x)$
different perspective: solution of linearized condition
- $\nabla f(x+v)$ (add v to point x, so that $\nabla f(x+v)$ close to zero)
- $\nabla f(x+v)\approx\nabla f(x)+\nabla^2 f(x)v=0$ (linear approx. of gradient)
$\rightarrow v=-\nabla^2 f(x)^{-1}\nabla f(x)=\Delta x_{nt}$ → from
Newton decrement
$\lambda(x)=[\nabla f(x)^T\nabla^2 f(x)^{-1}\nabla f(x)]^{0.5}$ $(\Delta x_{nt}^T\nabla^2 f(x)\Delta x_{nt})$
Justification of stopping criterion
- $f(x)-p^*=f(x)-\inf_y f(y)$ $f(y)\approx\hat f(y)$
- $\inf_y\hat f(y)=\inf_y\left(f(x)+\nabla f(x)^T(y-x)+\frac12(y-x)^T\nabla^2 f(x)(y-x)\right)$
$\hat f(y)$ inf @ $y=-\nabla^2 f(x)^{-1}\nabla f(x)+x$
$\inf_y\hat f(y)=f(x)-\frac12\nabla f(x)^T\nabla^2 f(x)^{-1}\nabla f(x)$
- plug back in

- LP duality
maximize c^Tx
s.t. $Ax\le b$, $x\ge 0$
⇕
minimize b^Ty
s.t. $A^Ty\ge c$, $y\ge 0$

Analysis for GD BLS
- $\tilde f(t)=f(x-t\nabla f(x))$
- $\tilde f(t)\le f(x)-\alpha t\|\nabla f(x)\|_2^2$
- $0\le t\le 1/M\rightarrow -t+\frac{Mt^2}{2}\le -\frac t2$
- $\hookrightarrow\tilde f(t)\le f(x)-t\|\nabla f(x)\|_2^2+\frac{Mt^2}{2}\|\nabla f(x)\|_2^2\le f(x)-(t/2)\|\nabla f(x)\|_2^2\le f(x)-\alpha t\|\nabla f(x)\|_2^2$
since $\alpha<1/2$
- BLS terminates $t=1$ or $t\ge\beta/M$
- gives lower bound on decrease in objective f
$\rightarrow f(x^+)\le f(x)-\alpha\|\nabla f(x)\|_2^2$
$f(x^+)\le f(x)-(\beta\alpha/M)\|\nabla f(x)\|_2^2$
$\rightarrow f(x^+)\le f(x)-\min\{\alpha,\beta\alpha/M\}\|\nabla f(x)\|_2^2$
$\rightarrow f(x^+)-p^*\le f(x)-p^*-\min\{\alpha,\beta\alpha/M\}\|\nabla f(x)\|_2^2$
- combine $\|\nabla f(x)\|_2^2\ge 2m(f(x)-p^*)$
$\rightarrow$ get:
$f(x^{(k)})-p^*\le C^k(f(x^{(0)})-p^*)$
$C=1-\min\{2m\alpha,2\beta\alpha m/M\}<1$

Sublinear convergence rate for GD w/o strong convexity
- $\|\nabla f(x)-\nabla f(y)\|_2\le L\|x-y\|_2$
- $x^{(k+1)}=x^{(k)}-t\nabla f(x^{(k)})$
$\rightarrow\forall k$: $f(x^{(k)})-p^*\le\frac{\|x^{(0)}-x^*\|_2^2}{2tk}=O(\frac1k)$

Algorithms (contd) Ch 9.10.11 **Ch 9.10.11!**

equality constrained minimization

minimize $f(x)$ $f: R^n \to R$ — ①
subject to $Ax=b$ $A \in R^{p \times n}$ w/ rank $A = p$

- suppose Slater's conditions hold
  i.e., $\exists x \in$ relint D such that $Ax=b$
- $\tilde{x}$ is optimal i.f.f. $\exists \tilde{\nu}$, such that $(\tilde{x}, \tilde{\nu})$ satisfies KKT!

