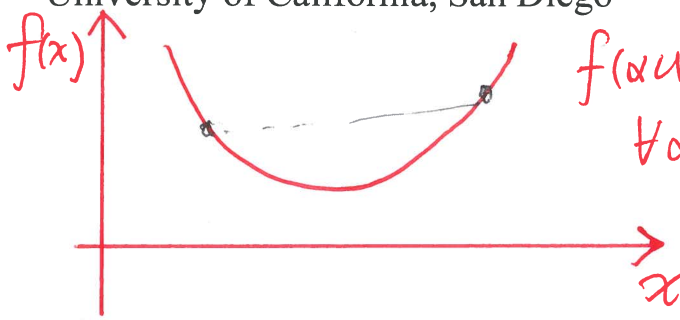
CSE203B Convex Optimization: Lecture 3: Convex Function

CK Cheng
Dept. of Computer Science and Engineering
University of California, San Diego



$f(\alpha u+\beta v) \le \alpha f(u)+\beta f(v)$
$\forall \alpha+\beta=1, \alpha, \beta \ge 0.$

Outlines

1. Definitions: Convexity, Examples & Views
2. Conditions of Optimality Taylor's exp.
   1. First Order Condition
   2. Second Order Condition
3. Operations that Preserve the Convexity
   1. Pointwise Maximum
   2. Partial Minimization
4. Conjugate Function
5. Log-Concave, Log-Convex Functions

W3A

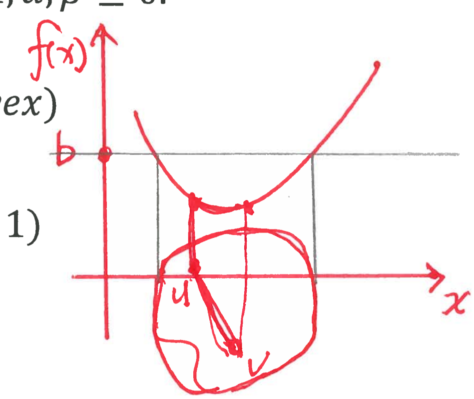
Outlines

1. Definitions
    1. Convex Function vs. Convex Set Obj. & Constraints
    2. Examples
        1. Norm
        2. Entropy
        3. Affine
        4. Determinant
        5. Maximum
    3. Views of Functions and Related Hyperplanes

1. Definitions: Convex Function vs Convex Set

Theorem: Given $S = \{x | f(x) \leq b\}$
If function $f(x)$ is convex, then S is a convex set.
Proof: We prove by the definition of convex set.
For every $u, v \in S$, i.e. $f(u) \leq b, f(v) \leq b$,
We want to show that $\alpha u + \beta v \in S, \quad \forall \alpha + \beta = 1, \alpha, \beta \geq 0$.
We have

$$
\begin{aligned}
f(\alpha u + \beta v) &\leq \alpha f(u) + \beta f(v) \quad (f \text{ is convex}) \\
&\leq \alpha b + \beta b \qquad\qquad (\alpha, \beta \geq 0) \\
&= (\alpha + \beta) \cdot b = b \quad (\alpha + \beta = 1)
\end{aligned}
$$



Thus $\alpha u + \beta v \in S$

Remark: Convex function => Convex Set

$S_1 = \{x \mid f(x) \leq b\}$ => Convex Set

$S_2 = \{x \mid f(x) \geq b\}$ => ?

1. Convex Function Definitions: Examples

$f: R^n \rightarrow R$ is convex if $dom\ f$ is a convex set and

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

$$\forall x, y \in dom\ f, 0 \leq \theta \leq 1$$

Example on R:

