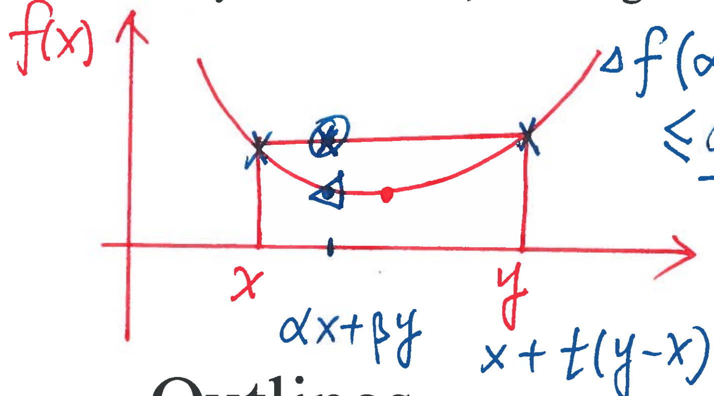
CSE203B Convex Optimization: Lecture 3: Convex Function

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



$f(\alpha x+\beta y) \leq \alpha f(x)+\beta f(y)$

Outlines

1. Definitions: Convexity, Examples & Views
2. Conditions of Optimality   Taylor's Ex.
    1. First Order Condition $f(x)=f(x_0)+\nabla f(x_0)^T(x-x_0)$
    2. Second Order Condition $+\frac{1}{2}(x-x_0)^T\nabla^2 f(x_0)(x-x_0)$ $+\cdots$
3. Operations that Preserve the Convexity
    1. Pointwise Maximum $g(x)=\max_i f_i(x)$ if $f_i(x)$ is convex for all i then $g(x)$ is convex
    2. Partial Minimization $h(x)=\min_y f(x,y)$ if $f(x,y)$ is convex then $h(x)$ is a convex function.
4. Conjugate Function
5. Log-Concave, Log-Convex Functions

$g(x)=\max_y f(x,y)$ if $f(x,y)$ is convex w.r.t. x then $g(x)$ is convex.

Outlines

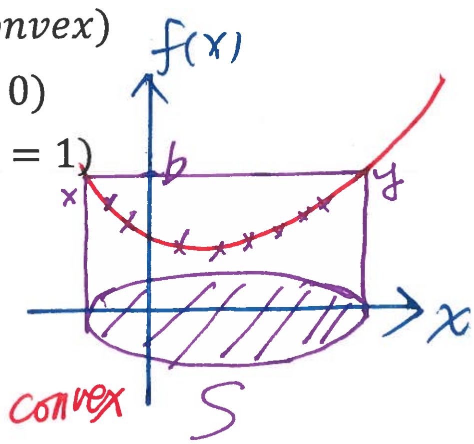
1. Definitions
   1. Convex Function vs Convex Set
   2. Examples
      1. Norm
      2. Entropy
      3. Affine
      4. Determinant
      5. Maximum
   3. Views of Functions and Related Hyperplanes

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

$S_2 = \{x \mid f(x) \geq b\}$ => ? May Not be convex if $f(x)$ is convex

is convex when $f(x)$ is concave

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: $dom\ f \in R$

Convex Functions

Affine: $ax + b\ on\ R$ for any $a, b \in R$

Exponential: e^{ax} for any $a \in R$

Power: $x^{\alpha}\ on\ R_{++}$ for $\alpha \geq 1$ or $\alpha \leq 0$

$|x|^p\ on\ R$ for $p \geq 1$

Concave Functions

Affine: $ax + b\ on\ R$ for any $a, b \in R$

Power: $x^{\alpha}\ on\ R_{++}$ for $0 \leq \alpha \leq 1$

Logarithm: $logx\ on\ R_{++}$

1. Convex Function Definitions: Examples

Example: $dom\ f \in R^n$

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: $dom\ f \in R^{m \times n}$

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:

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

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

$$g(t) = log\ det\ (X + tV) = log\ det\ X + logdet(I + tX^{-\frac{1}{2}}VX^{-\frac{1}{2}})$$
$$= log\ det\ X + \sum_{i=1}^{n} log(1 + t\lambda_i)$$
$$\lambda_i: eigenvalue\ of\ X^{-\frac{1}{2}}VX^{-\frac{1}{2}}$$

