MATH 3904 Ch.1 Basic properties of solutions and algorithms

Geometric method: Find distance from pt. to Ω

e.g. $\min\ (x_1-3)^2 + (x_2-2)^2$ s.t. $x_1^2 - x_2 - 3 \leq 0$

$x_2 - 1 \leq 0$

$x_1 \geq 0$

Algebraic method: e.g. $\min\ x^T A x$, A symmetric s.t. $x^T x = 1$

$A = Q^T(\lambda_1, \lambda_2, \dots, \lambda_n) Q$

$x^T A x = y^T(\lambda_1 \lambda_2 \dots \lambda_n) y = \sum_{i=1}^{n} \lambda_i y^2$

$\geq \lambda_n$ (1)

Aim: Find a lower bound for function

Guess/Solve parameters for lower bound

holds eq. for $y = (0, 0, \dots 1)^T$

$\rightarrow$ optimal. $x = Q^T y$

Local min: Def: There exists $\delta > 0$ s.t. $f(x^*) \leq f(x)$

- Suppose $f(x)$ has unique LExP x^*. Then x^* is Gmin/Gmax.

- If $f^{(k)}(x^*) = 0$ for $k = 1, 2 \dots n$, $f^{(n+1)}(x^*) \neq 0$,

$\Rightarrow x^*$ is LEx iff $n+1$ is even

if $f^{(n+1)}(x^*) > 0$, Lmin

$f^{(n+1)}(x^*) < 0$, Lmax.

Gradient: $\nabla f = \begin{pmatrix} \partial f / \partial x_1 \\ \partial f / \partial x_2 \\ \partial f / \partial x_3 \end{pmatrix}$

$\nabla^2 f = \begin{pmatrix} \partial x_1 & \partial x_1 & \partial x_1 \\ \partial x_2 & \partial x_2 & \partial x_2 \\ \partial x_3 & \partial x_3 & \partial x_3 \end{pmatrix}$

Re: chain rule! $\frac{\partial f}{\partial \alpha} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial \alpha} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial \alpha} \dots$

$x(\alpha) = x^* + \alpha d$, $g(\alpha) = f(x(\alpha))$. $\Rightarrow g'(\alpha) = \nabla f(x(\alpha))^T d$

$g''(\alpha) = d^T \nabla^2 f(x(\alpha)) d$

$f(x^* + d) = f(x^*) + \nabla(x^*)^T d + \frac{1}{2} d^T \nabla^2 f(x^* + \theta d) d$

~ Taylor's Thm

Feasible direction: $f(\bar{x} + \alpha d)$ $\bar{x} + \alpha d \in \Omega$

1^{st} order NC: $\nabla f(x^*)^T d \geq 0$ for any feasible dir. $d \in R^n$ of x^*.

↳ $\lim_{\alpha \to 0^+} \frac{f(x^*+d) + f(x^*)}{\alpha} = \nabla f(x^*)$

$\boxed{\nabla f(x^*) = 0}$ (interior pt, unconstrained case)

↳ n eqt, n variables → solve!

2^{nd} order NC: $d^T \nabla^2 f(x^*) d \geq 0$ whenever $\nabla f(x^*)^T = 0$.

i.e. $\nabla^2 f(x^*) d$ is PSD

2^{nd} order SC: $\nabla^2 f(x^*)$ is PD

sub!

" Use NC: Find candidate for SC.
SC: ensure that a certain pt. is LMin"

$\boxed{\begin{array}{l} \text{PD / PSD determination} \\ \text{① By definition} \;\Big|\; \begin{array}{l} PD: x^TAx > 0 \\ PSD: x^TAx \geq 0 \end{array} \\ \text{② All eigenvalue +ve: PD/PSD} \\ \text{③ PD: det(principal submatrix)} > 0 \\ \quad\; \text{PSD: det(principal submatrix)} \geq 0 \end{array}}$

Convex! $\iff f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y) \to f(\sum_{i=1}^{k} \alpha_i x_i) \leq \sum_{i=1}^{k} \alpha_i f(x_i)$

(bowl-shaped) $\iff f(y) \geq f(x) + \nabla f(x)^T (y-x)$

→ two points

→ tangent

$\iff \nabla^2 f(x)$ is PSD throughout Ω

$\Leftarrow f''(x) > 0$, for all $x \in \Omega$, f strictly convex

Convex min! set where f achieves min is convex

LMin of f is also GMin.