Convex Functions

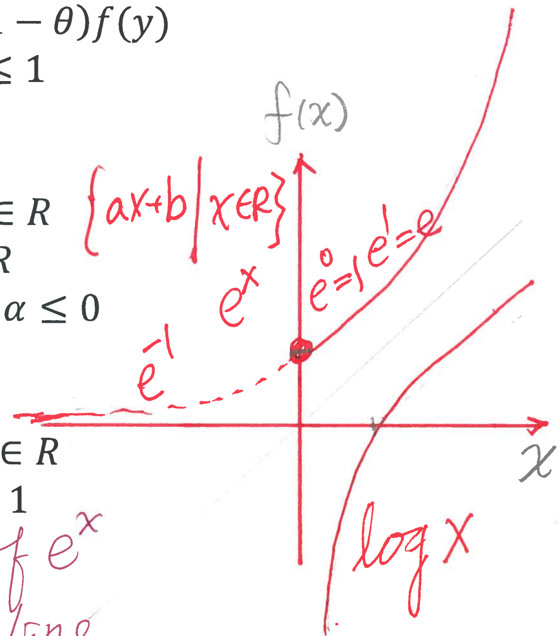
Affine:	$ax + b$ on R	for any $a, b \in R$
Exponential:	e^{ax}	for any $a \in R$
Power:	x^{α} on R_{++}	for $\alpha \geq 1$ or $\alpha \leq 0$
	$\lvert x\rvert^p$ on R	for $p \geq 1$

Concave Functions

Affine:	$ax + b$ on R	for any $a, b \in R$
Power:	x^{α} on R_{++}	for $0 \leq \alpha \leq 1$
Logarithm:	$log x$ on R_{++}	−Mirror of e^x at 45° line

Inverse funct. of e^x i.e. $e^{\log x} = x$

$\log e^x = x$



1. Convex Function Definitions: Examples

Example on R^n:

Affine: $f(x) = a^T x + b$

Norms: $\|x\|_p = (\sum_{i=1}^{n} |x_i|^p)^{1/p}$ for $p \geq 1$;

$\|x\|_{\infty} = \max_k |x_k|$

Example on $R^{m \times n}$:

Affine: $f(X) = tr(A^T X) = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} x_{ij}$

Spectral (max singular value):

$$f(X) = \|X\|_2 = \sigma_{max}(X) = (\lambda_{max}(X^T X))^{1/2}$$

1. Convex Function Definitions: Examples

Concave Functions:

$X^{1/2} X^{1/2}$

Log Determinant: $f(X) = \log\det X$, $dom\ f = S_{++}^n$

Proof: Let $g(t) = f(X + tV)$ $(V \in S^n)$

$$g(t) = log\ det\ (X + tV) = log\ det\ X + logdet(I + tX^{-\frac{1}{2}}VX^{-\frac{1}{2}})$$

$$= log\ det\ X + \sum_{i=1}^n log(1 + t\lambda_i)$$

$1+t\lambda_i > 0.$ $D\Sigma D^T$ $DD^T = I$

λ_i: $eigenvalue\ of\ X^{-\frac{1}{2}}VX^{-\frac{1}{2}}$

g is concave in $t \Rightarrow f$ is concave

$\log ab = \log a + \log b$

$\det AB = \det A \det B$

$\det X = \prod_{i=1}^{n} \lambda_i$

Convex function examples: norm, max, expectation $f_1(x) + \|x\|_1$

norm: If $f: R^n \to R$ is a norm and $0 \le \theta \le 1$

$$f(\theta x + (1-\theta)y) \le f(\theta x) + f((1-\theta)y) \quad \textit{triangle inequality}$$

$$= \theta f(x) + (1-\theta)f(y) \quad \textit{scalability}$$

$\max\left(\begin{bmatrix}1\\2\\3\end{bmatrix} + \begin{bmatrix}4\\3\\2\end{bmatrix}\right) = 5$

Max function: $f(x) = \max_i x_i$, $x = [x_1, x_2, \dots, x_n]^T$

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

example

$$\le \theta \max_i x_i + (1-\theta)\max_i y_i$$

$\left(\max\begin{bmatrix}1\\2\\3\end{bmatrix} + \max\begin{bmatrix}4\\3\\2\end{bmatrix}\right) = 7$

$$= \theta f(x) + (1-\theta)f(y) \quad \text{for } 0 \le \theta \le 1$$

Probability: (Expectation)

