Ch2 Convex sets

Line/line segments
$y = \theta x_1 + (1-\theta)x_2$
$x_1 \neq x_2$, $\theta \in R$

Affine sets
$C \to$ affine
$x_1, \dots, x_k \in C$
$\theta_1 + \dots + \theta_k = 1$
$\to \theta_1 x_1 + \dots + \theta_k x_k \in C$
$\theta_i \in R$

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
$C = \{x \mid Ax = b\}$
$x_1 \in C$, $x_2 \in C$
$Ax_1 = b$, $Ax_2 = b$
$\Rightarrow A(\theta x_1 + (1-\theta)x_2)$
$= A\theta x_1 + A(1-\theta)x_2$
$= \theta b + (1-\theta)b$
$= b$
$\therefore \theta x_1 + (1-\theta)x_0 \in C$

$C \subseteq R^n$
affine hull
$\text{aff}\, C = \{\theta_1 x_1 + \dots + \theta_k x_k \mid x_1 \dots x_k \in C,\ \theta_1 + \dots + \theta_k = 1\}$
$C \subseteq S$, 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
$= \{x \in R^3 \mid -1 < x_1 < 1,\ -1 < x_2 < 1,\ x_3 = 0\}$

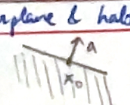
- if int $C \neq \emptyset$, relint C = int C
- all int $C \subset$ relint C
- relative boundary of C: cl $C \setminus$ relint C

convex sets
$C \to$ convex
$x_1, \dots x_k \in C$
$\theta_1 + \dots + \theta_k = 1$
$\theta_i \ge 0 \in R$
$\to \theta_1 x_1 + \dots + \theta_k x_k \in C$

$C \subseteq R^n$
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 $\mathcal{E}$ $\sqrt{\lambda_i}$, $A = P^{1/2}$
$P \in PD$ $P = P^T$

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

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

$\Rightarrow P = \{x \mid Ax \preceq b,\ Cx = d\}$
$A = \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix}$ $C = \begin{bmatrix} c_1^T \\ \vdots \\ c_p^T \end{bmatrix}$

eg. simplex
$= C = \text{conv}\{v_0 \dots v_k\}$
$v_0 \dots v_k \to$ affinely independent.

simplex $\to$ polyhedron representation

simplex:
$C = \text{conv}\{v_0 \dots v_k\}$
$= \{\theta_0 v_0 + \dots + \theta_k v_k \mid \theta \succeq 0,\ 1^T\theta = 1\}$
$x = \theta_0 v_0 + \theta_1 v_1 + \dots + \theta_k v_k$
$= (1 - \theta_1 - \theta_2 \dots - \theta_k)v_0 + v_1\theta_1 + \dots + v_k\theta_k$
$= v_0 + (v_1 - v_0)\theta_1 + (v_2 - v_0)\theta_2 + \dots + (v_k - v_0)\theta_k$
let $B = [v_1 - v_0 \cdots v_k - v_0]$
$y = [\theta_1, \dots \theta_k]$
$\therefore x = v_0 + By$
$AB = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} B = \begin{bmatrix} I \\ 0 \end{bmatrix}$
$A \to$ non-singular matrix exists

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

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.
$X = \begin{bmatrix} x_1 & x_2 \\ x_2 & x_3 \end{bmatrix} \in S_+^2$
$\Updownarrow$
$x_1 \ge 0$ $x_3 \ge 0$
$x_1 x_3 \ge x_2^2$


Operation preserve convexity

Intersection
- S_1, S_2 convex
$S_1 \cap S_2$ convex

eg.1
polyhedron is intersection of halfspaces & halfplanes.

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
$f(x_1, x_2) = x_1 + x_2$
$\to S_1 + S_2$
- 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)
$A(x) = x_1 A_1 + \dots + x_n A_n \preceq B$
$B, A_i \in S^m$
- $\{x \mid A(x) \preceq B\}$
$= \{x \mid f(x) = B - A(x)$
$f: R^n \to S^m\}$
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
$\mathcal{E} = \{x \mid (x - x_c)^T P^{-1}(x - x_c) \le 1\}$
$P \in S_{++}^n$
$P = \{f(u) \mid f(u) = P^{1/2}u + x_c,\ \|u\|_2 \le 1\}$

