

1.3 Views of Functions and Related Hyperplanes

Given $f(x), x \in R^n$, we plot the function in R^n or R^{n+1} space.

1. Draw function in R^n space

Equipotential surface: **tangent plane** $\nabla f(\tilde{x})^T (x - \tilde{x}) = 0$ at \tilde{x}

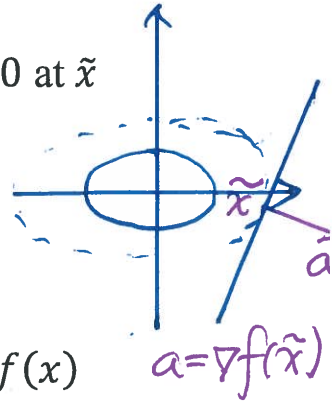
2. Draw function in R^{n+1} space

2.1 Graph of function: $\{(x, h) | x \in \text{dom } f, h = f(x)\}$

hyperplane $(h = \nabla f(\tilde{x})^T (x - \tilde{x}) + f(\tilde{x}))$

Supporting hyperplane $[\nabla f(\tilde{x})^T \quad -1] \left(\begin{bmatrix} x \\ h \end{bmatrix} - \begin{bmatrix} \tilde{x} \\ f(\tilde{x}) \end{bmatrix} \right) = 0$

Example: $f(x) = x^2$. We show the hyperplane with $\nabla f(x)$



2.2. Epigraph (set): $\text{epi } f: \{(x, t) | x \in \text{dom } f, f(x) \leq t\}$

A function is convex iff its epigraph is a convex set.

Example: $f(x) = \max\{f_i(x) | i = 1 \dots r\}$, $f_i(x)$ are convex.

Since $\text{epi } f$ is the intersect of $\text{epi } f_i$, $\text{epi } f$ is convex.

Thus, function f is convex.

If S_i is convex then $\bigcap_i S_i$ is convex.

If f_i is convex then $\max_i f_i(x)$ is convex.

2. Conditions of Optimality: First Order Condition

Definition: f is differentiable if $\text{dom } f$ is open and

$\nabla f(x) \equiv \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \dots, \frac{\partial f(x)}{\partial x_n} \right)^T$ exists at each $x \in \text{dom } f$

Theorem: Differentiable f with convex domain is convex

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

$t = f(x_0) + \nabla f(x_0)^T (x - x_0)$

Proof => If f is convex

Then $(1 - t)f(x) + tf(y) \geq f((1 - t)x + ty), \forall 0 \leq t \leq 1$

$$t[f(y) - f(x)] \geq f(x + t(y - x)) - f(x)$$

$$f(y) - f(x) \geq \frac{1}{t} (f(x + t(y - x)) - f(x))$$

$$= \nabla f(x)^T (y - x) \text{ when } t \rightarrow 0$$

<= Given $f(y) \geq f(x) + \nabla f(x)^T (y - x), \forall x, y \in \text{dom } f$

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

where $\begin{cases} f(x) \geq f(z) + \nabla f(z)^T (x - z) \\ f(y) \geq f(z) + \nabla f(z)^T (y - z) \end{cases}$

Thus $(1 - t)f(x) + tf(y) \geq f(z)$

$f(z) + \nabla f(z)^T (z - z)$

2. Conditions: Second Order Condition

Definition: f is twice differentiable if $dom f$ is open and the Hessian $\nabla^2 f(x) \in S^n$

$$\nabla^2 f(x)_{ij} \equiv \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \quad i, j = 1, \dots, n \text{ exists at each } x \in dom f$$

Theorem: Twice Differentiable f with convex domain is convex
iff $\nabla^2 f(x) \succeq 0, \forall x \in dom f$

$$x^T A x \geq 0 \quad \forall x \in \mathbb{R}^n$$

$$A \in S_+^n$$

Proof: Using Lagrange remainder, we can find a z
 $f(x + t(y - x))$

$$= f(x) + \nabla f(x)^T t(y - x) + \frac{1}{2} t^2 (y - x)^T \nabla^2 f(z) (y - x),$$

$$\forall 0 \leq t \leq 1, z \text{ is between } x \text{ and } x + t(y - x)$$

Since the last term is always positive by assumption, the first order condition is satisfied.

