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 columns of A

Schrug complement

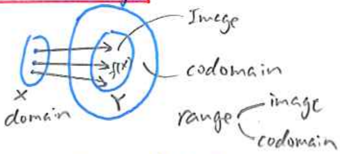
$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0$$
$$\Updownarrow$$
$$C - B^T A^{-1} B \succeq 0$$

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

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

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

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

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

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

$(ABC\ldots)^T = \ldots C^T B^T A^T$

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

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

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

$tr(AB) = tr(BA)$

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

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

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

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

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

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

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

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

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

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

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

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

PD matrix M

$eig(M)_i > 0$, symmetric

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

interior point : $\exists\, \varepsilon > 0$

x s.t. $B(x, \varepsilon) \subset C$

$\rightarrow$ int C

Δ limit point of C

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

$(B(x, \varepsilon) \mid \{x\}) \cap C \neq \emptyset$

or

x is limit point of set S

if $\forall\, \epsilon > 0$, $\exists\, y \in S \setminus \{x\}$

w/ $d(x, y) < \epsilon$

Δ 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

$C = cl(C)$

Δ 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. $bd(C) \subset C$

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

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

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

Δ $det[\lambda I - A]$ polynomial $\rightarrow$ find eigenvalues

$det[\lambda B - A]$ .. $\rightarrow$ find generalized eigenvalues

linear-fractional programming

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

subject to $Gx \preceq h$

$Ax = b$

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

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

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. $\left\| \begin{bmatrix} 2A^{\frac{1}{2}} x \\ 1 - y \end{bmatrix} \right\|_2 \le 1 + y$

$Bx \le b$

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

$x_1, \dots, x_k \in C$

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

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

$\theta_i \in R$

e.g. 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

e.g. 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

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

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

Affine dimension & relative interior

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

e.g.

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

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

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

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

Convex sets

$C \to$ convex

$x_1, \cdots x_k \in C$

$\theta_1 + \cdots + \theta_k = 1$

$\theta_i \ge 0 \in R$

$\Rightarrow \theta_1 x_1 + \cdots + \theta_k x_k \in C$

$C \subseteq R^n$

convex hull:

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

Convex cone

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

$\theta_1 x_1 + \cdots + \theta_k x_k$

$C \subseteq R^n$

conic hull:

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

hyperplane & halfspaces

$\nRightarrow$ affine

$=$ convex

$a^T x \le b$

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

e.g. 1

polyhedron is intersection of halfspaces & half-planes.

e.g. 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 ($z_1^T X z_1 \ge 0$, $\vdots$, $z_\infty^T X z_\infty \ge 0$)

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)

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

$B\ A_i \in S^m$

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

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

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

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

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

$P = \{f(u) \mid f(u) = P^{1/2} u + x_c,\ \|u\|_2 \le 1\}$

polyhedra

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

- intersection of "FINITE" number of halfspaces & halfplanes

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

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

e.g. simplex

$C = \operatorname{conv}\{v_0 \cdots v_k\}$

$v_0 \cdots v_k \to$ affinely independent

simplex $\to$ polyhedron representation

simplex:

$C = \operatorname{conv}\{v_0 \cdots v_k\}$

$= \{x \mid \theta_0 v_0 + \theta_1 v_1 + \cdots + \theta_k v_k \mid \theta \succeq 0,\ 1^T \theta = 1\}$

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

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

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

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

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

$\therefore x = v_0 + By$

$AB = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} B = \begin{bmatrix} I \\ 0 \end{bmatrix}$

$\hookrightarrow$ non-singular matrix exists

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

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

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

$A_1 x \succeq A_1 v_0$

$A_2 x = A_2 v_0$

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 $z \in R^n$

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

e.g.

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

$\Downarrow$

$x_1 \ge 0\ x_3 \ge 0$

$x_1 x_3 \ge x_2^2$

Linear-fractional & perspective functions

① perspective function $P: R^{n+1} \to R^n$, $\operatorname{dom} P = R^n \times R_{++}$

$\Rightarrow$

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

- $C \subseteq \operatorname{dom} P$, 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)

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

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

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

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

image, inverse image are all convex

proper cones & generalized inequalities

- Assume $K \subseteq R^n$, 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.