If $f(x)$ is convex with $p(x)$ a probability at x,

i.e. $p(x) \ge 0, \forall x$ and $\int p(x)\,dx = 1$

Then $f(Ex) \le Ef(x)$,

where $Ex = \int x\, p(x) dx$ $\longrightarrow \sum p_i(x_i) x_i \to \sum \theta_i x_i$

$Ef(x) = \int f(x)\, p(x) dx$ $\quad \sum f(x_i) p(x_i) \to \sum \theta_i f(x_i)$

$\sum_i \theta_i f(x_i) \ge f(\sum \theta_i x_i)$

1.3 Views of Functions and Related Hyperplanes

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

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: $\{(x,h) | x \in dom\, f, h = f(x)\}$

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

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

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

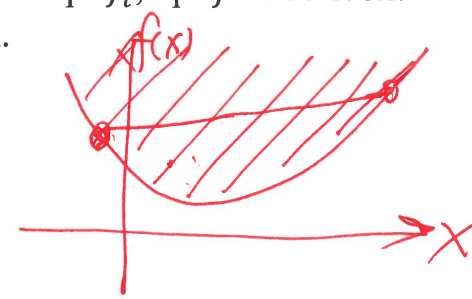
   2.2. Epigraph: epi f: $\{(x,t) | x \in dom\, f, f(x) \le t\}$

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

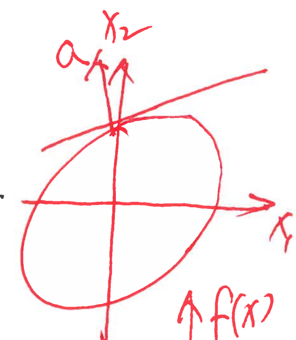
   Example: $f(x) = \max\{f_i(x) | \, i = 1 \ldots r\}$, $f_i(x)$ *are convex.*

   Since epi f is the intersect of epi f_i, epi f is convex.

   Thus, function f is convex.