$$f(x + t(y - x)) \geq f(x) + \nabla f(x)^T t(y - x) \quad \text{if } \nabla^2 f(z) \succeq 0$$

11 $\forall z.$

2. Conditions: Second Order Condition

Example: Negative Entropy:

$$f(x) = x \log x, \quad x \in \mathbb{R}_{++}$$

$$f'(x) = \frac{x}{x} + \log x = 1 + \log x, \quad f''(x) = \frac{1}{x}$$

Since $x \in \mathbb{R}_{++}, f''(x) > 0 \Rightarrow f(x)$ is convex

Show the plot of $x \log x$

Remark:

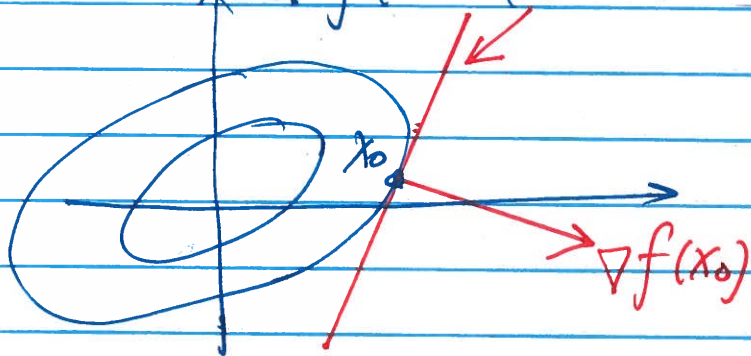
- 1st order condition can be used to design and prove the property of opt. algorithms.
- 2nd order condition implies the 1st order condition
- 2nd order condition can be used to prove the convexity of the functions.