$\nabla f(x^*)^T (x - x^*) \geq 0 \;\; \forall x \in \Omega$

Convex Max! - At extreme pt (no x_1, x_2 s.t. $x = \alpha x_1 + (1-\alpha) x_2$)

e.g. $f(x)$ $[a,b]$ max of $\max(f(a), f(b))$

- Suppose $f(x)$ is convex over R^n. if x^* is GMaxP, f is const. function

Convex → unimodal

MATH 2014 Memo:

$A = \frac{\partial f^2}{\partial^2 x_1}\Big|_{x^*}$ $B = \frac{\partial f^2}{\partial x_1 \partial x_2}$ $C = \frac{\partial f^2}{\partial^2 x_2}$

$H = \det\begin{pmatrix} A & B \\ B & C \end{pmatrix} = AC - B^2$.

$H > 0$, $A > 0$ min, $A < 0$ max; $H < 0$ saddle; $H = 0$ indeterminated.

MATH 3904 Ch.2 Basic Descent Methods

Order of convergence p : $\lim_{k\to\infty} \frac{\|x_{k+1}-x^*\|}{\|x_k - x^*\|^p} = \beta < \infty$

↳ higher, more efficient

linear convergence: $p=1$, $\beta < 1$.

superlinear converge: $p=1$, $\beta = 0$

solve convergence: $x^{k+1} = x^k = x^*$.

1-dimensional search

(A) Without derivatives

ℓ : interval of uncertainty

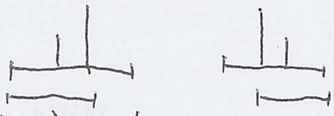
ε : distinguishability number

Assume: Function is unimodal (strictly convex)

① Golden selection method

1. $\boxed{\begin{aligned}\lambda_1 &= a_1 + (1-\alpha)(b_1 - a_1)\\ \mu_1 &= a_1 + \alpha(b_1 - a_1)\end{aligned}}$

$\alpha = 0.618$



a λ μ b

↳ only 1 extra obsv.

2. if $b_k - a_k < \ell$, stop.
3. select new $a/b/\lambda/\mu$

$0.618^k \le \frac{\ell}{b_1 - a_1}$

linear convergence, ratio $= 0.618$

② Fibonacci

$\boxed{\begin{aligned}\lambda_k &= a_k + \frac{F_{n-k-1}}{F_{n-k+1}}(b_k - a_k)\\ \mu_k &= a_k + \frac{F_{n-k}}{F_{n-k+1}}(b_k - a_k)\end{aligned}}$

$F_0 = F_1 = 1$

$n = \frac{|a-b|}{\ell} - 1$

$b_{k+1} - a_{k+1} = \frac{F_{n-k}}{F_{n-k+1}}(b_k - a_k)$

1. select new $a/b/\lambda/\mu$
2. Last observation: $\lambda_n = \lambda_{n-1}$, $\mu_n = \lambda_{n-1} + \varepsilon$

$\frac{1}{F_n} \le (0.618)^{n-1}$

→ Fib more efficient.

(B) ① With derivatives: Newton's method

$|f'(x_{k+1})| < \varepsilon$ or $|x_{k+1} - x_k| < \varepsilon$

termination scalar

$\boxed{x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}}$

not globally convergent!

$g(x^*) = f'(x^*) = 0$, $g'(x^*) \ne 0 \Rightarrow$ converge to x^* with order ≥ 2.