2. Conditions of Optimality: First Order Condition

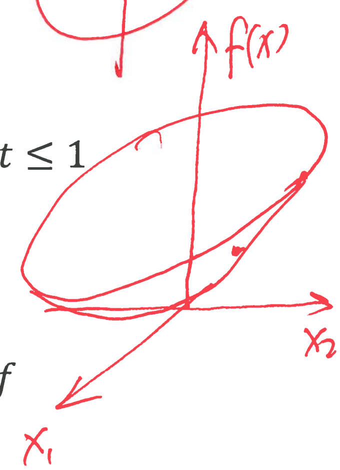


Defintion: f is differentiable if $dom f$ is open and

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

Theorem: Differentiable f with convex domain is convex

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



Proof => If f is convex

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

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

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

$$= \nabla f(x)(y-x) \quad when\ t \to 0$$

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

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

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

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

W3B

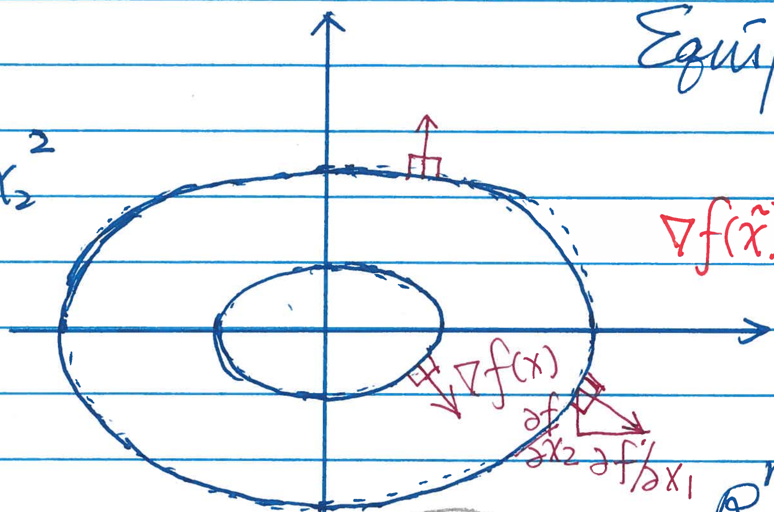
View of Functions

$f(x_1,x_2) = ax_1^2+bx_2^2$

$(a,b>0)$

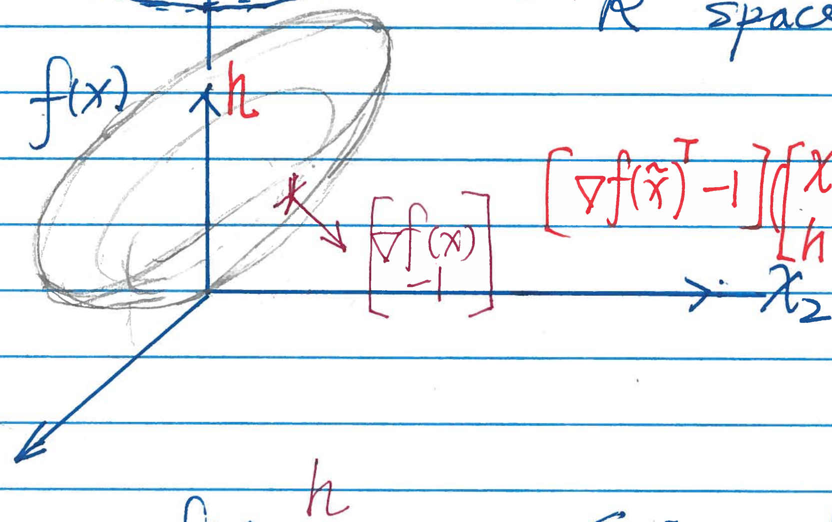
Equipotential Surface

$\nabla f(\tilde{x})^T(x-f(\tilde{x}))=0$

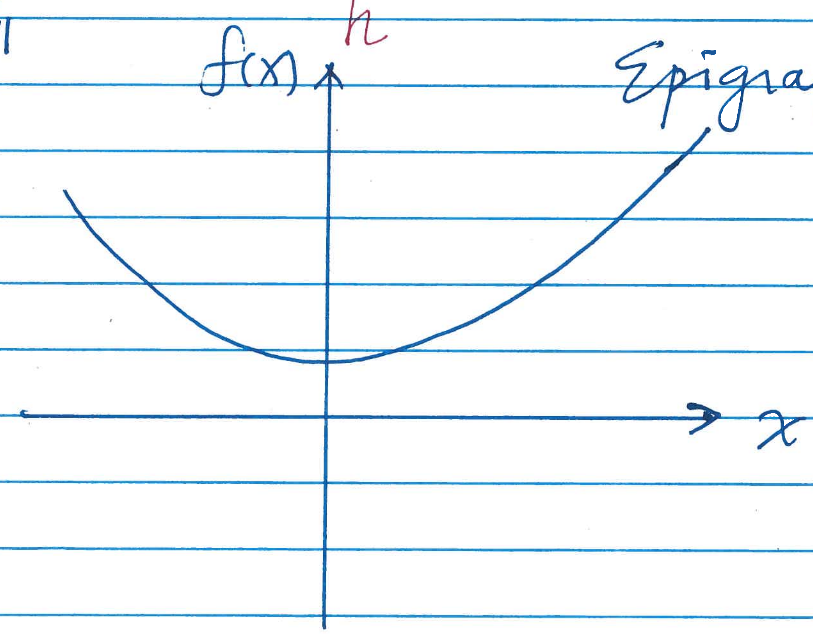


R^{n+1} space

$$\begin{bmatrix}\nabla f(\tilde{x})^T & -1\end{bmatrix}\left(\begin{bmatrix}x\\ h\end{bmatrix}-\begin{bmatrix}\tilde{x}\\ f(\tilde{x})\end{bmatrix}\right)=0$$