$ax+b$

$$f(x) \approx f(x_0) + \nabla f(x_0)^T (x - x_0)$$

1. Equal Potential ($f(x) = f(x_0)$)

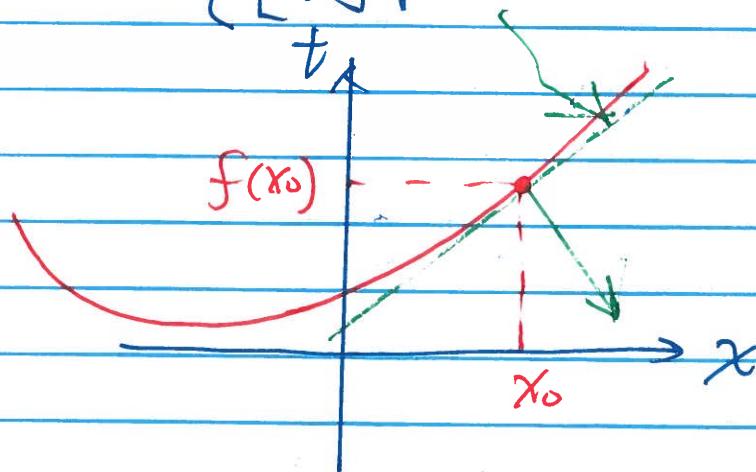
$$\nabla f(x_0)^T (x - x_0) = 0.$$

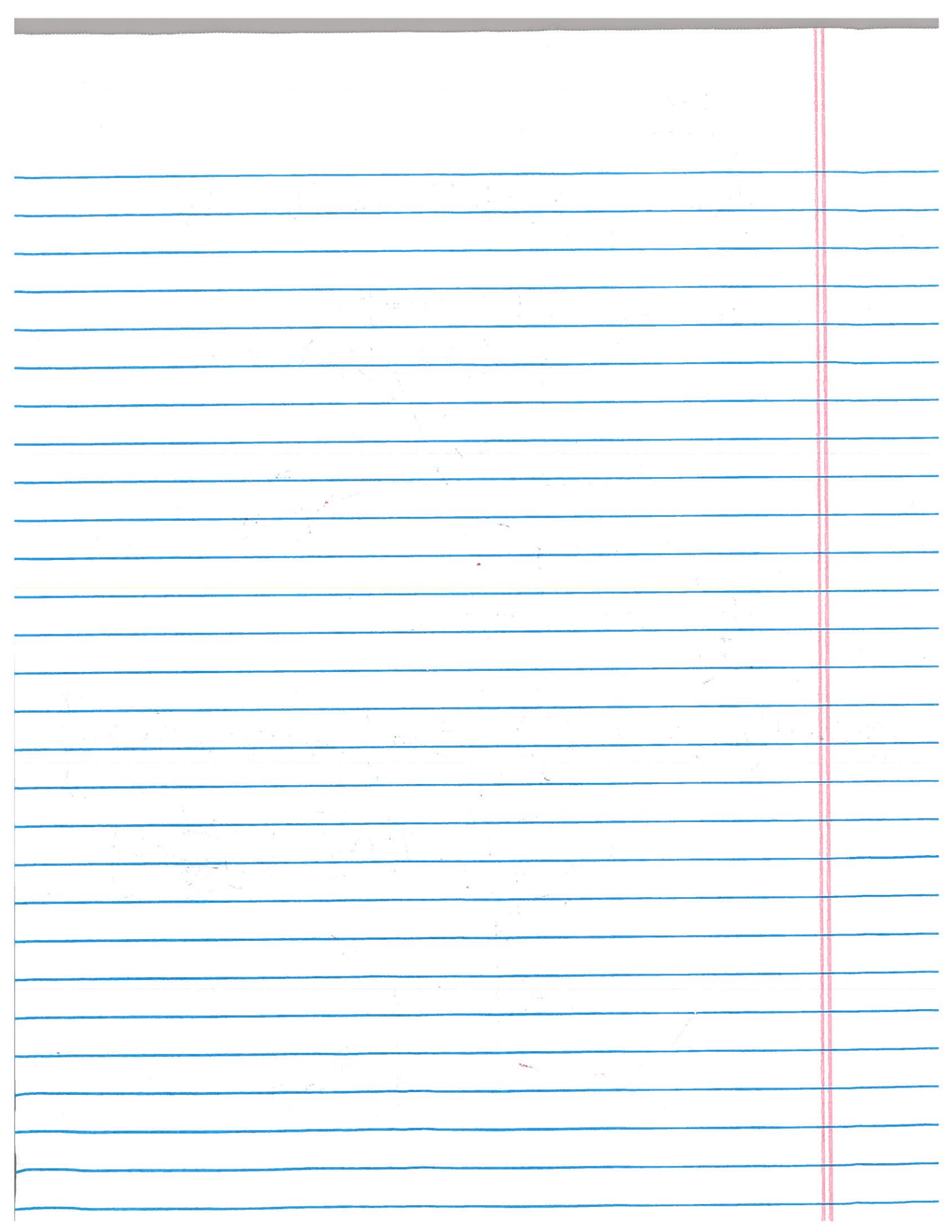


2. $\begin{bmatrix} x \\ t \end{bmatrix}$ where $t = f(x)$.

Supporting Hyperplane $\left\{ \begin{bmatrix} x \\ t \end{bmatrix} \mid t = f(x_0) + \nabla f(x_0)^T (x - x_0) \right\}$

$$= \left\{ \begin{bmatrix} x \\ t \end{bmatrix} \mid \nabla f(x_0)^T \begin{bmatrix} x - x_0 \\ t - f(x_0) \end{bmatrix} = 0 \right\}$$





2. Conditions: Examples

- Quadratic Function: $f(x) = \frac{1}{2}x^T Px + q^T x + r, P \in S^n$
 $\nabla f(x) = Px + q, \nabla^2 f(x) = P$
- Least Square: $f(x) = \|Ax - b\|_2^2$
 $\nabla f(x) = 2A^T(Ax - b), \nabla^2 f(x) = A^T A$
- Quadratic over linear: $f(x, y) = \frac{x^2}{y}, y > 0$