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

Convex function examples: norm, max, expectation

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

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

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

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

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

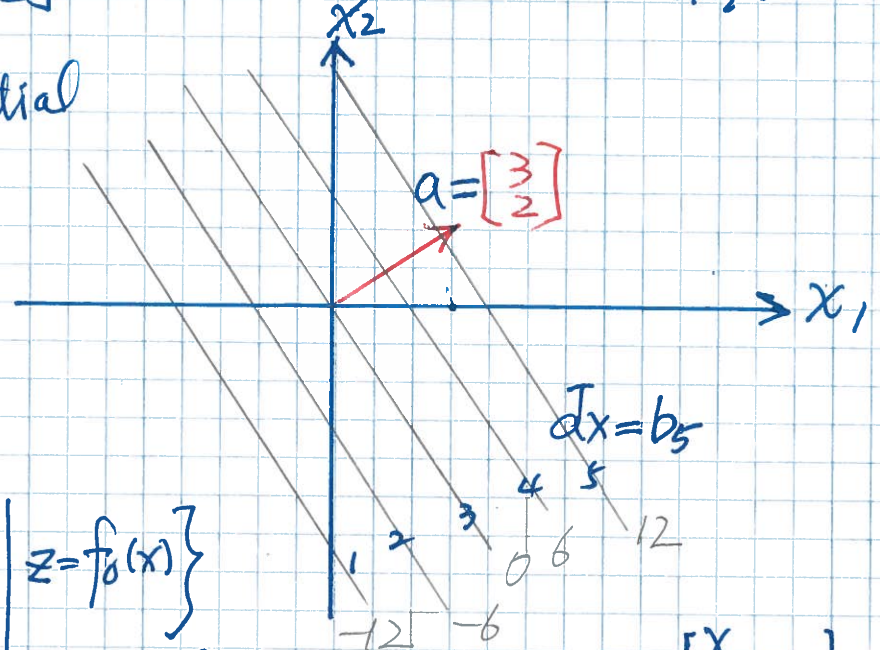
where $Ex = \int x\, p(x)dx$

$Ef(x) = \int f(x)\, p(x)dx$

Ex: $f_0(x) = a^T x \quad a = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \quad \nabla f_0(x) = a$

$$f_0(x) = \begin{bmatrix} 3 \\ 2 \end{bmatrix}^T x = 3x_1 + 2x_2 \qquad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Equal Potential Plot.

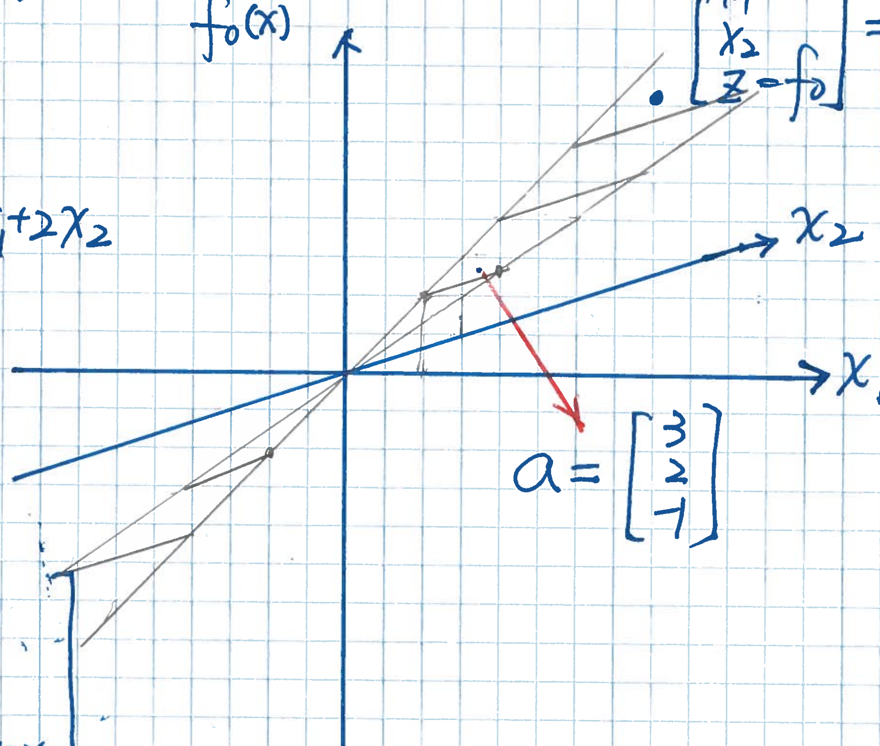


Set Plot $\left\{ \begin{bmatrix} x_1 \\ x_2 \\ z \end{bmatrix} \middle| z = f_0(x) \right\}$