Linear-fractional & perspective functions
1. perspective function $P: R^{n+1} \to R^n$
dom $P = R^n \times R_{++}$
$\Rightarrow$
$P(z, t) = \dfrac{z}{t}$
- $C \subseteq \text{dom}\, P$, convex
$\to P(C) = \{P(x) \mid x \in C\}$, convex
$\to P^{-1}(C) = \{(x, t) \in R^{n+1} \mid x/t \in C,\ t > 0\}$ convex

2. Linear-fractional function
see remark & examples
$g: R^n \to R^{m+1}$ (affine)
$g(x) = \begin{bmatrix} A \\ c^T \end{bmatrix} x + \begin{bmatrix} b \\ d \end{bmatrix}$
$f: R^n \to R^m$, $f = P \circ g$
$f(x) = \dfrac{(Ax + b)}{c^T x + d}$
dom $f = \{x \mid c^T x + d > 0\}$

image
inverse image
are all convex

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.

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

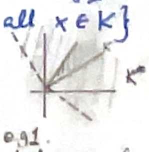
- properties:
  - $x \preceq_K y$, $u \preceq_K v \to x + u \preceq_K y + v$
  - $x \preceq_K y$, $y \preceq_K z \to x \preceq_K z$
  - $x \preceq_K y$, $\alpha \ge 0$, $\alpha x \preceq_K \alpha y$
  - $x \preceq_K x$
  - $x \preceq_K y$, $y \preceq_K x$, $x = y$
  - $x_i \preceq_K y_i$ $i = 1, 2, \dots$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

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

- 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
- $C \cap D = \emptyset$
$\exists\ a^T x \le b,\ x \in C$
$a^T x \ge b,\ x \in D$
$\{x \mid a^T x = b\} \to$ 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 int $C \neq \emptyset$,
C is convex,
$x_0 \in \text{bd}\, C$
$\exists$ hyperplane to C at x_0

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

eg2 R^2
nonnegative orthant
$y^T x \ge 0$ $\forall$ all $x \succeq 0 \Leftrightarrow y \succeq 0$
self dual

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 $\begin{cases} Y \notin K \to Y \notin K^* \\ Y \in K \to Y \in K \end{cases}$
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 $X \in S_+^n$
$\Rightarrow \text{tr}(XY) = \text{tr}(X[\sum \lambda_i v_i v_i^T]) = \text{tr}(\sum \lambda_i X v_i v_i^T) = \sum \lambda_i \text{tr}(X v_i v_i^T)$

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

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
exer 2.31

Show that maximum of a convex function f over the polyhedron P = conv$\{v_1 \dots v_k\}$ is achieved at one of its vertices
$\sup_{x \in P} f(x) = \max f$

Ch3 Convex functions

(Definition)

$f: R^n \to R$

if dom f is convex set

$x, y \in$ 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.d.f.

dom f convex

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

(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)

$\to$ Exponential e^{ax} on R convex

$\to$ Power x^a on R_{++} $a \ge 1$, $a \le 0$ convex

x^a -- $0 \le a \le 1$ concave

$\to$ Power of absolute value $|x|^p$, $p \ge 1$

$\to$ logarithm $\log x$ concave

$\to$ negative entropy $x \log x$ convex

on R^n

$\to$ Norms convex

$\to$ Max function convex

$f(x) = \max\{x_1, \dots, x_n\}$

$\to$ Quadratic-over-linear function

$f(x) = x_1^2/x_2$, $x \in R^n, x_2 > 0$ convex

$\to$ log-sum-exp

$f(x) = \log(e^{x_1} + \dots + e^{x_n})$ convex

$\to$ geometric mean

$f(x) = (\prod_{i=1}^n x_i)^{1/n}$ concave

$\to$ log-determinant

$f(x) = \log \det(X)$ on dom $f = S^n_{++}$

concave

Method in Sum:

1. Check basic inequality
2. 2^{nd}-order: Hessian Matrix
3. restrict to an abritary line & verify convexity on R

e.g. $g(t) = \log\det(Z + tV)$

4. operations

(sublevel sets)

$\to \alpha$-sublevel set: $f: R^n \to R$

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

$\to \alpha$-superlevel set:

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

(Epigraph)

epi f

$\to$ graph of $f: R^n \to R$

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

$\to$ epigraph of $f: R^n \to R$ convex

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

$\to$ a function is convex i.f.f. epigraph is convex set.