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

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

- 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

- $C \subseteq R^n$

$x_0 \in \operatorname{bd} C$

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

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

$\forall x \in C$

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

$\downarrow$

supporting hyperplane

$\downarrow$

$\forall$ any $\operatorname{int} C \neq \emptyset$, C is convex, $x_0 \in \operatorname{bd} C$

$\exists$ hyperplane to C @ x_0

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 $\lambda^T x \le \mu^T x$, $\forall x \succeq_K 0$

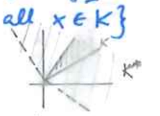
Dual cones

K is a cone

K^* is a dual cone

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

- Dual cone always convex
  Exer 2.31



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

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

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

self-dual

e.g. 3 Positive semidefinite cone

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

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

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

proof: self-dual

if $Y \notin K$ ($Y \notin S_n^+$) $\Rightarrow \exists z \mid z^T Y z < 0$, $\operatorname{tr}(\underbrace{z z^T}_{\in S_n^+} Y) < 0$, $\operatorname{tr}(Y) < 0$

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

$\Rightarrow \operatorname{tr}(XY) = \operatorname{tr}(X[\sum_{i=1}^n \lambda_i v_i v_i^T]) = \operatorname{tr}(\sum_{i=1}^n \lambda_i X v_i v_i^T) = \sum_{i=1}^n \lambda_i \operatorname{tr}(X v_i v_i^T)$, $\lambda_i \ge 0$

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

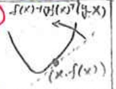
$\sup_{x \in P} f(x) = \max f$

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.f.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. 2nd-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)

$\to$ α-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
$\to$ α-superlevel set:
$C_\alpha = \{x \in \text{dom} f \mid f(x) \ge \alpha\}$ C_α of concave func. is convex

(Epigraph)

$\to$ 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)

$f(\theta x + (1-\theta) y) \le \theta f(x) + (1-\theta) f(y)$
$\Rightarrow f(\theta_1 x_1 + \dots + \theta_k x_k) \le \theta_1 f(x_1) + \dots + \theta_k f(x_k)$, $\Sigma \theta_i = 1$
$\Rightarrow f(\int_S p(x) x \, dx) \le \int_S f(x) p(x) \, dx$
$\Rightarrow f(Ex) \le E f(x)$
$\Rightarrow$
convex inequality:
$\text{prob}(X = x_1) = \theta$, $\text{prob}(X = x_2) = (1-\theta)$
$\therefore f(\theta x_1 + (1-\theta) x_2) \le \theta f(x_1) + (1-\theta) f(x_2)$

(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 $g = \{x \mid Ax+b \in \text{dom} f\}$
$\begin{bmatrix} x_1 \\ x_m \end{bmatrix} = f(\underbrace{Ax+b}_{R^{n \times m} R^m + R^n})$
$\to$ if f convex, g is convex
e.g. $f(x) = -\sum_{i=1}^m \log(b_i - a_i^T x)$
$\Rightarrow \varphi(\delta) = \Sigma - \log \delta_i$

(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_i non-decreasing, g_i convex
f convex: h convex, h_i non-increasing, g_i concave
f concave: h concave, h_i non-decreasing, g_i 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
$\to (x,t,s) \in \text{epi}\, g \Leftrightarrow t f(x/t) \le s$
$\Leftrightarrow f(x/t) \le s/t$
$\Leftrightarrow (x/t, s/t) \in \text{epi} f$
$\therefore$ epi $g \leftrightarrow$ epi f
is perspective mapping (operation preserve convex set)

(Conjugate function)

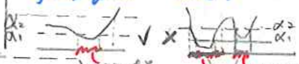
$f: R^n \to R$
$f^*: R^n \to R$
$f^*(y) = \boxed{\sup_{x \in \text{dom} f} (y^T x - f(x))}$
what kind of x makes $f^*(y)$ sup.

$\frac{d}{dx}(y^T x - f(x)) = 0$ (get maximum)
$y - f'(x) = 0$
$f'(x) = y$
occurs @ ★
$f^*(y) = $ 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.

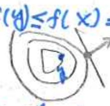


convex / not convex

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

(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
$f(x) = \sup_{y \in C} (\underbrace{w(y)}_{\ge 0} \underbrace{g(x,y)}_{\text{quasiconvex}})$
$\to$ composition
$g: R^n \to R$ quasiconvex
$R \to R$ 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
induce $\preceq_K$
$f: R^n \to R^m$ 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 functions)