$$\begin{cases} A\tilde{x}=b \\ \nabla f(\tilde{x})+A^T\tilde{\nu}=0 \end{cases} \quad — ②$$

solve ① = solve ②

Newton's method for equality constraints

- Derivation of Newton Step

① @ x
- addition of v: $x+v$
- $x+v$ should be feasible
  $A(x+v)=b$
  $Av=0$

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

from ① & ② we get

minimize $f(x)+\nabla f(x)^Tv+\frac{1}{2}v^T\nabla^2 f(x)v$
subject to $Av=0$ — ③

③ $\tilde{v}$ is optimal i.f.f. $\exists w$, such that $(\tilde{v}, w)$ satisfies KKT!

$$\begin{cases} A\tilde{v}=0 \\ \nabla f(x)+\nabla^2 f(x)\tilde{v}+A^Tw=0 \end{cases}$$

$$\Leftrightarrow \underbrace{\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix}}_{\text{KKT matrix}} \begin{bmatrix} \tilde{v} \\ w \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix} \quad — ④$$

$\Delta x_{nt} = \tilde{v}$

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

- Derivation of New Step (Again!)

① @ x
- addition of v: $x+v$
- $(x+v, w)$ satisfies ② (KKT)!

$$\begin{cases} A(x+v)=b \\ \nabla f(x+v)+A^Tw=0 \end{cases}$$

Taylor 1-st order linearization

$$\Leftrightarrow \begin{cases} A(x+v)=b \\ \nabla f(x)+\nabla^2 f(x)v+A^Tw=0 \end{cases}$$

$$\Leftrightarrow \begin{cases} Av=0 \\ \nabla f(x)+\nabla^2 f(x)v+A^Tw=0 \end{cases}$$

$$\Leftrightarrow \begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} v \\ w \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix}$$

$\Delta x_{nt} = \tilde{v}$

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

The Newton decrement
- pretty similar to unconstrained case
- $\lambda(x) = (\Delta x_{nt}^T \nabla^2 f(x) \Delta x_{nt})^{0.5}$

descent direction
- from ④
  $\nabla^2 f(x)\Delta x_{nt}+A^Tw = -\nabla f(x)$
- for descent direction
  $\nabla f(x)^T\Delta x_{nt} < 0$

$$\Rightarrow -(\nabla^2 f(x)\Delta x_{nt}+A^Tw)\Delta x_{nt}$$
$$= -\Delta x_{nt}^T\nabla^2 f(x)\Delta x_{nt} + A^Tw\Delta x_{nt}$$
$$= -\Delta x_{nt}^T\nabla^2 f(x)\Delta x_{nt} + 0$$
$$= -\lambda^2(x)$$
$$< 0$$

Justification of stopping criterion

$$f(x)-p^*$$
$$= f(x)-\inf_y\{f(y) \mid Ay=b\}$$
$$\Rightarrow f(x)-\inf_y\{\hat{f}(y) \mid Ay=b\}$$
$$= f(x)-\inf_v\{\hat{f}(x+v) \mid A(x+v)=b\}$$
$$= f(x)-\inf_v\{\hat{f}(x+v) \mid Av=0\}$$
$$= f(x)-\inf_v\{f(x)+\nabla f(x)^Tv+\tfrac{1}{2}v^T\nabla^2 f(x)v \mid Av=0\}$$

from ③ $v=\Delta x_{nt}$

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

Newton's method

1. (given) $x^{(0)} \in \mathbf{dom} f$, $Ax=b$
2. (repeat)
   a. $\Delta x_{nt}^{(k)} := \begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix}^{-1}\begin{bmatrix} -\nabla f(x) \\ w \end{bmatrix}$ (1:n, :)
   $\lambda^{(k)2} := \nabla f(x^{(k)})^T\nabla^2 f(x^{(k)})^{-1}\nabla f(x^{(k)})$
   b. (until) $(\lambda^{(k)})^2/2 \le \epsilon$
   c. backtracking line search get $t^{(k)}$
   d. $x^{(k+1)} := x^{(k)}+t^{(k)}\Delta x^{(k)}$

convergence analysis
similar to unconstrained

Interior Point Method

minimize $f_0(x)$ — ①
subject to $f_i(x) \le 0$ $i=1\ldots m$
$Ax=b$ $A \in R^{p \times n}$ rank$(A)=p$