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

$$\nabla^2 f(x, y) = \begin{bmatrix} \frac{2}{y} & -\frac{2x}{y^2} \\ -\frac{2x}{y^2} & \frac{2x^2}{y^3} \end{bmatrix} = \frac{2}{y^3} \begin{bmatrix} y & -x \\ -x & x^2 \end{bmatrix} \begin{bmatrix} y & -x \end{bmatrix}$$

13

2. Conditions: Examples

- Log-sum-exp: $f(x) = \log \sum_{k=1}^n e^{x_k}$ (Smooth max of softmax function)

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

$$v^T \nabla^2 f(x) v = \frac{1}{(1^T z)^2} [(\sum_{i=1}^n z_i)(\sum_{i=1}^n v_i^2 z_i) - (\sum_{i=1}^n v_i z_i)^2] \geq 0,$$

for all $v \in R^n$ (Cauchy-Schwarz inequality)

Thus, $f(x)$ is a convex function

Cauchy-Schwarz inequality: $[(a^T a)(b^T b) \geq (a^T b)^2, a_i = \sqrt{z_i}, b_i = v_i \sqrt{z_i}]$

Proof 1: Let $z = a - \frac{a^T b}{b^T b} b$, or $a = z + \frac{a^T b}{b^T b} b$

We have

$$a^T a = z^T z + \frac{(a^T b)^2}{(b^T b)^2} b^T b \geq \frac{(a^T b)^2}{(b^T b)^2} b^T b = \frac{(a^T b)^2}{b^T b}$$

Proof 2: By induction

14

3. Operations that preserve convexity

- Nonnegative multiple: αf , where $\alpha \geq 0$, f is convex
- Sum: $f_1 + f_2$, where f_1 , and f_2 are convex
- Composition with affine function: $f(Ax + b)$, where f is convex

Proof: $\nabla_x^2 f(Ax + b) = A^T \nabla_y^2 f(y|y = Ax + b)A$

E.g. $f(x) = -\sum_{i=1}^m \log(b_i - a_i^T x_i)$,

$dom f = \{x | a_i^T x < b_i, i = 1, \dots, m\}$

$f(x) = \|Ax + b\|$ (if f is twice differentiable)

$f(y)$ is convex $\Leftrightarrow \nabla_y^2 f(y) \geq 0$.

$f(Ax + b)$

$\nabla_x^2 f(Ax + b) \geq 0$ (to prove)

3. Operations that preserve convexity

- Pointwise maximum: $f(x) = \max\{f_1(x), \dots, f_r(x)\}$, f_i are convex
- Pointwise supremum:
 $g(x) = \sup_{y \in C} f(x, y)$, where $f(x, y)$ is convex in x and C is an arbitrary set

