Linear Algebra & Convex Optimization – Lecture 12

Convex Functions

Text : Convex Optimization , S. Boyd

Matrix Calculus

Some Common Derivatives

(in denominator layout convention)

u & **v** are functions of vector **x**
A & **b** are independent of vector **x**

$$\frac{\partial \mathbf{Ax}}{\partial \mathbf{x}} = \mathbf{A}^\top$$

$$\frac{\partial \mathbf{u}^\top \mathbf{v}}{\partial \mathbf{x}} = \frac{\partial \mathbf{u}}{\partial \mathbf{x}}\mathbf{v} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}}\mathbf{u}$$

$$\frac{\partial \mathbf{x}^\top \mathbf{A}}{\partial \mathbf{x}} = \mathbf{A}$$

$$\checkmark \quad \frac{\partial \mathbf{b}^\top \mathbf{Ax}}{\partial \mathbf{x}} = \mathbf{A}^\top \mathbf{b}$$

$$\frac{\partial \mathbf{Au}}{\partial \mathbf{x}} = \frac{\partial \mathbf{u}}{\partial \mathbf{x}}\mathbf{A}^\top$$

$$\frac{\partial \mathbf{x}^\top \mathbf{Ax}}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}^\top)\,\mathbf{x}$$

$$Ax = y \qquad f(x) = \|Ax - y\|^2$$

$$= (Ax - y)^T (Ax - y)$$

$$= (x^T A^T - y^T)(Ax - y)$$

$$f(x) = \underbrace{x^T A^T A x} - \underbrace{x^T A^T y} - \underbrace{y^T A x} + \underbrace{y^T y}$$

$$\nabla f(x) = 2A^T A x - 2A^T y$$

$$\frac{\partial}{\partial x} x^T A^T A x = (A^T A + (A^T A)^T)\, x = 2A^T A x$$

$$\frac{\partial}{\partial x} y^T A x = A^T y$$

$$\frac{\partial}{\partial x} x^T A^T y = A^T y$$

$$\nabla f(x) = 2A^T A x - 2A^T y = 0$$

$$A^T A x = A^T y$$

$$x = \underbrace{(A^T A)^{-1} A^T}\, y$$

$$[(x^T A^T y)]^T = y^T \underbrace{Ax}$$

General Optimization Formulation

$x \in \mathbb{R}^n$

General Formulation:

$x^2 - 2 < 0$

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

$$p^\star = \inf\{f_0(x) \mid f_i(x) \le 0,\ i = 1, \ldots, m,\ h_i(x) = 0,\ i = 1, \ldots, p\}$$

x is **feasible** if $x \in \mathbf{dom}\, f_0$ and it satisfies the constraints

a feasible x is **optimal** if $f_0(x) = p^\star$; X_{opt} is the set of optimal points

Implicit Constraints:

$$x \in \mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom}\, f_i \ \cap \ \bigcap_{i=1}^{p} \mathbf{dom}\, h_i$$

Explicit Constraints:

$$f_i(x) \le 0,\ h_i(x) = 0$$

Un-Constraint Problem:

$$m = p = 0$$

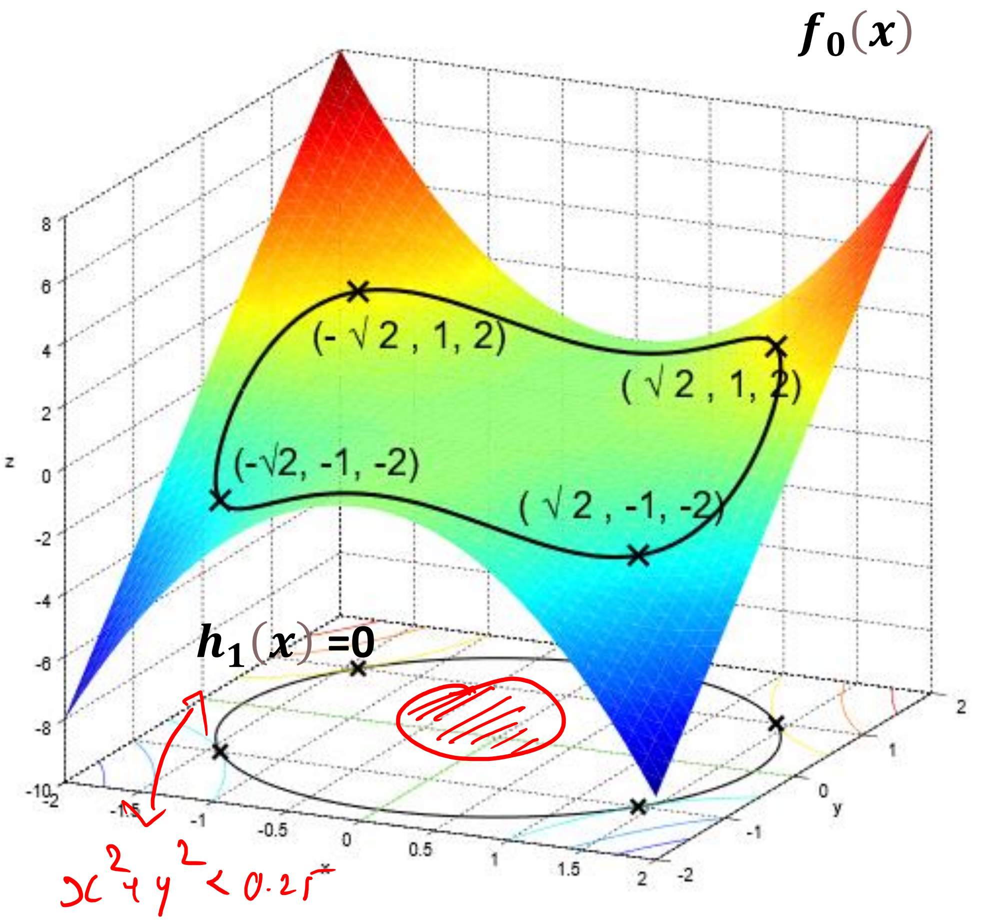
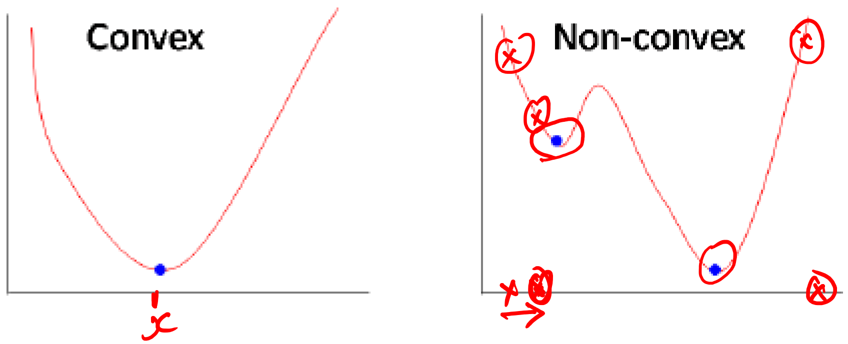


Illustration – 1: $f_0: \mathbb{R}^2 \to \mathbb{R}$

Non - Convex Optimization

graph of loss fn.



Local Optimization Methods:

- Find a point that minimizes $f_0(x)$ among feasible points near it
- Requires initial guess
- Provides no information on distance to global optimum.