- assume strictly feasible
  $\exists x \in D$ such that $\begin{cases} f_i(x)<0 & i=1\ldots m \\ Ax=b \end{cases}$
- ∴ transfer inequality to equality

Logarithmic Barrier

- $I_-(u) = \begin{cases} 0 & \text{if } u \le 0 \\ +\infty & \text{if } u>0 \end{cases}$
- minimize $f_0(x)+\sum_{i=1}^m I_-(f_i(x))$
  such that $Ax=b$ — ②
- define $\phi(x) = -\sum_{i=1}^m \log(-f_i(x))$
- where:
  $\nabla\phi(x) = \sum_{i=1}^m \frac{1}{-f_i(x)}\nabla f_i(x)$
  $\nabla^2\phi(x) = \sum_{i=1}^m \frac{1}{f_i(x)^2}\nabla f_i(x)\nabla f_i(x)^T + \sum_{i=1}^m \frac{1}{-f_i(x)}\nabla^2 f_i(x)$
- ∴ minimize $tf_0(x)+\phi(x)$ — ③
  subject to $Ax=b$
  $\phi(x) = -\sum_{i=1}^m \log(-f_i(x))$

Central Path

- $-\frac{1}{t}\phi(x)$, $t\uparrow$, closer to $I_-(u)$
- optimal solution of ③
  $\underset{\text{central point}}{x^*(t)}$
- central path: $\{x^*(t) \mid t>0\}$
- Question:
  How good is $x^*(t)$?

Ⓐ minimize $tf_0(x)+\phi(x)$ solution
subject to $Ax=b$
- $x^*(t) \to$ strictly feasible:
  $Ax^*(t)=b$ $f_i(x^*(t))<0$
- $\exists \hat{\nu} \in R^p$ such that
  $t\nabla f_0(x^*(t))+\nabla\phi(x^*(t))+A^T\hat{\nu}=0$

$$\Rightarrow t\nabla f_0(x^*(t))+\sum_{i=1}^m \frac{1}{-f_i(x^*(t))}\nabla f_i(x^*(t))+A^T\hat{\nu}=0$$

Ⓑ $\nabla f_0(x^*(t))+\sum_{i=1}^m \underbrace{\frac{1}{-tf_i(x^*(t))}}_{\lambda_i^*(t)}\nabla f_i(x^*(t))+A^T\underbrace{\frac{\hat{\nu}}{t}}_{\nu^*(t)}=0$

- we at central point
- we got dual feasible at $x^*(t)$ with $\lambda^*(t), \nu^*(t)$

$$\Rightarrow \nabla f_0(x^*(t))+\sum_{i=1}^m \lambda_i^*(t)\nabla f_i(x^*(t))+A^T\nu^*(t)=0$$
$$\Rightarrow x^*(t) = \arg\min_x L(x,\lambda,\nu) = f_0(x)+\sum_{i=1}^m \lambda_i f_i(x)+\nu^T(Ax-b)$$
$$\Rightarrow g(\lambda^*(t),\nu^*(t)) = f_0(x^*(t))+\sum_{i=1}^m \lambda_i^*(t)f_i(x^*(t))+\nu^*(t)^T(Ax^*(t)-b)$$
$$= f_0(x^*(t))+\sum_{i=1}^m \frac{1}{-tf_i(x^*(t))}f_i(x^*(t))+0 = f_0(x^*(t))-\frac{m}{t}$$

Ⓒ $g(\lambda^*(t),\nu^*(t)) = d^*(t) \le d^* \le p^*$

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

Barrier method

original problem:
minimize $f_0(x)$
subject to $f_i(x) \le 0$
$Ax=b$

⇓

new problem:
minimize $tf_0(x)+\phi(x)$
subject to $Ax=b$
$\phi(x) = -\sum_{i=1}^m \log(-f_i(x))$