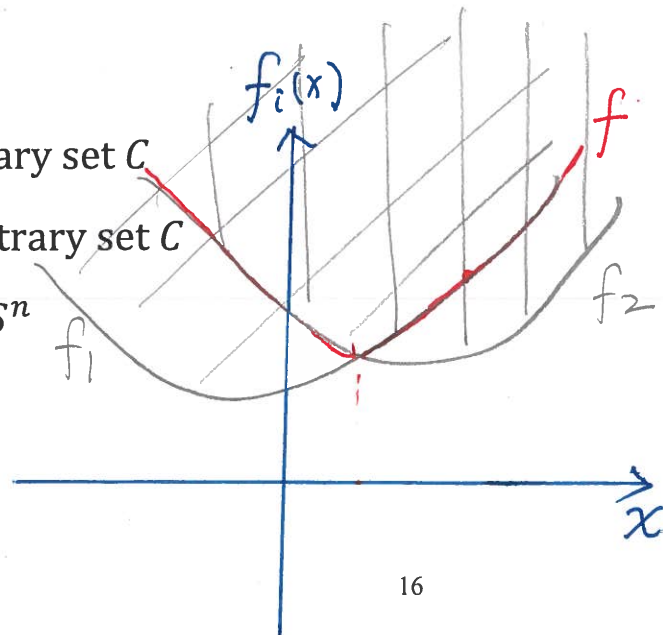
Examples

- $S_c(x) = \sup_{y \in C} y^T x$, for an arbitrary set C

- $f(x) = \sup_{y \in C} \|x - y\|$, for an arbitrary set C

- $\lambda_{max}(X) = \sup_{\|y\|_2=1} y^T X y$, $X \in S^n$

$\sum_{i,j} x_{ij} y_i y_j$



3. Operations that preserve convexity: Dual norm

Example:

$$f(x) = \max_{\|y\|_2 \leq 1} y^T x$$

$$f(x) = \max_{\|y\|_1 \leq 1} y^T x$$

$$f(x) = \max_{\|y\|_p \leq 1} y^T x$$

$$\|y\|_p = \left(\sum_i |y_i|^p \right)^{1/p}$$

17

3. Operations that preserve convexity: max function

Theorem: Pointwise maximum of convex functions is convex

Given $f(x) = \max\{f_1(x), f_2(x)\}$, where f_1 and f_2 are convex and $\text{dom } f = \text{dom}\{f_1\} \cap \text{dom}\{f_2\}$ is convex, then $f(x)$ is convex.

Proof: For $0 \leq \theta \leq 1$, $x, y \in \text{dom } f$

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

$$= \max\{f_1(\theta x + (1 - \theta)y), f_2(\theta x + (1 - \theta)y)\}$$

$$\leq \max\{\theta f_1(x) + (1 - \theta)f_1(y), \theta f_2(x) + (1 - \theta)f_2(y)\}$$

$$\leq \theta \max\{f_1(x), f_2(x)\} + (1 - \theta) \max\{f_1(y), f_2(y)\}$$

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

$$\text{i.e. } f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$$

Thus, function $f(x)$ is convex.

18

3. Operations that preserve convexity: minimization

Theorem: Partial minimization

If $g(x, y)$ is convex in x and y , and a set C is convex

Then $f(x) = \min_{y \in C} g(x, y)$ is convex.

Proof: Let $y_1 \in \{y \mid \min_{y \in C} g(x_1, y)\}$ and $y_2 \in \{y \mid \min_{y \in C} g(x_2, y)\}$,

we can write

$$\begin{aligned} & \theta f(x_1) + (1 - \theta)f(x_2) \\ &= \theta g(x_1, y_1) + (1 - \theta)g(x_2, y_2) \\ &\geq g(\theta x_1 + (1 - \theta)x_2, \theta y_1 + (1 - \theta)y_2) \text{ (} g \text{ is convex)} \\ &\geq \min_{y \in C} g(\theta x_1 + (1 - \theta)x_2, y) \text{ (} C \text{ is convex)} \\ &= f(\theta x_1 + (1 - \theta)x_2) \end{aligned}$$

i.e. we have $\theta f(x_1) + (1 - \theta)f(x_2) \geq f(\theta x_1 + (1 - \theta)x_2)$

Therefore, $f(x) = \min_{y \in C} g(x, y)$ is convex.

19

3. Operations that preserve convexity

Examples for Partial Minimization

$$\text{Given } f(x, y) = \begin{bmatrix} x^T & y^T \end{bmatrix} \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

$$x \in R^n, y \in R^m, A \in S_+^n, C \in S_+^m, \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \in S_+^{n+m}$$

$$\text{Let } g(x) = \min_y f(x, y) = x^T (A - BC^+B^T)x,$$

C^+ : **pseudo inverse** of matrix C . (**Drazin inverse**, or **generalized inverse**)

We can claim that function $g(x)$ is convex.

Proof:

(1) $f(x, y)$ is convex

(2) $y \in R^m$ where R^m is a convex non-empty set

(3) Therefore, $g(x)$ is convex, i.e. $A - BC^+B^T \succcurlyeq 0$

20