$$\begin{bmatrix} x_1 \\ x_2 \\ z = f_0 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ 3x_1 + 2x_2 \end{bmatrix}$$

$$a \begin{bmatrix} x_1 \\ x_2 \\ z \end{bmatrix} = 0$$

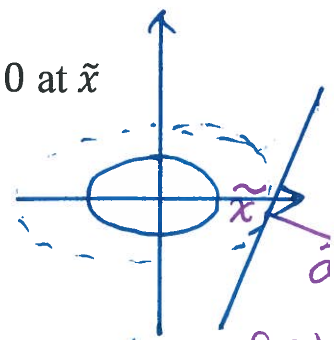
where $z = 3x_1 + 2x_2$



$$\begin{bmatrix} 3 & 2 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ 3x_1 + 2x_2 \end{bmatrix}$$

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

   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 $\nabla f(x)$ $a = \nabla f(\tilde{x})$

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

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

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

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

   Thus, function f is convex. $\left\{\begin{bmatrix} x \\ t \end{bmatrix} \middle| \; t \ge f_i \; \forall i\right\}$

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

If f_i is convex $\forall i$ then $\max_i f_i(x)$ 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

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

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

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

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

<= *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 $(1-t)f(x) + tf(y) \ge f(z)$

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 $\nabla^2 f(x) \succcurlyeq 0, \forall x \in dom f$

Proof: Using Lagrange remainder, we can find a z

$$f(x + t(y - x))$$
$$= f(x) + \nabla f(x)^T t(y - x) + \frac{1}{2} t^2 (y - x)^T \nabla^2 f(z)(y - x),$$
$$\forall 0 \le t \le 1, z \text{ is between } x \text{ and } x + t(y - x)$$

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

2. Conditions: Second Order Condition

Example: Negative Entropy:

$f(x) = x \log x, x \in R_{++}$

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

Since $x \in R_{++}, f''(x) > 0 \Rightarrow f(x)$ 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.

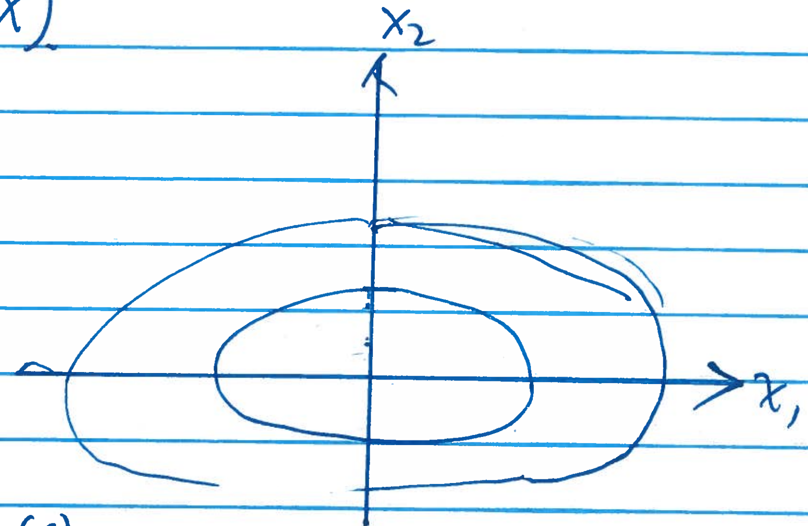
View of function $f(x)$

Equal Potential Plot.