$\to$ hypograph of $f: R^n \to R$ concave

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

(Jensen's inequality & extensions)

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

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

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

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

$\Rightarrow$ 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)

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

(Composition w/ affine mapping)

$\to f: R^n \to R$

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

$g: R^m \to R$

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

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

$\to$ if f convex, g is convex

e.g. $f(x) = -\sum \log(b_i - a_i^T x)$

$\Rightarrow \phi(z) = z - \log z_i$

(Pointwise maximum)

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

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)

$\to$ 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)

$h: R^k \to R$

$g: R^n \to R^k$

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

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

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)

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

$= h(g_1(x), \dots, g_k(x))$

$h: R^k \to R$

$g_i: R^n \to R$

dom $g = R$

dom $h = R^k$

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)

$\to$ if $f(x,y)$ convex

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

(Perspective of a function)

$f: R^n \to R$.

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

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

$\to f$ convex, g convex

$\to (x,t,s) \in$ epi $g \Leftrightarrow t f(x/t) \le s$

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

$\Leftrightarrow (x/t, s/t) \in$ epi f

$\therefore$ epi $g \Leftrightarrow$ epi f is perspective mapping (operation preserve convex set)

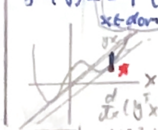
(Conjugate function)

$f: R^n \to R$

$f^*: R^n \to R$

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

what kind of x makes $f(y)$ sup.



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

$y - f'(x) = 0$

$f'(x) = y$

occurs $@$ x

$f^*(y) =$ max difference

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

$= f^*(y)$

f^* convex

pointwise supremum of affine of y

(Quasiconvex definition)

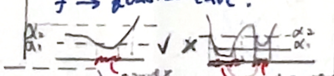
$f: R^n \to R$

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

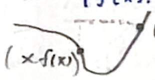
$f \to$ quasiconvex

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

$f \to$ quasiconcave.



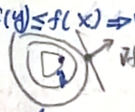
$-f(\theta x + (1-\theta)y) \le \max\{f(x), f(y)\}$



(1st-order condition for quasiconvex func.)

$f: R^n \to R$ dom f convex $x, y \in$ 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)

$\to$ nonnegative weighted maximum

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

$w_i \ge 0$,

f_i: quasiconvex

$\to$ pointwise supremum

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

$\to$ composition

$g: R^n \to R$ quasiconvex

$h: R \to R$ nondecreasing

$f = h \circ g$ is quasiconvex

$\to$ minimization

$f(x,y)$ is quasiconvex jointly

C convex

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

(log-concave -convex)

$f: R^n \to R$

$f(x) > 0$

$x \in$ dom f

$\log f$ convex

$\to f$ is log-convex

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

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

(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, $x \in$ dom f

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

$\to$ multiplication, addition, integration } log-convex preserve

(Convexity w.r.t. GI)

$K \subseteq R^m$ proper cone induces $\preceq_K$

$f: R^n \to R^m$ 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: $f^*(y) = \sup(y^T x - f(x))$

a. $f(x) = -\log x$ dom $f = R_{++}$

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

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

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

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

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

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

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

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

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

has 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_{++}

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

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

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

$X = I + tvv^T$,

$tr(YX) + \log\det X$

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

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

$t \to \infty$

$\to$ if $Y \prec 0$,

$\nabla_X (tr(YX) + \log\det X)$

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

$X = -Y^{-1}$

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

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

$g(t) = tr((Z + tV)^{-1})$ let $Z^{-\frac{1}{2}} V Z^{-\frac{1}{2}}$

$= tr[Z(I + tZ^{-1/2} V Z^{-1/2})]^{-1}$ eigenvalue decomp.