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

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 $y < 0$ $\frac{d}{dx}(y^T x + \log x) = 0$
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_{++}
$f^*(Y) = \sup_{X \succ 0} (\underbrace{\text{tr}(YX)}_{\text{inner product of matrix } S^n} + \log\det X)$
$\text{tr}(YX) + \log\det X$ unbounded above unless $Y \prec 0$:
$\to$ if $Y \not\prec 0$, Y has v $\|v\|_2 = 1$, eigenvalue $\lambda \ge$
$X = I + tvv^T$,
$\text{tr}(YX) + \log\det X$
$= \text{tr}\, Y + t\lambda + \log\det(I + tvv^T)$
$= \text{tr}\, Y + t\lambda + \log(1+t)$
$t \to \infty$
$\to$ 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. $f(X) = \text{tr}(X^{-1})$
$g(t) = \text{tr}((Z + tV)^{-1})$
$= \text{tr}[Z(I + tZ^{-1/2} V Z^{-1/2})]^{-1}$
$= \text{tr}(Z^{-1}(I + tQ\Lambda Q^T)^{-1})$
$= \text{tr}(Z^{-1}(QQ^T + tQ\Lambda Q^T)^{-1})$
$= \text{tr}(Z^{-1} Q (I + t\Lambda)^{-1} Q^T)$
$= \text{tr}(Q^T Z^{-1} Q (I + t\Lambda)^{-1})$
$= \sum_{i=1}^n [(Q^T Z^{-1} Q)]_{ii} (1 + t\lambda_i)$
let $Z^{-\frac{1}{2}} V Z^{-\frac{1}{2}}$ eigenvalue decomposition
$Z^{-\frac{1}{2}} V Z^{-\frac{1}{2}} = Q\Lambda Q^T$

③ matrix convexity $f(X) = X^{-1}$
$g(X) = y^T X^{-1} y$ (if g is convex $f(x)$ is convex)
epi $g = \{(X, t) \mid X \succ 0, y^T X^{-1} y \le t, y \in R^n\}$
$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0 \Leftrightarrow C - B^T A^{-1} B \succeq 0$
$t - y^T X^{-1} y \ge 0$ $\therefore \underbrace{\begin{bmatrix} X & y \\ y^T & t \end{bmatrix} \succeq 0}_{\text{LMI}}$
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)

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

$p^*=\infty$ if problem infeasible

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

b.t.w $D=\bigcap_{i=0}^{m}\mathrm{dom} f_i \cap \bigcap_{i=1}^{p}\mathrm{dom} h_i$

(optimal & locally optimal points)

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

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

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

→ optimal value attained / achieved → solvable

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

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

→ feasibility problem

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

(Convex Optimization)

minimize $f_0(x)$ convex

subject to $f_i(x)\le 0,\ i=1,\dots,m$ convex

$a_i^Tx=b_i,\ i=1,\dots,p$ affine

feasible set $D=\bigcap_{i=0}^{n}\mathrm{dom} f_i$

→ minimize a convex objective function over a convex set

(Local & global optima)

locally optima = global optima

→ x is locally optimal if x feasible

$f_0(x)=\inf\{f_0(z)\mid z \text{ feasible},\ \|z-x\|_2\le R\}$

$R>0$

→ proof:

if x not globally optimal,

$\exists y\quad f_0(y)<f_0(x),\ \|y-x\|_2>R$

also

$z=(1-\theta)x+\theta y\quad \theta=\frac{R}{2\|y-x\|_2}\to\|z-x\|_2=R/2<R$

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

↕ contradicts

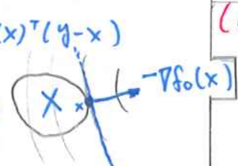
$f_0(x)=\inf\{f_0(z)\mid z \text{ feasible},\ \|z-x\|_2\le R\}$

(optimality criterion)

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

x is optimal iff.

$\nabla f_0(x)^T(y-x)\ge 0$



(more on vector optimization)

Scalarization $\in R^n$

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

minimize $\lambda^Tf_0(x)$

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

$h_i(x)=0,\ i=1,\dots,p$

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

→ 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$_x$ $f_0(x)$

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

$h_i(x)=0$ $i=1,\dots,p$

↕

minimize$_z$ $f_0(Fz+x_0)$

subject to $f_i(Fz+x_0)\le 0$, $i=1..m$

(introducing equality constraints)

minimize$_x$ $f_0(A_0x+b_0)$

subject to $f_i(A_ix+b_i)\le 0$, $i=1..m$

↕

minimize$_{x,y}$ $f_0(y_0)$

subject to $f_i(y_i)\le 0$, $i=1..m$

$y_i=A_ix+b_i$, $i=0,1..m$

(introducing slack variables for linear inequalities)