$$f(x_1, x_2) = a x_1^2 + b x_2^2$$

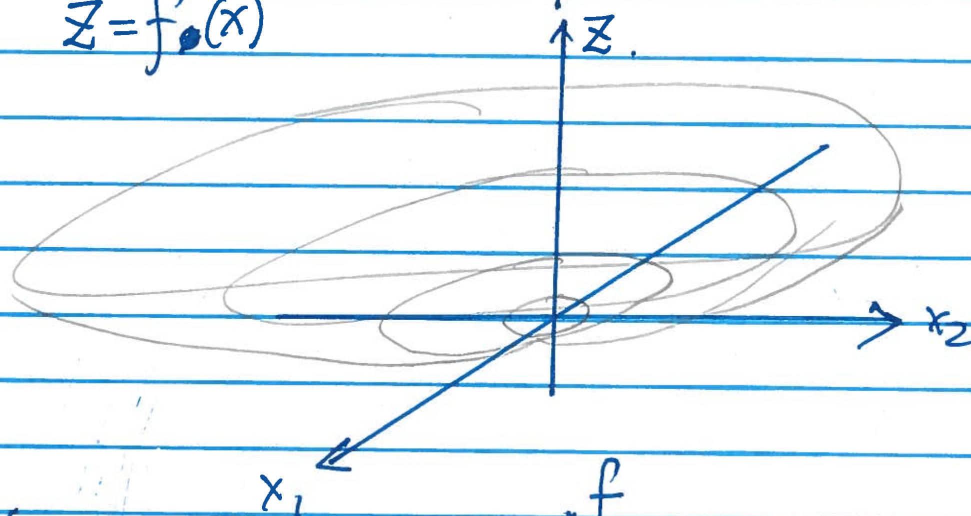
$a \geq 0$

$b \geq 1$



Extra Dimension $z = f_0(x)$

$$\begin{bmatrix} x \\ z \end{bmatrix}$$

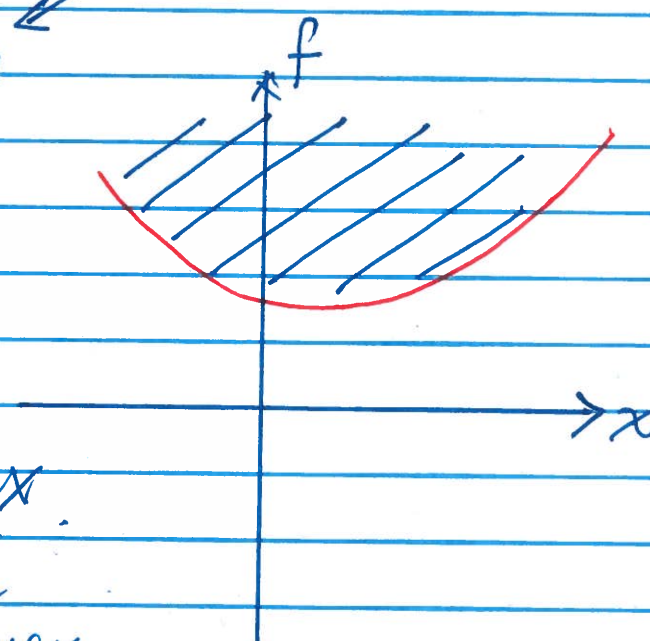


Epigraph (Set)

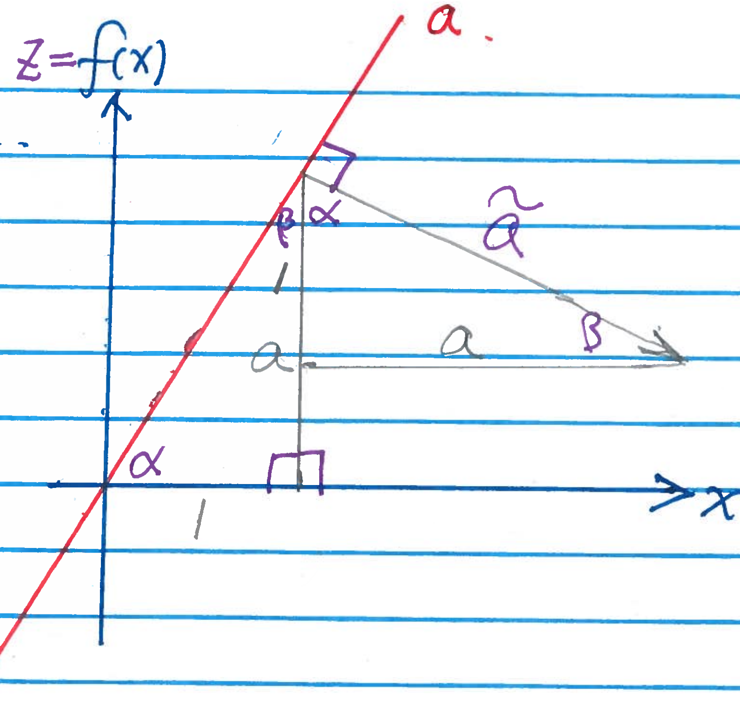
$$E = \left\{ \begin{bmatrix} x \\ z \end{bmatrix} \middle| \; z \geq f(x) \right\}$$

If $f(x)$ is convex then E is convex.

If E is convex then $f(x)$ is convex.