$= tr(Z^{-1}(I + tQ\Lambda Q^T)^{-1})$ $Z^{-\frac{1}{2}} V Z^{-\frac{1}{2}}$

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

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

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

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

③ matrix convexity $f(X) = X^{-1}$

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

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

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

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

epi g convex

$g(x)$ convex $\to f(X)$ 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)

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

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

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

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

$\to$ optimal value infeasible ∞ / unbounded $-\infty$

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

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

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

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

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

(Local & global optima)

locally optima = global optima

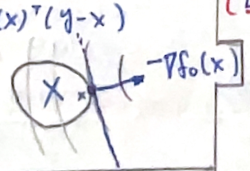
$\to x$ is locally optimal be x feasible
$f_0(x) = \inf\{f_0(z) \mid z \text{ feasible},\ \|z-x\|_2 \le R\}$
$R>0$

$\to$ proof:
if x not globally optimal,
$\exists y$ $f_0(y) < f_0(x)$, $\|y-x\|_2 > R$
also
$z = (1-\theta)x + \theta y$ $\theta = \frac{R}{2\|y-x\|_2}$ $\to \|z-x\|_2 = R/2 < R$
$\therefore f_0(z) \le (1-\theta)f_0(x) + \theta f_0(y) < \theta f_0(x)$
$\Updownarrow$ 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 $f_i(x) \le 0,\ i=1,\dots,m$
$h_i(x) = 0,\ i=1,\dots,p$

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

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

(equivalent convex problems)

(Eliminating equality constraints)

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

(introducing slack variables for linear inequalities)

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

(epigraph problem form)

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

(minimizing over some variables)

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

(Quasiconvex function)

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

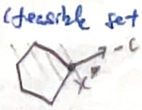
$f_0(x) \le t \Leftrightarrow \phi_t(x) \le 0$
t-sublevel set

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

$\to$ suppose f_0 is differentiable
Let $\bar{X}$ be the feasible set.
if $x \in \bar{X}$ & $\nabla f_0(x)^T(y-x) > 0$
$\forall$ all $y \in \bar{X} \setminus \{x\}$:
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 $Gx \preceq h$
$Ax=b$

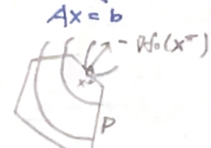
$f_0(x) = \frac{c^Tx+d}{e^Tx+f}$ $\operatorname{dom} f_0 = \{x \mid e^Tx+f > 0\}$

is equivalent
$\to$ 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 $\frac{1}{2}x^TP_ix + q_i^Tx + r_i \le 0$
$i=1,\dots,m$
$Ax=b$ $P_i \in S_+^n$ $i=0\sim m$

over a feasible region that's intersection of ellipsoids

(Second-order cone programming)

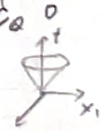
minimize f^Tx
subject to $i=1\dots m$
$\|A_ix + b_i\|_2 \le c_i^Tx + d_i$
$Fx = g$

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

$\Leftrightarrow (Ax+b, c^Tx+d) \succeq_Q 0$

SOC: $k=2$

$\begin{bmatrix} x_1 \\ x_2 \\ t \end{bmatrix} \Leftrightarrow \|x\|_2 \le t$



(Geometric Programming)

$f: R^n \to R$ $\operatorname{dom} f = R_{++}^n$

$f(x) = c x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}$

$c > 0$, $a_i \in R$

$\to$ monomial function

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

$c_k > 0$
$\to$ posynomial function

closed under (posynomial) { addition, multiplication, nonnegative scaling }

(monomial) { multiplication, division }

posynomial $\times$ monomial $\to$ posynomial
posynomial / monomial $\to$ 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.

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

transform to convex

let $y_i = \log x_i$ $x_i = e^{y_i}$

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

$f(x) = \sum_{k=1}^{K} c_k x_1^{a_{1k}} x_2^{a_{2k}} \cdots x_n^{a_{nk}}$

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

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

(Generalized inequality constraints)

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

$R^n \to R^{k_i}$
K_i-convex

(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)
$Ax = b$

$G, F_1, \dots, F_n \in S^k$ (if they are all diagonal, SDP $\to$ LP)
$A \in R^{p \times n}$

LP$\to$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$\to$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 $\begin{bmatrix} I & \frac{1}{\sqrt{2}}(P^{\frac{1}{2}})^Tx \\ \frac{1}{\sqrt{2}}x^TP^{\frac{1}{2}} & tI \end{bmatrix} \succeq 0$
$\operatorname{diag}(Gx-h) \preceq 0$
$Ax = b$

QCQP$\to$SDP

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

SOCP$\to$SDP

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

(Vector optimization)

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

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

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

$g(\lambda,\nu)=\inf_{x\in D} L(x,\lambda,\nu)\le L(\tilde{x},\lambda,\nu)\le f_0(\tilde{x})$ (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