minimize $f_0(x)$

subject to $a_i^Tx\le b_i$, $i=1,\dots,m$

↕

minimize$_{x,s}$ $f_0(x)$

subject to $a_i^Tx+s_i=b_i$, $i=1..m$

$s_i\ge 0$, $i=1..m$

(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 $f_i(x_1)\le 0$, $i=1..m$

↕

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

$\underbrace{f_0(x)\le t}_{t\text{-sublevel set}}\iff\phi_t(x)\le 0$

formulate as feasibility problem

find x

subject to $\phi_t(x)\le 0$

$f_i(x)\le 0$ $i=1,\dots,m$

$Ax=b$

→ 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 $Gx\preceq h$

$Ax=b$

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

is equivalent

→ minimize c^Ty+dz

subject to $Gy\preceq hz$

$Ay=bz$

$e^Ty+fz=1$

$z\ge 0$

(Quadratic Program)

minimize $\frac12 x^TPx+q^Tx+r$

subject to $Gx\preceq h$

$Ax=b$

$P\in S^n_+$

(Quadratically constrained quadratic program)

minimize $\frac12 x^TP_0x+q_0^Tx+r_0$

subject to $\frac12 x^TP_ix+q_i^Tx+r_i\le 0$ $i=1,\dots,m$

$Ax=b$ $P_i\in S^n_+$ $i=0..m$

over a feasible region that's intersection of ellipsoids

(Second-order cone programming)

minimize f^Tx

subject to $i=1..m$

$\|A_ix+b_i\|_2\le c_i^Tx+d_i$

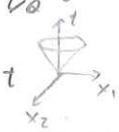
$Fx=g$

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)

$f: R^n\to R$ $\mathrm{dom} f=R^n_{++}$

$f(x)=cx_1^{a_1}x_2^{a_2}\cdots x_n^{a_n}$,

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

→ monomial function

sum of monomials

$f(x)=\sum_{k=1}^{K}c_kx_1^{a_{1k}}x_2^{a_{2k}}\cdots x_n^{a_{nk}}$

$c_k>0$

→ 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.

$\mathrm{dom} f=R^n_{++}$ − $x\succ 0$ (implicit)

transform to convex

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

$\therefore f(x)=f(e^{y_1},\dots,e^{y_n})$

$=c(e^{y_1})^{a_1}\cdots(e^{y_n})^{a_n}$

$=e^{a^Ty+b}$ $b=\log c$

$f(x)=\sum_{k=1}^{K}c_kx_1^{a_{1k}}x_2^{a_{2k}}\cdots x_n^{a_{nk}}$

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

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 $f_i(x)\preceq_{K_i}0$, $i=1\dots m$

$Ax=b$

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

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 $x_1F_1+\dots+x_nF_n+G\preceq 0$ LMI

$Ax=b$

$G,F_1,\dots,F_n\in S^k$ (if they are all diagonal, SDP→LP)

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

LP→SDP

minimize c^Tx+d

subject to $Gx\preceq h$

$Ax=b$

↕

minimize c^Tx+d

subject to $Gx-h\preceq 0$

$Ax=b$

↕

minimize c^Tx+d

subject to $\mathrm{diag}(Gx-h)\preceq 0$

$Ax=b$

QP→SDP

minimize $\frac12 x^TPx+q^Tx+r$

subject to $Gx\preceq h$

$Ax=b$

↕

minimize$_{t,x}$ $t+q^Tx+r$

subject to $\frac12 x^TPx\le t$

$Gx\preceq h$

$Ax=b$

↕

minimize$_{t,x}$ $t+q^Tx+r$

subject to $\begin{bmatrix} & \frac{1}{\sqrt2}(P^{\frac12})^Tx\\ \frac{1}{\sqrt2}x^TP^{\frac12} & tI\end{bmatrix}\succeq 0$

$\mathrm{diag}(Gx-h)\preceq 0$

$Ax=b$

QCQP→SDP

minimize $\frac12 x^TP_0x+q_0^Tx+r_0$

subject to $\frac12 x^TP_ix+q_i^Tx+r_i$ $i=1\sim m$

$Ax=b$

↕

minimize$_{u,t,x}$ $t+q_0^Tx+r_0$

subject to $\frac12 x^TP_0x\le t$

$\frac12 x^TP_ix\le u_i$ $i=1..m$

$u_i+q_i^Tx+r_i\le 0$ $i=1..m$

$Ax=b$

↕

minimize$_{u,t,x}$ $t+q_0^Tx+r_0$

subject to $\begin{bmatrix}I & \frac{1}{\sqrt2}(P^{\frac12})^Tx\\ \frac{1}{\sqrt2}x^TP^{\frac12} & tI\end{bmatrix}\succeq 0$