Epigraph



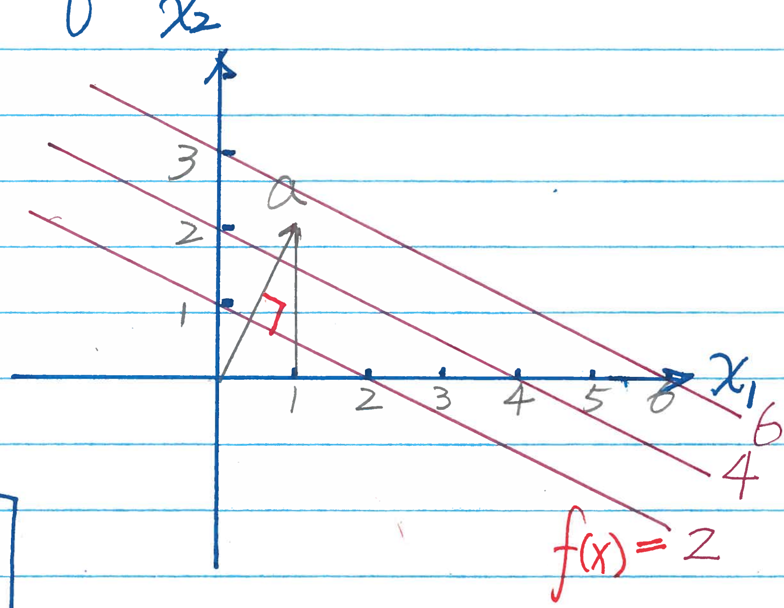
A.9.1

Equipotential Surface

Sensitivity

$$f(x) = a^T x$$

e.g. $\begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$



or $\nabla f(x) = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

(1) $\Delta x_1 = 1 \Rightarrow f\left(\begin{bmatrix} x_1 + \Delta x_1 \\ x_2 \end{bmatrix}\right) - f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = 1$

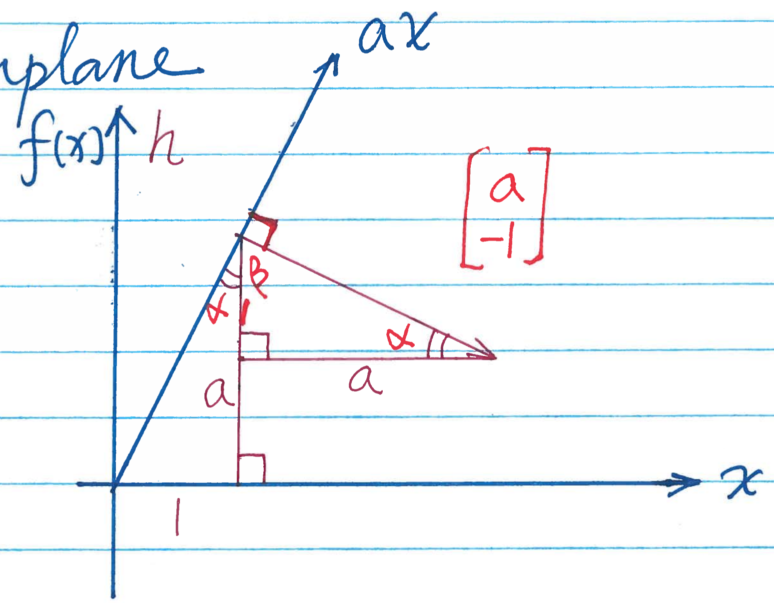
(2) $\Delta x_2 = 1 \Rightarrow f\left(\begin{bmatrix} x_1 \\ x_2 + \Delta x_2 \end{bmatrix}\right) - f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = 2$

Remark

$a^T x = b$ is an equipotential hyperplane of $f(x) = a^T x$.

$a = \nabla f(x)$ is the sensitivity of $f(x) = a^T x$.