1. (given) $x^{(0)}$ feasible
   $t^{(0)} := t^{(0)} > 0$
   $\mu > 1$
   $\epsilon > 0$
2. (repeat)
   a. solve $x^*(t)$ of $tf_0+\phi$ s.t. $Ax=b$ w/ $t^{(0)}$
   b. $x := x^*(t)$
   c. (until) $m/t < \epsilon$
   d. $t := \mu t$

hyperbolic constraints

- $x^Tx \le yz$ $y \ge 0$ $z \ge 0$
- $4x^Tx \le 4yz$ $y \ge 0$ $z \ge 0$
- $4x^Tx+y^2+z^2 \le y^2+z^2+4yz$
  $y \ge 0$ $z \ge 0$
- $4x^Tx+y^2+z^2-2yz \le y^2+z^2+2yz$
  $y \ge 0$ $z \ge 0$
- $4x^Tx+(y-z)^2 \le (y+z)^2$
  $y \ge 0$, $z \ge 0$
- $\left\| \begin{bmatrix} 2x \\ y-z \end{bmatrix} \right\|_2 \le y+z$
  $y \ge 0$, $z \ge 0$

Newton fitting

- minimize $-\sum_{i=1}^n \log(\alpha_i+x_i)$
  s.t. $x \succeq 0$ $1^Tx=1$
- KKT: $x^* \succeq 0$
  $1^Tx^*=1$
  $\lambda^* \succeq 0$
  $\lambda_i^*x_i^*=0$
  $-1/(\alpha_i+x_i^*)-\lambda_i^*+\nu^*=0$
- → $x^* \succeq 0$ $1^Tx^*=1$ $\nu^* \ge 1/(\alpha_i+x_i^*)$
  $x_i^*(\nu^*-1/(\alpha_i+x_i^*))=0$

solving x_i^*:
- if $\nu^* < 1/\alpha_i$: only $x_i^*>0$, $\nu^* = 1/(\alpha_i+x_i^*)$, $x_i^* = 1/\nu^*-\alpha_i$
- if $\nu^* \ge 1/\alpha_i$: $x_i^*>0$ impossible as $\nu^* \ge 1/\alpha_i > 1/(\alpha_i+x_i^*)$
  ↳ violate slackness
  ∴ $x_i^*=0$ if $\nu^* \ge 1/\alpha_i$

∴ $x_i^* = \begin{cases} 1/\nu^*-\alpha_i & \nu^*<1/\alpha_i \\ 0 & \nu^* \ge 1/\alpha_i \end{cases}$

- Matrix game.
  prob$(k=i)=u_i$
  prob$(l=i)=v_i$
  $\sum_{k=1}^n\sum_{l=1}^m u_kv_lP_{kl}$
  $= u^TPv$

player 2
$\sup\{u^TPv \mid v \succeq 0, 1^Tv=1\}$
$= \max_{i=1\ldots n}(P^Tu)_i$

player I
minimize $\max_{i=1\ldots m}(P^Tu)_i$
s.t. $u \succeq 0$, $1^Tu=1$

⇕

minimize t
s.t. $u \succeq 0$ $1^Tu=1$
$P^Tu \preceq t1$

$L() = t+\lambda^T(P^Tu-t1)-\mu^Tu+\nu(1-1^Tu)$
$= \nu+(1-1^T\lambda)t+(P\lambda-\nu 1-\mu)^Tu$

vice versa.

player I
$\inf\{u^TPv \mid u \succeq 0, 1^Tu=1\}$
$= \min_{i=1\ldots n}(Pv)_i$

player 2
maximize $\min_{i=1\ldots n}(Pv)_i$
s.t. $v \succeq 0$, $1^Tv=1$

⇕

$g(\lambda,\mu,\nu) = \begin{cases} \nu & 1^T\lambda=1 \;\; P\lambda-\nu 1=\mu \\ -\infty & \text{o.w.} \end{cases}$

maximize ν
s.t. $\lambda \succeq 0$ $1^T\lambda=1$ $\mu \succeq 0$
$P\lambda-\nu 1=\mu$

⇕

maximize ν
s.t. $\lambda \succeq 0$
$1^T\lambda=1$
$P\lambda \succeq \nu 1$

LO, Li-Yu
20997405
lloac@connect.ust.hk

Linear Algebra Basics :