$g(\lambda,\nu)\le p^*=f_0(\tilde{x})\ \forall\lambda\succeq 0$

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)

$=\inf_x\left(f_0(x)+\sum_{i=1}^m\lambda_i^* f_i(x)+\sum_{i=1}^p\nu_i^* h_i(x)\right)$

$\le f_0(x^*)+\sum_{i=1}^m\lambda_i^* f_i(x^*)+\sum_{i=1}^p\nu_i^* h_i(x^*)$

$\le f_0(x^*)$ (≥ 0, ≤ 0, $=0$)

① x^* is one of the minimizer of $L(x,\lambda^*,\nu^*)$

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!

Duality w/ Generalized Inequalities

minimize $f_0(x)$
subject to $f_i(x)\preceq_{K_i} 0$ (proper cone)
$h_i(x)=0$

$K_i\subseteq R^{k_i}$ → 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 $\sum\lambda_i^{*T}f_i(x^*)=0\Rightarrow f_i(x^*)=0$

③ $\lambda_i^*\succ_{K_i^*}0\Rightarrow f_i(x^*)=0$

$f_i(x^*)\prec_{K_i}0\Rightarrow\lambda_i^*=0$

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 $f_i(x)\le u_i$
$h_i(x)=v_i$

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

$u\in R^m$, $v\in R^p$

△ 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): $\nabla p^*(u,v)$ @ $(u,v)=(0,0)$

$\lambda_i^*=-\frac{\partial p^*(0,0)}{\partial u_i}$ ($\partial p^*(0,0)=-\lambda_i^*\partial u_i$)

$\nu_i^*=-\frac{\partial p^*(0,0)}{\partial v_i}$ ($\partial p^*(0,0)=-\nu_i^*\partial v_i$)

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

△ $p^*(u,v)\ge p^*(0,0)-\sum_i\lambda_i^{*T}u_i-\nu^{*T}v$

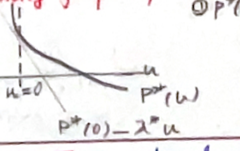
global sensitivity

$\lambda_i^*=-\nabla_{u_i}p^*(0,0)$ (local sensitivity)

vector here

sensitivity graphic presentation

① $p^*(u)\ge p^*(0)-\lambda^* u$



Convergence 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
  - $l\ge k$, $f(x^{(l)})-p^*\le\frac{2m^3}{L^2}\left(\frac{1}{2}\right)^{2^{l-k+1}}$

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

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

FACT I

$f(y)\ge f(x)+\nabla f(x)^T(y-x)+\frac{m}{2}\|y-x\|_2^2$

proof

△ Taylor:

$f(y)=f(x)+\nabla f(x)^T(y-x)+\frac{1}{2}(y-x)^T\nabla^2 f(\theta x+(1-\theta)y)(y-x)$ ($\succeq mI$)

$\ge f(x)+\nabla f(x)^T(y-x)+\frac{m}{2}\|y-x\|_2^2$

FACT II

$f(x)\le p^*+\frac{1}{2m}\|\nabla f(x)\|_2^2$ (smaller $\nabla f(x)$, closer to optimality)

proof

$p^*=f(x^*)\ge f(x)+\nabla f(x)^T(x^*-x)+\frac{m}{2}\|x^*-x\|_2^2$

$=f(x)+\frac{m}{2}\left(\|x^*-x\|_2^2+\frac{2}{m}(x^*-x)^T\nabla f(x)\right)$

$=f(x)+\frac{m}{2}\left(\|x^*-x+\frac{1}{m}\nabla f(x)\|_2^2-\frac{1}{m^2}\|\nabla f(x)\|_2^2\right)$

$\ge f(x)-\frac{1}{2m}\|\nabla f(x)\|_2^2$

FACT III

$\forall x$: $\|x^*-x\|_2\le\frac{2}{m}\|\nabla f(x)\|_2$

proof — from Cauchy

$f(x^*)\ge f(x)+\nabla f(x)^T(x^*-x)+\frac{m}{2}\|x^*-x\|_2^2$

$\ge p^*-\|\nabla f(x)\|_2\|x^*-x\|_2+\frac{m}{2}\|x^*-x\|_2^2$

$\Rightarrow\frac{m}{2}\|x^*-x\|_2^2\le\|\nabla f(x)\|_2\|x^*-x\|_2$

FACT IV upper bound: (i) $\forall x,y$: $f(y)\le f(x)+\nabla f(x)^T(y-x)+\frac{M}{2}\|x-y\|_2^2$