Slope & Hyperplane



$f(x) = ax$

$\nabla f(x) = a$

$$\hat{a}^T\left(\begin{bmatrix} x \\ h \end{bmatrix} - \begin{bmatrix} x_0 \\ f(x_0) \end{bmatrix}\right) = 0$$

$$\hat{a}^T\left(\begin{bmatrix} x_0 + \Delta x \\ f(x_0) + \Delta h \end{bmatrix} - \begin{bmatrix} x_0 \\ f(x_0) \end{bmatrix}\right) = 0$$

Since $\Delta h = \nabla f(x)^T \Delta x$

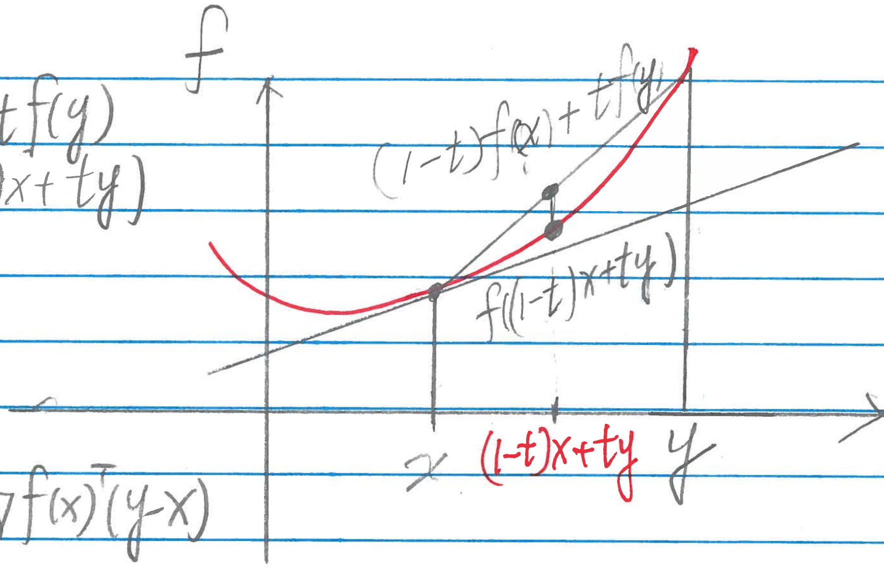
or $\nabla f(x)^T \Delta x - \Delta h = 0$

We have $\hat{a} = \begin{bmatrix} \nabla f(x) \\ -1 \end{bmatrix}$

A.9.3

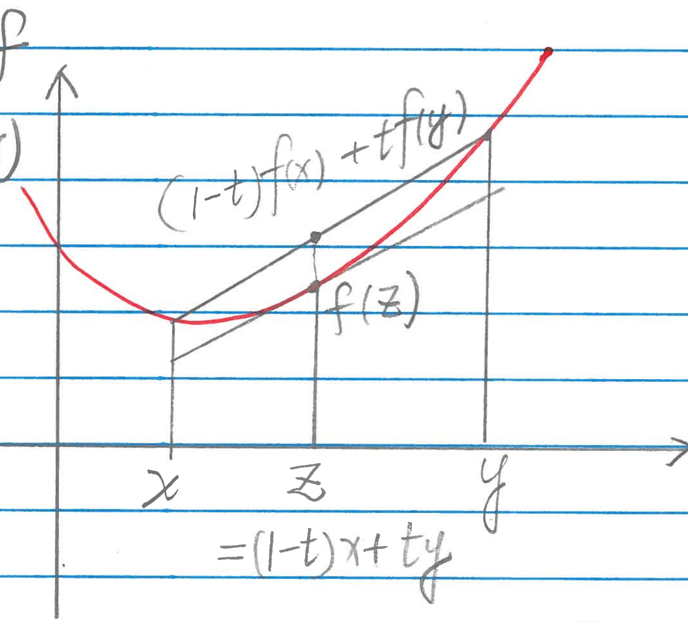
First Order Condition

If $(1-t)f(x)+tf(y) \geq f((1-t)x+ty)$



Then $f(y) \geq f(x)+\nabla f(x)^T(y-x)$

If $f(y) \geq f(x)+\nabla f(x)^T(y-x)$



Then $(1-t)f(x)+tf(y) \geq f((1-t)x+ty)$

① $f(x) \geq f(z)+\nabla f(z)^T(x-z)$

② $f(y) \geq f(z)+\nabla f(z)^T(y-z)$