4 ways to write a linear system

① $\begin{cases} 2x_1 + 3x_2 - 2x_3 = 7 \\ x_1 - x_2 - 3x_3 = 5 \end{cases}$ system of equation

② $\left(\begin{array}{ccc|c} 2 & 3 & -2 & 7 \\ 1 & -1 & -3 & 5 \end{array} \right)$ augmented matrix

③ $x_1 v_1 + x_2 v_2 + \cdots + x_n v_n = b$

$x_1 \begin{pmatrix} 2 \\ 1 \end{pmatrix} + x_2 \begin{pmatrix} 3 \\ -1 \end{pmatrix} + x_3 \begin{pmatrix} -2 \\ -3 \end{pmatrix} = \begin{pmatrix} 7 \\ 5 \end{pmatrix}$ vector equation

④ $Ax = b$ matrix equation

$\begin{pmatrix} 2 & 3 & -2 \\ 1 & -1 & -3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 7 \\ 5 \end{pmatrix}$

$Ax = b$ $\exists$ soln i.f.f
b is in the span of column of A

Schrug complement

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0$$
$$\Updownarrow$$
$$C - B^T A^{-1} B \succeq 0$$

Grammatrix

	a	b	c
a	a^2	ab	ac
b	ab	b^2	bc
c	ac	bc	c^2

for R^3

domain, range

Integration by change of variables.

$\int_0^x f(t)dt$ $\quad t = sx$ $\quad \int_0^1 f(sx)x\,ds$

$\int_{g(a)}^{g(b)} f(u)du$ $\quad u = g(x)$ $\quad \int_a^b f(g(x))g'(x)dx$

Matrix manipulation

$(AB)^{-1} = B^{-1}A^{-1}$
$(ABC\ldots)^{-1} = \ldots C^{-1}B^{-1}A^{-1}$
$(A^T)^{-1} = (A^{-1})^T$
$(A+B)^T = A^T + B^T$
$(AB)^T = B^T A^T$
$(ABC\ldots)^T = \ldots C^T B^T A^T$

$tr(A) = \sum_i A_{ii}$
$tr(A) = \sum_i \lambda_i \quad \lambda_i = eig(A)$
$tr(A) = tr(A^T)$
$tr(AB) = tr(BA)$
$tr(A+B) = tr(A) + tr(B)$
$tr(ABC) = tr(BCA) = tr(CAB)$
$a^T A = tr(aa^T)$

$det(A) = \prod_i \lambda_i, \lambda_i = eig(A)$
$det(cA) = c\,det(A)$
$det(A^T) = det(A)$
$det(AB) = det(A)det(B)$
$det(A^{-1}) = 1/det(A)$
$det(A^n) = det(A)^n$
$det(I + uv^T) = 1 + u^T v$

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

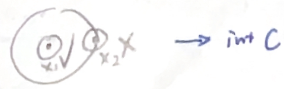
$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1}\frac{\partial f(x)}{\partial x_1} & \frac{\partial}{\partial x_1}\frac{\partial f(x)}{\partial x_2} & \cdots & \frac{\partial}{\partial x_1}\frac{\partial f(x)}{\partial x_n} \\ \frac{\partial}{\partial x_2}\frac{\partial f(x)}{\partial x_1} & \frac{\partial}{\partial x_2}\frac{\partial f(x)}{\partial x_2} & \ddots & \frac{\partial}{\partial x_2}\frac{\partial f(x)}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial}{\partial x_n}\frac{\partial f(x)}{\partial x_1} & \frac{\partial}{\partial x_n}\frac{\partial f(x)}{\partial x_2} & \cdots & \frac{\partial}{\partial x_n}\frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

PD matrix M

$eig(M)_i > 0$, symmetric

△ $B(x, \varepsilon) = \{ y \in R^n \mid \|y - x\|_2 \le \varepsilon \}$

interior point : $\exists \varepsilon > 0$
x $\quad$ s.t. $B(x, \varepsilon) \subset C$



△ limit point of C
if $\forall \varepsilon > 0$, excluding
$(B(x, \varepsilon) \setminus \{x\}) \cap C \ne \emptyset$
or
x is limit point of set S
if $\forall \epsilon > 0$, $\exists y \in S \setminus \{x\}$
w/ $d(x, y) < \epsilon$