$\begin{bmatrix}I & \frac{1}{\sqrt2}((P_i)^{\frac12})^Tx\\ \frac{1}{\sqrt2}x^TP_i^{\frac12} & u_iI\end{bmatrix}\succeq 0$

$u_i+q_i^Tx+r_i$ $i=1\sim m$

$Ax=b$

SOCP→SDP

minimize f^Tx

subject to $\|A_ix+b_i\|_2\le c_i^Tx+d_i$

$Fx=g$ $i=1\sim m$

↕

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 $f_i(x)\preceq 0$ $i=1\sim m$

$h_i(x)=0$ $i=1\sim p$

$R^n\to R$ $R^n\to R$

① $f_0(x)$ K-convex

② $f_1\sim f_m$ convex

③ h_i affine

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

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

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

$P\subseteq\mathcal{O}\cap\mathrm{bd}\,\mathcal{O}$ 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 x feasible, (λ,ν) feasible)

The Lagrangian

minimize $f_0(x)$ — λ_i

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

$h_i(x)=0$ $i=1,\dots,p$ — ν_i

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

$\mathrm{dom}\,L = D\times R^m \times R^p$

A **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. $f_i(x)\le 0$)

B **Optimality conditions**

- Complementary slackness

$f_0(x^*) = g(\lambda^*,\nu^*)$ (zero duality gap, 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**

$\to x^*$ 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: – strong duality holds – x^* primal, (λ^*,ν^*) dual must satisfy KKT!

$f_i(x^*)\le 0$

$h_i(x^*)=0$

$\lambda_i^*\ge 0$

$\lambda_i^* f_i(x^*)=0$

$\nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^*\nabla f_i(x^*) + \sum_{i=1}^p \nu_i^*\nabla h_i(x^*) = 0$

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

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

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

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 nonnegative

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

A **dual problem w/ G.I.**

maximize $g(\lambda,\nu)$

subject to $\lambda_i\succeq_{K^*} 0$

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 $f_i(x)\prec_{K_i} 0$, $Ax=b$

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 $\sum_{i=1}^m \lambda_i^{*T} f_i(x^*) = 0$

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

$\to x^*$ 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
- $p^*(u,v)\ge p^*(0,0) - \sum_{i=1}^m \lambda_i^{*T}u_i - \nu^{*T}v$

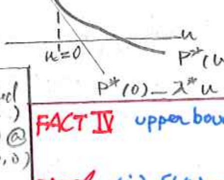
(global sensitivity)

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

↳ vector here

sensitivity graphic presentation

① $p^*(u)\ge p^*(0) - \lambda^* u$



On 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 $\alpha\beta mML$
   - $K\le \frac{f(x^{(0)})-p^*}{\gamma}$
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:

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

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

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) $f(y) = f(x) + \nabla f(x)^T(y-x) + \frac{1}{2}(y-x)^T\underbrace{\nabla^2 f(\theta x + (1-\theta)y)}_{\preceq MI}(y-x)$

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

(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)

Algorithms Ch9, 10, 11

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

we have

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

- 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

$\to f(x^{(k)}) \le p^* + \frac{1}{2m}\|\nabla f(x)\|_2^2$ (from previous)

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

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

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

$\to f(x^{(k+1)}) - p^* \le (1-\frac{m}{M})(f(x^{(k)}) - p^*)$ ※

- 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 $mI\preceq\nabla^2 f(x)\preceq MI$

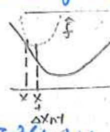
↳ $\lambda_{min}(\nabla^2 f(x))\ge m$

↳ $\lambda_{max}(\nabla^2 f(x))\le M$

- $1\le K(\nabla^2 f(x)) = \frac{\lambda_{max}(\nabla^2 f(x))}{\lambda_{min}(\nabla^2 f(x))} \le$

$\therefore$ if $M >> m$, converge sl

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 optimality 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)
  $\to v = -\nabla^2 f(x)^{-1}f(x) = \Delta x_{nt}$ → from

Newton decrement

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

Justification of stopping criterion.