(ii) $\forall x$: $p^*\le f(x)-\frac{1}{2M}\|\nabla f(x)\|_2^2$

proof (i) $f(y)=f(x)+\nabla f(x)^T(y-x)+\frac{1}{2}(y-x)^T\nabla^2 f(\theta x+(1-\theta)y)(y-x)$

$\le f(x)+\nabla f(x)^T(y-x)+\frac{M}{2}\|y-x\|_2^2$ ($\preceq MI$)

(ii) $y=x-\frac{1}{M}\nabla f(x)$

$f(x-\frac{1}{M}\nabla f(x))\le f(x)+\nabla f(x)^T(-\frac{1}{M}\nabla f(x))+\frac{M}{2}\cdot\frac{1}{M^2}\|\nabla f(x)\|_2^2$

$=f(x)-\frac{1}{2M}\|\nabla f(x)\|_2^2$ (similar to FACT II)

Convergence Analysis

- Convergence of Gradient Descent w/ exact line search

△ assume $mI\preceq\nabla^2 f(x)\preceq MI$, $0<m\le M$

△ $f(x^{(k)})-p^*\le\left(1-\frac{m}{M}\right)^k\left(f(x^{(0)})-p^*\right)$

△ linear convergence rate

proof

$f(x^{(k)}-t\nabla f(x^{(k)}))\le f(x^{(k)})+\nabla f(x^{(k)})^T(-t\nabla f(x^{(k)}))+\frac{M}{2}t^2\|\nabla f(x^{(k)})\|_2^2$

let $t=\frac{1}{M}$, we have $=f(x^{(k)})+\left(-t+\frac{Mt^2}{2}\right)\|\nabla f(x^{(k)})\|_2^2$

$f(x^{(k)}-\frac{1}{M}\nabla f(x^{(k)}))\le f(x^{(k)})-\frac{1}{2M}\|\nabla f(x^{(k)})\|_2^2$

△ 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)

$\hookrightarrow\|\nabla f(x^{(k)})\|_2^2\ge 2m\left(f(x^{(k)})-p^*\right)$ — ②

from ① & ② $f(x^{(k+1)})\le f(x^{(k)})-\frac{m}{M}\left(f(x^{(k)})-p^*\right)$

$\rightarrow f(x^{(k+1)})-p^*\le\left(1-\frac{m}{M}\right)f(x^{(k)})+\left(1-\frac{m}{M}\right)(-p^*)$

$\rightarrow f(x^{(k+1)})-p^*\le\left(1-\frac{m}{M}\right)\left(f(x^{(k)})-p^*\right)$

- Convergence of GD w/ back-tracking line search

$f(x^{(k)})-p^*\le C^k\left(f(x^{(0)})-p^*\right)$, $C=1-\min\{2m\alpha,\frac{2\beta\alpha m}{M}\}$

remark

△ from $mI\preceq\nabla^2 f(x)\preceq MI$ △ $1\le K(\nabla^2 f(x))=\frac{\lambda_{max}(\nabla^2 f(x))}{\lambda_{min}(\nabla^2 f(x))}\le\frac{M}{m}$

$\hookrightarrow\lambda_{min}(\nabla^2 f(x))\ge m$
$\hookrightarrow\lambda_{max}(\nabla^2 f(x))\le M$

$\therefore$ 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)

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

$\Rightarrow$ 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)

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

Newton decrement

$\lambda(x)=\left[\nabla f(x)^T\nabla^2 f(x)^{-1}\nabla f(x)\right]^{0.5}$ → from $(\Delta x_{nt}^T\nabla^2 f(x)\Delta x_{nt})$

Justification of stopping criterion.