Slope & Hyperplane.



$f(x) = a^T x,\ x \in R.$

$\nabla f(x) = a.$

Extra Dimension

$$E = \left\{ \begin{bmatrix} x \\ z \end{bmatrix} \middle| z = f(x) \right\}$$

$$= \left\{ \begin{bmatrix} x \\ z \end{bmatrix} \middle| z = a^T x \right\} = \left\{ \begin{bmatrix} x \\ z \end{bmatrix} \middle| \begin{bmatrix} a & -1 \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} = 0 \right\}$$

$$a^T x - z = 0.$$

$$\Rightarrow \begin{bmatrix} a^T & -1 \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} = 0.$$

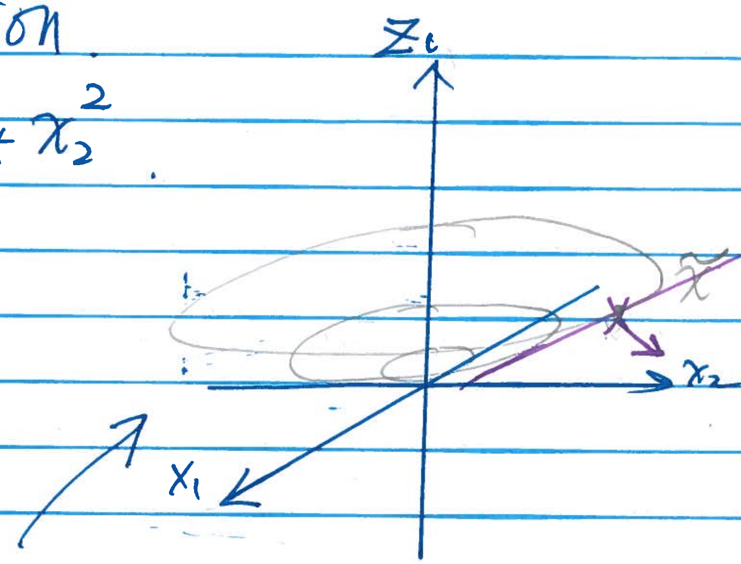
$$E = \left\{ \begin{bmatrix} x \\ z \end{bmatrix} \middle| \tilde{a}^T \begin{bmatrix} x \\ z \end{bmatrix} = 0 \right\}$$

where $\tilde{a} = \begin{bmatrix} a \\ -1 \end{bmatrix} \left(\begin{bmatrix} \nabla f(x) \\ -1 \end{bmatrix} \right)$

(Supporting Hyperplane)

Draw in R^{n+1} dimension.

Ex: $f\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right) = x_1^2 + x_2^2$.



$$\left\{ \begin{bmatrix}x_1\\x_2\\z\end{bmatrix} \;\middle|\; \begin{array}{l} z = x_1^2 + x_2^2 \\ z-\tilde{z} = \nabla f\left(\begin{bmatrix}\tilde{x}_1\\\tilde{x}_2\end{bmatrix}\right)^T \begin{bmatrix}x_1-\tilde{x}_1\\x_2-\tilde{x}_2\end{bmatrix} \end{array} \right\}$$

Hyperplane at $\tilde{x} \in \tilde{x}_1$, $\tilde{z} = \tilde{x}_1^2 + \tilde{x}_2^2$.

$$\nabla f(\tilde{x}) = 2\tilde{x}_1 + 2\tilde{x}_2.$$

$$\cancel{\left\{ \begin{bmatrix}\tilde{x}_1\\\tilde{x}_2\\\tilde{z}\end{bmatrix} \;\middle|\; h = 2\tilde{x}_1 + 2\tilde{x}_2 \right.}$$

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

Hyperplane $\left\{ \begin{bmatrix}x_1\\x_2\\z\end{bmatrix} \;\middle|\; h = z - \tilde{z} \quad h = (x_1-\tilde{x}_1)(2\tilde{x}_1) + (x_2-\tilde{x}_2)2\tilde{x}_2 \right.$

$$= \left\{ \begin{bmatrix}x_1\\x_2\\z\end{bmatrix} \;\middle|\; \begin{bmatrix}\nabla f(\tilde{x})\\-1\end{bmatrix}^T \left(\begin{bmatrix}x_1\\x_2\\z\end{bmatrix} - \begin{bmatrix}\tilde{x}_1\\\tilde{x}_2\\\tilde{z}\end{bmatrix} \right) = 0 \right\}$$