$\boxed{x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k)}$

Multi-dimensional search

② Steepest descent

- Direction of descent: $f(x+\alpha d) < f(x)$ for all $0<\alpha<\delta$ for some $\delta > 0$

$f'(x,d) < 0$

- move along d, where $|d| = 1$, to minimize $f'(x,d)$

↳ optimal sol: $d = -\frac{\nabla f(x)}{\|\nabla f(x)\|}$ (direction of steepest descent)

- line search along $g(\alpha) = f(x_k + \alpha d_k)$, $d_k = -\nabla f(x_k)$ (GSM/FM)

$g'(\alpha) = 0$

$d_k^T \nabla f(x_k + \alpha_k d_k) = 0$

$d_k^T \cdot d_{k+1} = 0$

$\Rightarrow$ search dir orthogonal.

Steepest descent - Quadratic case

Q : PD

Minimize $f(x) = \frac{1}{2}x^TQx - b^Tx$

$d_k = b - Qx_k$, $\nabla f(x) = Qx_k - b$

$b = Qx + d_k$

$g(\alpha) = f(x_k + \alpha d_k)$

$g'(\alpha) = x_k^T Q d_k + \alpha d_k Q d_k - d_k^T b = 0$

$$\boxed{\alpha_k = \frac{d_k^T d_k}{d_k^T Q d_k}}$$

$$x_{k+1} = x_k + \frac{d_k^T d_k}{d_k^T Q d_k} d_k$$

Unique min pt. Gradient $= d_k = 0$, $Qx^* = b$.

Define $E(x) = \frac{1}{2}(x-x^*)^T Q (x - x^*) = f(x) + \underbrace{\frac{1}{2}x^{*T}Qx^*}_{\text{const.}}$

$$E(x_{k+1}) = \left(1 - \frac{(d_k^T d_k)^2}{(d_k^T Q d_k)(d_k^T Q^{-1} d_k)}\right)$$

$$E(x_{k+1}) \le \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^2 E(x_k) \to \text{converge to } x^*$$

Useful theorems in linear algebra:

A symmetric $\Rightarrow$ $A = Q^T D Q$, Q orthogonal ($Q^TQ = I$)

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{\det A}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Q is PD $\Rightarrow$ $Q = PP$, where P is PD (PD square root)

technique: let $y = Ax$, then transform x.
$x = A^{-1}y$

function $f(x + \alpha d)$: direct explode! then differentiate.

Eigenvector: $Ax = \lambda x$, $x\ y$ of symmetric matrix orthogonal.
$(y^T A x = \lambda_1 y^T x)$

MATH 3904. Ch.3 Conjugate direction methods | Keyword: Q-orthogonal, g_k.

Solve $\min \frac{1}{2}x^TQx - bx$ (sol. to $Qx=b$)

From any x_0, $d_0 = -g_0 = b - Qx_0$

$x_{k+1} = x_k + \alpha_k d_k$ → recursively call x_k.

$$\boxed{\alpha_k = -\frac{g_k^T d_k}{d_k^T Q d_k}}$$

Solve 2x2: use eigenvectors as d_0, d_1. no need calculate β.

$d_{k+1} = -g_{k+1} + \beta_k d_k$

$$\beta_k = \frac{g_{k+1}^T Q d_k}{d_k^T Q d_k}$$

$g_k = Qx_k - b = \nabla f(x_k)$

Properties:
1. converge to unique solution x^* after n steps, $x_n = x^*$ ↑ dimension ($\mathbb{R}^n$)
2. all d_k linearly independent →, $d_i^T Q d_j = 0$ → $d_i^T Q d_i > 0$
3. $x^* - x_0 = \alpha_0 d_0 + \alpha_1 d_1 + \dots + \alpha_{n-1} d_{n-1}$

   $x_k = x_0 + \alpha_0 d_0 + \alpha_1 d_1 + \dots + \alpha_{k-1} d_{k-1}$

Proofs related: Try to build back $d_i^T Q d_j = 0$!

e.g. Solve $f(x_1, x_2) = 2x_1^2 + 2x_2^2 - 3x_1 x_2 + x_1 - 3x_2$

① Standard form $\frac{1}{2} x^T Q x - b^T x$

$$Q = \begin{pmatrix} 4^{2\cdot x_1^2} & -3_{x_1 x_2} \\ -3_{x_1 x_2} & 4_{2\cdot x_2^2} \end{pmatrix}, \quad b = \begin{pmatrix} -1^{-x_1} \\ 3^{3x_2} \end{pmatrix}$$

② Eigenvectors. $\begin{pmatrix} 4-\lambda & -3 \\ -3 & 4-\lambda \end{pmatrix}$ $\lambda = 1, 7.$ $d_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $d_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$

$\downarrow$

$(4-\lambda)x_1 + (-3)x_2 = 0$

since Q symmetric, d_0 d_1 Q-orthogonal.