- $f(x)-p^*=f(x)-\inf_y f(y)$, $f(y)\approx\hat{f}(y)$
- $\inf_y\hat{f}(y)=\inf_y\left(f(x)+\nabla f(x)^T(y-x)+\frac{1}{2}(y-x)^T\nabla^2 f(x)(y-x)\right)$

$\hat{f}(y)$ inf @ $y=-\nabla^2 f(x)^{-1}\nabla f(x)+x$

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

- plug back in

- LP duality

maximize $c^T x$
s.t. $Ax\le b$, $x\ge 0$

⇕

minimize $b^T y$
s.t. $A^T y\ge c$, $y\ge 0$

Analysis for GD BLS

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

$f(x^{(k)})-p^*\le C^k\left(f(x^{(0)})-p^*\right)$

$C=1-\min\{2m\alpha,2\beta\alpha m/M\}<1$

Sublinear convergence rate for GD w/o strong convexity

- $\|\nabla f(x)-\nabla f(y)\|_2\le L\|x-y\|_2$
- $x^{(k+1)}=x^{(k)}-t\nabla f(x^{(k)})$

$\rightarrow\forall k$: $f(x^{(k)})-p^*\le\frac{\|x^{(0)}-x^*\|_2^2}{2tk}=O\left(\frac{1}{k}\right)$

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!

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

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

KKT matrix.

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

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

- 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

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

from ③ $v=\Delta x_{nt}$

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

Newton's method

1. (given) $x^{(0)}\in$ 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 $f_i(x)\le 0$ $i=1\dots m$
$Ax=b$ $A\in R^{p\times n}$ rank$(A)=p$

- assume strictly feasible
$\exists x\in D$ such that $\begin{cases} f_i(x)<0 & i=1\dots m \\ Ax=b \end{cases}$
- ∴ 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 $x^*(t)$ ← central point
- central path: $\{x^*(t)\mid t>0\}$
- Question: How good is $x^*(t)$?

Ⓐ minimize $tf_0(x)+\phi(x)$ solution
subject to $Ax=b$
- $x^*(t)$ → 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

- $x^Tx\le yz$ $y\ge0$ $z\ge0$
- $4x^Tx\le 4yz$ $y\ge0$ $z\ge0$
- $4x^Tx+y^2+z^2\le y^2+z^2+4yz$ $y\ge0$ $z\ge0$
- $4x^Tx+y^2+z^2-2yz\le y^2+z^2+2yz$ $y\ge0$ $z\ge0$
- $4x^Tx+(y-z)^2\le(y+z)^2$ $y\ge0$, $z\ge0$
- $\left\|\begin{bmatrix} 2x \\ y-z \end{bmatrix}\right\|_2\le y+z$ $y\ge0$, $z\ge0$

Water-filling

- minimize $-\sum_{i=1}^n\log(\alpha_i+x_i)$
s.t. $x\succeq0$ $1^Tx=1$
- KKT:
$x^*\succeq0$
$1^Tx^*=1$
$\lambda^*\succeq0$
$\lambda_i^*x_i^*=0$
$-1/(\alpha_i+x_i^*)-\lambda_i^*+\nu^*=0$
→ $x^*\succeq0$ $1^Tx^*=1$ $\nu^*\ge1/(\alpha_i+x_i^*)$
$x_i^*(\nu^*-1/(\alpha_i+x_i^*))=0$

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 $\nu^*\ge1/\alpha_i$: $x_i^*>0$ impossible as $\nu^*\ge1/\alpha_i>1/(\alpha_i+x_i^*)$ ↳ violate slackness
∴ $x_i^*=0$ if $\nu^*\ge1/\alpha_i$

∴ $x_i^*=\begin{cases} 1/\nu^*-\alpha_i & \nu^*<1/\alpha_i \\ 0 & \nu^*\ge1/\alpha_i \end{cases}$

Matrix game

- prob$(k=i)=u_i$
- prob$(l=i)=v_i$
- $\sum_{k=1}^n\sum_{l=1}^m u_k v_l P_{kl}$
$=u^TPv$
- player 2
$\sup\{u^TPv\mid v\succeq0$, $1^Tv=1\}$
$=\max_{i=1..n}(P^Tu)_i$
- player I
minimize $\max_{i=1..m}(P^Tu)_i$; s.t. $u\succeq0$, $1^Tu=1$
⇓
minimize t ①
s.t. $u\succeq0$ $1^Tu=1$
$P^Tu\preceq t1$
$L()=t+\lambda^T(P^Tu-t1)-\mu^Tu+\nu(1-1^Tu)$
$=\nu+(1-1^T\lambda)t+(P\lambda-\nu 1-\mu)^Tu$

vice versa.