△ closure
$cl(C) = C \cup L(C)$ includes bd (C)
$cl(C)$, closed
$cl(C)$, smallest closed set contains C
$C \subset S$,
$cl(C) \subset S$
set C is closed i.f.f
$C = cl(C)$

△ Boundary
▲ $bd\,C = cl(C) \setminus int(C)$
▲ $int(C) \subset C \subset cl(C)$
▲ C open i.f.f $C \cap bd\,C = \emptyset$
▲ C closed i.f.f. $bd(C) \subset C$

△ $f_0(x) = \frac{1}{2} x^T P x + q^T x + r$
$\nabla f_0(x) = \frac{1}{2}(P + P^T)x + q^T$

△ $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 3 & 6 \end{bmatrix}$ eg. of SPD

△ $det[\lambda I - A]$ polynomial → find eigenvalues
$det[\lambda B - A]$ 〃 → find generalized eigenvalues

linear-fractional programming

minimize $f_0(x) = \dfrac{c^T x + d}{e^T x + f}$ quasiconvex function $\quad dom\,f_0 = \{x \mid e^T x + f > 0\}$
subject to $Gx \preceq h$
$Ax = b$

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

minimize$_{y, z}$ $c^T y + dz$ $\qquad y = \dfrac{x}{e^T x + f} \quad z = \dfrac{1}{e^T x + f}$
subject to $Gy - hz \preceq 0$
$Ay - bz = 0$
$e^T y + fz = 1$
$z \ge 0$

convert QP to SOCP

minimize $x^T A x + a^T x$
subject to $Bx \le b$

$\Rightarrow$ minimize$_{x, y}$ $y + a^T x$
s.t. $Bx \le b$
$x^T A x \le y$

$\Rightarrow$ $y \ge x^T A x$
$\Rightarrow 0 \ge x^T A x - y$
$\Rightarrow 0 \ge 4x^T A x - 4y$
$\Rightarrow 0 \ge 4x^T A x + (1-y)^2 - (1+y)^2$
$\Rightarrow (1+y)^2 \ge 4x^T A x + (1-y)^2$

$\Rightarrow$ minimize$_{x, y}$ $y + a^T x$
s.t. $\left\| \begin{bmatrix} 2A^{\frac{1}{2}}x \\ 1 - y \end{bmatrix} \right\|_2 \le 1 + y$
$Bx \le b$

Unconstrained

Descent Method

step size
direction
search length

$\triangle$ $x^{(k+1)} = x^{(k)} + \underbrace{t^{(k)}}_{\text{search length}} \underbrace{\Delta x^{(k)}}_{\substack{\text{step (search direction)} \\ \text{search direction}}}$

$\triangle$ $f(x^{(k+1)}) < f(x^{(k)})$

$\triangle$ General Descent method

1. (given) $x^{(0)} \in \operatorname{dom} f$
2. (repeat)
   a. get descent direction Δx
      $\Delta x^{(k)}$ search direction
   b. line search
      $t^{(k)}$ search length
   c. update
      $x := x + t\Delta x$
3. (until)
   $\|\nabla f(x)\|_2 \leq \eta$ (oftenly)

1. Exact Line Search

$$t = \operatorname*{argmin}_{s \geq 0} f(x + s\Delta x)$$

2. Backtracking Line Search
   1. given
      - $\Delta x^{(k)}$ @ $f(x^{(k)})$
        $x^{(k)} \in \operatorname{dom} f$
      - $\alpha \in (0, 0.5)$
      - $\beta \in (0, 1)$
   2. $t := 1$
   3. (while)
      $f(x + t\Delta x) > f(x) + \alpha t \nabla f(x)^T \Delta x$
      (do)
      $t := \beta t$

Gradient Descent Method

1. (given) $x^{(0)} \in \operatorname{dom} f$
2. (repeat)
   a. $\Delta x^{(k)} = -\nabla f(x^{(k)})$
   b. backtracking line search get $t^{(k)}$
   c. $x^{(k+1)} := x^{(k)} + t\Delta x^{(k)}$
3. (until)
   $\|\nabla f(x)\|_2 \leq \eta$