③ Let x_0. Let $x_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ $g_0 = Qx_0 - b = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$

$$\alpha_0 = \frac{-g_0^T d_0}{d_0^T Q d_0} = -\frac{(1,-3)(1,1)^T}{(1,1)(1,1)^T} = 1$$

no need calculate d_0, d_1!

d_0, d_1 is eigenvector.

$x_1 = x_0 + \alpha_0 d_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ $g_1 = Qx_1 - b = \begin{bmatrix} 2 \\ -2 \end{bmatrix}$

$$\alpha_1 = -\frac{g_1^T d_1}{d_1^T Q d_1} = \frac{2}{7}.$$

$x_2 = x_1 + \alpha_1 d_1 = \begin{bmatrix} 5/7 \\ 9/7 \end{bmatrix}$, optimal solution.

MATH3904 Ch. 4-5 Constrained minimization conditions, Lagrange

$$\min f(x), \ \text{s.t.} \ h(x)=0, \ g(x)\le 0$$

active: $g(x)=0$
inactive: $g(x)<0$ → combinations! check feasibility

Prop: all $y\in\mathbb{R}^n$ s.t. $\nabla h(x^*)^T y=0 \Rightarrow f(x^*)^T y=0$

1st order NC: (Kuhn - Tucker conditions)

$$\boxed{\begin{cases}\nabla f(x^*)+\nabla h(x^*)\lambda+\nabla g(x^*)\mu=0 \\ \lambda\in\mathbb{R}^m, \ \mu\ge 0 \qquad g(x^*)\mu=0\end{cases}}$$

multiple h → still h_1, h_2, multiple λ.
multiple g → $g_1, g_2 \cdots$

2nd order NC/SC: $L(x^*)=\nabla^2 f(x^*)+\sum_{i=1}^{m}\lambda_i\nabla^2 h_i(x^*)+\sum_{j=1}^{1}\mu_j\nabla^2 g_j(x^*)$ is PSD/PD

0. is it originally PD/PSD?
1. plane $y_1+y_2+y_3=0$ ← sub
2. $(y_1, y_2, y_3)^T L(x^*)\begin{pmatrix}y_1\\y_2\\y_3\end{pmatrix}$ ← on tangent plane of constraints $M=\{y: \nabla h(x^*)^T y=0, \nabla g_j(x^*)^T y=0 \ \forall j\in J\}$
3. check >0 / <0. → Convex → G.min

Skills: For λ as a vector, treat matrix as blocks!

e.g. $\min \frac{1}{2}x^T Qx+c^T x$ s.t. $Ax-b=0$

$$\begin{cases}Qx+c+A^T\lambda=0 \\ Ax-b=0\end{cases} \Rightarrow \begin{bmatrix}Q & A^T\\ A & 0\end{bmatrix}\begin{bmatrix}x\\ \lambda\end{bmatrix}=\begin{bmatrix}-c\\ b\end{bmatrix}$$

$$\begin{bmatrix}x\\ \lambda\end{bmatrix}=\begin{bmatrix}Q & A^T\\ A & 0\end{bmatrix}^{-1}\begin{bmatrix}-c\\ b\end{bmatrix}$$

Try to prove regular pt. ($\nabla h_i(x), \nabla g_i(x)$ linearly independent)

Q PD, $d_0 \cdots d_k$ Q orthogonal $\Rightarrow$ d_i linearly independent.