player I
$\inf\{u^TPv\mid u\succeq0$, $1^Tu=1\}$
$=\min_{i=1..n}(Pv)_i$

player 2
maximize $\min_{i=1..n}(Pv)_i$
s.t. $v\succeq0$, $1^Tv=1$
⇓
maximize ν
s.t. $\lambda\succeq0$ $1^T\lambda=1$ $\mu\succeq0$
$P\lambda-\nu1=\mu$
⇓
maximize ν
s.t. $\lambda\succeq0$
$1^T\lambda=1$
$P\lambda\succeq\nu1$

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

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 $x_1v_1+x_2v_2+\cdots+x_nv_n=b$

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

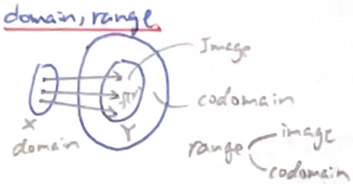
$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0$$
$$\Updownarrow$$
$$C-B^TA^{-1}B \succeq 0$$

Gram matrix

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

for R^3

domain, range



range — image / codomain

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

$(AB)^{-1}=B^{-1}A^{-1}$

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

$(A^T)^{-1}=(A^{-1})^T$

$(A+B)^T=A^T+B^T$

$(AB)^T=B^TA^T$

$(ABC\ldots)^T=\ldots C^TB^TA^T$

$tr(A)=\sum_i A_{ii}$

$tr(A)=\sum_i \lambda_i;\ \lambda_i=eig(A)$

$tr(A)=tr(A^T)$

$tr(AB)=tr(BA)$

$tr(A+B)=tr(A)+tr(B)$

$tr(ABC)=tr(BCA)=tr(CAB)$

$a^TA=tr(aa^T)$

$det(A)=\prod_i \lambda_i,\ \lambda_i=eig(A)$

$det(cA)=c\,det(A)$

$det(A^T)=det(A)$

$det(AB)=det(A)det(B)$

$det(A^{-1})=1/det(A)$

$det(A^n)=det(A)^n$

$det(I+uv^T)=1+u^Tv$

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

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

PD matrix M

$eig(M)_i > 0$, symmetric

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

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

$\rightarrow$ int C

△ limit point of C

if $\forall\, \epsilon>0$, excluding

$(B(x,\epsilon)\setminus\{x\})\cap C \neq \emptyset$

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. $bd(C)\subset C$

△ $f_0(x)=\frac{1}{2}x^TPx+q^Tx+r$

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

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

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

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

linear-fractional programming

quasiconvex function

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

subject to $Gx\preceq h$

$Ax=b$

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

minimize$_{y,z}$ c^Ty+dz $\quad y=\frac{x}{e^Tx+f} \quad z=\frac{1}{e^Tx+f}$

subject to $Gy-hz\preceq 0$

$Ay-bz=0$

$e^Ty+fz=1$

$z\ge 0$

convert QP to SOCP

minimize x^TAx+a^Tx

subject to $Bx\le b$

$\Rightarrow$ minimize$_{x,y}$ $y+a^Tx$

s.t. $Bx\le b$

$x^TAx\le y$

$\Rightarrow$ $y\ge x^TAx$

$\Rightarrow 0\ge x^TAx-y$

$\Rightarrow 0\ge 4x^TAx-4y$

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

$\Rightarrow (1+y)^2\ge 4x^TAx+(1-y)^2$

$\Rightarrow$ minimize$_{x,y}$ $y+a^Tx$

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

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

search length　　search direction

Δ $f(x^{(k+1)}) < f(x^{(k)})$

Δ General Descent method

1. (given) $x^{(0)} \in \mathrm{dom}\, f$
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 - $\Delta x^{(k)}$ @ $f(x^{(k)})$
      $x^{(k)} \in \mathrm{dom}\, f$
      - $\alpha \in (0, 0.5)$
      - $\beta \in (0, 1)$
   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) $x^{(0)} \in \mathrm{dom}\, f$
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) $x^{(0)} \in \mathrm{dom}\, f$
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) $\lambda^{(k)2}/2 \leq \epsilon$
   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) $x^{(0)} \in \mathrm{dom}\, f$

$Ax = b$

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) $x^{(0)} \in \mathrm{dom} f$

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:

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

$\downarrow$

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

$\downarrow$

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