Newton's Method

1. (given) $x^{(0)} \in \operatorname{dom} f$
2. (repeat)
   a. $\Delta x_{nt}^{(k)} := -\nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)})$
      $\lambda^{(k)2} := \nabla f(x^{(k)})^T \nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)})$
   b. (until) $\lambda^{(k)2}/2 \leq \epsilon$
   c. backtracking line search get $t^{(k)}$
   d. $x^{(k+1)} = x^{(k)} + t^{(k)} \Delta x^{(k)}$

Newton's method

equality constrained

1. (given) $x^{(0)} \in \text{dom} f$

   $Ax = b$

2. (repeat)

   a. $\Delta x_{nt}^{(k)} := \begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix}^{-1} \begin{bmatrix} -\nabla f(x) \\ w \end{bmatrix} (1:n, :)$

   $\lambda^{(k)2} := \nabla f(x^{(k)})^T \nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)})$

   b. (until) $\lambda^{(k)2}/2 \le \epsilon$

   c. backtracking line search get $t^{(k)}$

   d. $x^{(k+1)} = x^{(k)} + t^{(k)} \Delta x^{(k)}$

Newton's method (infeasible start)

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{pd} \\ \Delta \nu_{pd} \end{bmatrix} = - \begin{bmatrix} r_{dual} \\ r_{pri} \end{bmatrix} = - \begin{bmatrix} \nabla f(x) + A^T \nu \\ Ax - b \end{bmatrix} \quad \text{— } \#$$

1. (given) $x^{(0)} \in \text{dom} f$

   ν

   $\epsilon > 0$

   $\alpha \in (0, 0.5)$

   $\beta \in (0, 1)$

2. (repeat)

   a. $\Delta x_{nt}^{(k)} :=$ from $\#$

   $\Delta \nu_{nt}^{(k)} :=$ from $\#$

   b. backtracking line search on $\|r\|_2$

   1. $t := 1$
   2. while $\|r(x + t\Delta x_{nt}, \nu + t\Delta \nu_{nt})\|_2 > (1 - \alpha t)\|r(x, \nu)\|_2$

      $t := \beta t$

   c. $x := x + t\Delta x_{nt}$

   $\nu := \nu + t\Delta \nu_{nt}$

3. (until)

   $Ax = b$

   $\|r(x, \nu)\|_2 \le \epsilon$

Barrier method with logarithm

inequality constrained

original problem :

$$\begin{aligned} &\text{minimize} \quad f_0(x) \\ &\text{subject to} \quad f_i(x) \leq 0 \quad i=1,\dots,m \\ &\qquad\qquad\quad Ax=b \end{aligned}$$

new problem :

$$\begin{aligned} &\text{minimize} \quad t f_0(x) + \phi(x) \\ &\text{subject to} \quad Ax=b \end{aligned}$$

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

1. (given) $x^{(0)}$ (feasible)
   $t^{(0)} := t^{(0)} > 0$
   $\mu > 1$
   $\epsilon > 0$

2. (repeat)
   1. solve $x^*(t)$ of $t f_0 + \phi$ subject to $Ax=b$ with "t"
   2. $x := x^*(t)$
   3. (until) $m/t < \epsilon$
   4. $t := \mu t$

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

$$\nabla^2 \phi(x) = \sum_{i=1}^{m} \frac{1}{f_i(x)^2} \nabla f_i(x) \nabla f_i(x)^T + \sum_{i=1}^{m} \frac{1}{-f_i(x)} \nabla^2 f_i(x)$$

Phase I via infeasible start Newton method

inequality constrained w/ infeasible start.

original problem:

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \le 0, \quad i=1,\dots,m \\ & Ax = b \end{aligned}$$

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \le s, \quad i=1,\dots,m \\ & Ax = b \\ & s = 0 \end{aligned}$$

$$\begin{aligned} \text{minimize} \quad & t^{(0)} f_0(x) - \sum_{i=1}^{m} \log(s - f_i(x)) \\ \text{subject to} \quad & Ax = b \\ & s = 0 \end{aligned}$$