Q-type: is x^* (a,b,c) an optimal solution?

e.g. T8 Q2. $\min\ f(x) = e^{x_1} + x_1 x_2 + x_2^2 - 2x_2x_3 + x_3^2$

s.t. $x_1^2 + x_2^2 + x_3^2 \le 5$, $a^T x = -2$.

$$\nabla f(x) = \begin{bmatrix} e^{x_1} + x_2 \\ x_1 + 2x_2 - 2x_3 \\ -2x_2 + 2x_3 \end{bmatrix}, \quad \nabla h(x) = a, \quad \nabla g(x) = \begin{bmatrix} 2x_1 \\ 2x_2 \\ 2x_3 \end{bmatrix}$$

$$\nabla^2 f(x) = \begin{bmatrix} e^{x_1} & 1 & 0 \\ 1 & 2 & -2 \\ 0 & -2 & 2 \end{bmatrix}, \quad \nabla^2 h(x) = 0, \quad \nabla^2 g(x) = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

① Kuhn Tucker conditions: there exist $\mu \ge 0$, λ s.t.

$$\begin{cases} \nabla f(x^*) + \lambda \nabla h(x^*) + \mu g(x^*) = 0 \\ \mu g(x^*) = 0 \end{cases}, \quad h(x) = 0.$$

sub $x^* = (0,0,1)$, $\mu = 0$ $\lambda = 1$, $a = (-1,2,2)^T$,

② Put $L(x) = \nabla^2 f(x^*) + \lambda \nabla^2 h(x^*) + \mu \nabla^2 g(x^*)$

$$= \begin{bmatrix} 1 & 1 & 0 \\ 1 & -2 & -2 \\ 0 & -2 & 2 \end{bmatrix},$$ tangent space of active constants:

find point in M, $y^T L y < 0$, not satisfy.

MATH3404 Ch. 6-7 Penalty and Barrier methods

Penalty method: $\min f(x) + c\,p(x)$ $\qquad h(x)=0,\ g(x)\le 0$.

$$p(x) = \frac{1}{2}\sum_{i=1}^{m} h_i(x)^2 + \frac{1}{2}k\sum_{j=1}^{u} \max(\{0, g_j(x)\})^2 + \frac{1}{2}k\sum_{j=1}^{q} (h_j(x))^2$$

e.g. $\min f(x) = x_1 + x_2$, s.t. $g_1(x) = x_1^2 - x_2 \le 0$, $g_2(x) = -x_1 \le 0$

⓪ $\boxed{\phi(x,c) = f(x) + c\,q_1(x)^2 + c\,q_2(x)^2} \Rightarrow$ convex $\rightarrow$ equality: $c(h(x))^2$

① let $q_1(x) = \max\{0, x_1^2 - x_2\}$ $\qquad q_2(x) = \max\{0, -x_1\}$

Can omit max if $g(x)\ge 0$ e.g. norm

② $\frac{\partial q_1}{\partial x_1} = \begin{cases} 2x_1 & \text{if } x_1^2 - x_2 \ge 0 \\ 0 & \text{otherwise} \end{cases}$ $\qquad \frac{\partial q_1}{\partial x_2} \begin{cases} -1 & \text{if } x_1^2 - x_2 \ge 0 \\ 0 & \text{otherwise} \end{cases}$

$\frac{\partial q_2}{\partial x_1} = \begin{cases} -1 & \text{if } -x_1 \ge 0 \\ 0 & \text{otherwise} \end{cases}$ $\qquad \frac{\partial q_2}{\partial x_2} = 0$

③ $\frac{\partial \phi}{\partial x_1} = 1 + 2c\cdot 2x_1\cdot q_1(x) + 2c(-1)q_2(x)$ $\qquad \nabla f(x) + 2k\cdot \nabla g(x)\,q(x)$ (↑ max)

$\frac{\partial \phi}{\partial x_2} = 1 + 2c\cdot(-1)\,q_1(x) - 2c\,(0)\,q_2(x)$

④ set $\frac{\partial \phi}{\partial x_1} = \frac{\partial \phi}{\partial x_2} = 0$. Think! If this part is 0, can we get $\frac{\partial \phi}{\partial x} = 0$?

$\Rightarrow$ determine case of q_1, q_2.

$\Rightarrow$ alternatively, have 2^q cases, and check validity of x^*.

⑤ Solve x in terms of c. $\quad x = \left(-\frac{1}{2(1+c)}, \frac{1}{4(1+c)^2} - \frac{1}{2c}\right)^T$

⑥ $\lim_{c\to+\infty}$, $x = (0,0)^T$.