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

$\hat{f}(y)$ inf @

$y = -\nabla^2 f(x)^{-1}\nabla f(x)$

$\inf_y \hat{f}(y) =$

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

- 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

- $\hat{f}(t) = f(x - t\nabla f(x))$
- $\hat{f}(t) \le f(x) - \alpha t\|\nabla f$
- $0\le t\le 1/M \to -t + \frac{Mt^2}{2}$
- ↳ $\hat{f}(t)\le f(x) - t\|\nabla f(x)\|$
- $\frac{Mt^2}{2}\|\nabla f(x)\|_2^2 \le f(x) - ($
- $\|\nabla f(x)\|_2^2 \le f(x) - \alpha t\|$

since $\alpha<1/2$ $t=$

- BLS terminates $t\ge$
- gives lower bound decrease in objective
- $\to f(x^+)\le f(x) - \alpha\|\nabla f$
  $f(x^+)\le f(x) - (\beta\alpha/$
- $\to f(x^+)\le f(x) - \min\{\alpha,\beta$
  $\|\nabla f(x)\|_2^2$
- $\to f(x^+) - p^* \le f(x$
  $-\min(\alpha,\beta\alpha/M)\|\nabla f$
- combine $\|\nabla f(x)\|_2^2 \ge 2m($
  $\to$ get:

$f(x^k) - p^* \le C^k(f(x^0$

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

Sublinear convergence rate for GD w.o. strong convexity

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

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

Algorithms (cont'd) 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
- @ x,
- addition of v: $x+v$
- $x+v$ should be feasible
  $A(x+v) = b$
  $Av = 0$

④ $f(x+v) = f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x) v$

from ① & ④ we get

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

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

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

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

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

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

Taylor 1-st order linearization

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

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

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

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

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

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

$\Rightarrow -(\nabla^2 f(x) \Delta x_{nt} + A^T w) \Delta x_{nt}$

$= -\Delta x_{nt}^T \nabla^2 f(x) \Delta x_{nt} + A^T w \Delta x_{nt}$

$= -\Delta x_{nt}^T \nabla^2 f(x) \Delta x_{nt} + 0$

$= -\lambda^2(x)$

< 0

justification of stopping criterion

$f(x) - p^*$

$= f(x) - \inf_y \{ f(y) \mid Ay = b \}$

$\Rightarrow f(x) - \inf_y \{ \hat{f}(y) \mid Ay = b \}$

$= f(x) - \inf_v \{ \hat{f}(x+v) \mid A(x+v) = b \}$

$= f(x) - \inf_v \{ \hat{f}(x+v) \mid Av = 0 \}$

$= f(x) - \inf_v \{ f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x) v \mid Av = 0 \}$

from ③ $v = \Delta x_{nt}$

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

$= -(-\lambda^2(x)) - \frac{1}{2} \lambda(x)^2$

$= \frac{1}{2} \lambda(x)^2$

Newton's method

1. (given) $x^{(0)} \in \mathrm{dom} f$
   $Ax = b$
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. $x^{(k+1)} := x^{(k)} + t^{(k)} \Delta x^{(k)}$

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:

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

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

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

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

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

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

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

$\Rightarrow g(\lambda^*(t), \nu^*(t)) = f_0(x^*(t)) + \sum_{i=1}^m \lambda_i^*(t) f_i(x^*(t)) + \nu^*(t)^T \underbrace{(Ax^*(t) - b)}_{0}$

$= f_0(x^*(t)) + \sum_{i=1}^m \frac{1}{-t f_i(x^*(t))} f_i(x^*(t)) + 0 = f_0(x^*(t)) - \frac{m}{t}$

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

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

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

Barrier method

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

⇓

new problem:
minimize $t f_0(x) + \phi(x)$
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 $t f_0 + \phi$ s.t. $Ax = b$ w/ t
   b. $x := x^*(t)$
   c. (until) $m/t < \epsilon$
   d. $t := \mu t$

hyperbolic constraints

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

Water filling

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

if $\nu^* < 1/\alpha_i$
only $x_i^* > 0$
$\nu^* = 1/(\alpha_i + x_i^*)$

solving x_i^*
$x_i^* = 1/\nu^* - \alpha_i$
if $\nu^* < 1/\alpha_i$

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$(l = i) = v_i$
- $\sum_{k=1}^n \sum_{l=1}^m u_k v_l P_{kl}$
  $= u^T P v$

player 2
$\sup\{ u^T P v \mid v \succeq 0, 1^T v = 1 \}$
$= \max_{i=1 \dots m} (P^T u)_i$

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

vice versa.

player 1
$\inf\{ u^T P v \mid u \succeq 0, 1^T u = 1 \}$
$= \min_{i=1 \dots n} (Pv)_i$

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

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

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

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

⇓

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

⇓

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