Barrier method!

$\Downarrow$

no need to eliminate "max" function

$$\min \ \phi(x, c) = \boxed{f(x) + \frac{1}{c} \cdot B(x)}$$

$$B(x) = \sum_{i=1}^{p} -\frac{1}{g_i(x)}$$

or

$$\sum_{i=1}^{p} -\log(-g_i(x)) \Rightarrow \text{power 1 denominator!}$$

~~Ex~~.

e.g. $f(x) = x_1 + x_2$, $g_1(x) = x_1^2 - x_2 \le 0$, $g_2(x) = -x_1 \le 0$

① $\phi(x, c) = f(x) + \frac{1}{c} B(x)$

$$= x_1 + x_2 + \frac{1}{c}\left(-\log(-x_1^2 + x_2) - \log(x_1)\right)$$

② set $\frac{\partial \phi}{\partial x_1} = \frac{\partial \phi}{\partial x_2} = 0$, $\quad \boxed{\text{i.e. } \nabla f(x(c)) - \frac{1}{c} \sum_{i=1}^{m} \frac{\nabla g_i(x(c))}{g_i(x(c))} = 0}$

$$\begin{cases} 1 - \dfrac{-2x_1}{c(x_2 - x_1^2)} - \dfrac{1}{c x_1} = 0 \\ 1 - \dfrac{1}{c(x_2 - x_1^2)} = 0 \end{cases}$$

$$(x_1, x_2) = \left(\frac{-c \sqrt{c^2 + 8c}}{4c}, \frac{c + 12 - \sqrt{c^2 + 8c}}{8c} \right)$$

③ $\lim_{c \to \infty} (x_1, x_2) = (0, 0)^T$.

MATH 3904 Ch. 8 Duality

Primal problem (P): Min $f(x)$ s.t. $h(x)=0$

Lagrangian Dual (D): $\max_{\lambda}\left(\min_{x}\left(f(x)+h(x)\lambda\right)\right)$ $\qquad \max(\phi(\lambda,\mu))$

$\Rightarrow$ L Min of P, regular pt, $L(x^*)$ PD $\Rightarrow$ L Max of D

① define $\phi(\lambda) = \min\{f(x) + \lambda \cdot h(x)\}$, $\mu \geq 0$

② calculate $\nabla^2 l$, where $l(x,\lambda)$ is objective function of ϕ.

$\Rightarrow$ PD for certain range of λ. (prove convex)

③ Set $\frac{\partial l}{\partial x_1} = \frac{\partial l}{\partial x_2} = 0$, solve x_1, x_2 in terms of λ.

④ Sub x_1, x_2 into l, solve for max λ via diff. $\qquad$ diff again?

$\phi(\mu) = \begin{cases} \text{———} & \text{for some range of } \lambda \rightarrow \text{convex min} \\ -\infty & \text{otherwise} \end{cases}$

e.g. Min $x_1^2 + x_2^2$, s.t. $2x_1 + x_2 - 4 \leq 0$

$l(x) = x_1^2 + x_2^2 + \mu \cdot (2x_1 + x_2 - 4) \Rightarrow$ convex for $\mu \in \mathbb{R}$.

$\frac{\partial l}{\partial x_1} = 2x_1 + 2\mu = 0 \quad \Rightarrow x_1 = -\mu,\ x_2 = -\mu/2.$

$\frac{\partial l}{\partial x_2} = 2x_2 + \mu = 0$

$\phi(\mu) = (-\mu)^2 + (-\frac{\mu}{2})^2 + \mu(-2\mu - \frac{\mu}{2} + 4) = -\frac{5}{4}\mu^2 + 4\mu.$

max $\phi \Rightarrow \mu^* = \frac{8}{5}$, $x = (-\frac{8}{5}, -\frac{4}{5})$?

Convex duality: $f, h, g \in C^2$.

- L Min of P is G Min
- $-\phi(\lambda,\mu)$, $\mu \geq 0$ is convex
- P G Min $\Rightarrow$ D G Max.