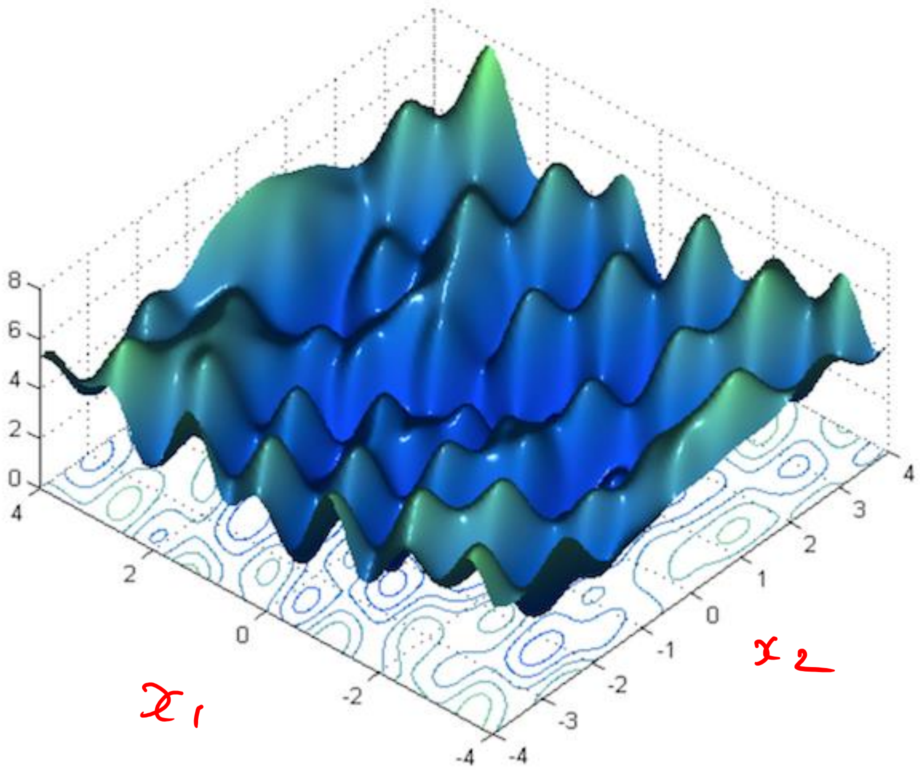


Fig: Loss function ($f: \mathbb{R}^N \to \mathbb{R}$) surface of a Neural Network ($\boldsymbol{N}$: # parameters)

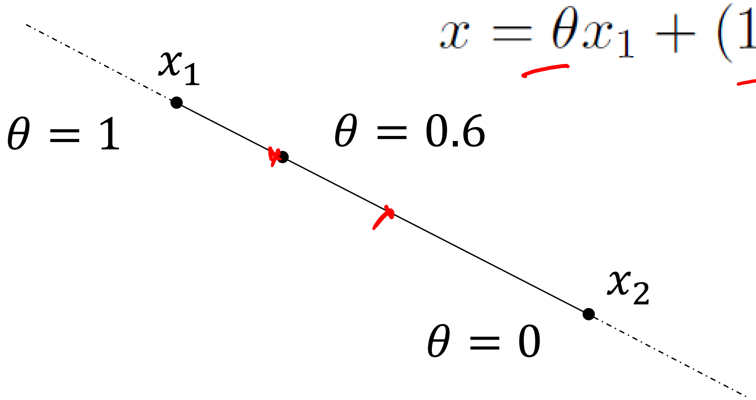
Convex Sets

$\theta_1 \geq 0 \quad \theta_2 \geq 0$

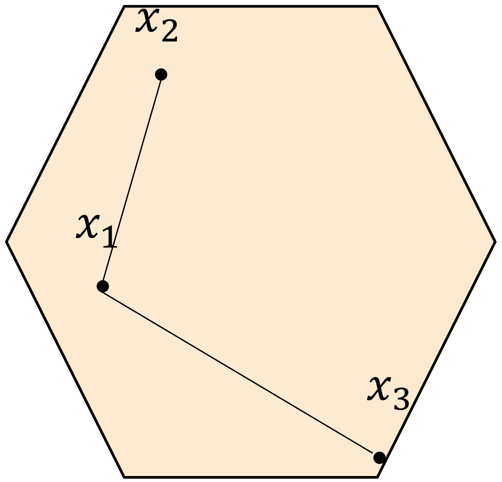
Definition :

Given any 2 points, x_1, x_2 in a set S , and any 2 non-negative numbers , θ_1, θ_2, such that $\theta_1 + \theta_2 = 1$, the affine combination of the points, $\theta_1 x_1 + \theta_2 x_2$, should lie inside the set S.

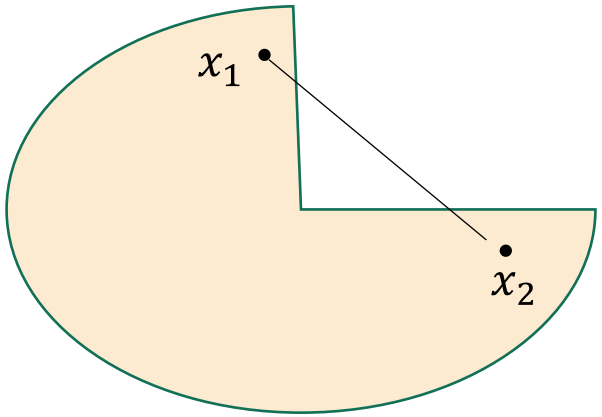
$$x = \theta x_1 + (1 - \theta) x_2$$



Convex Set

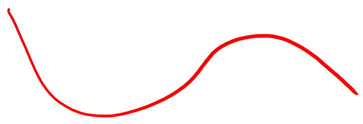


Convex Set



Non- Convex Set

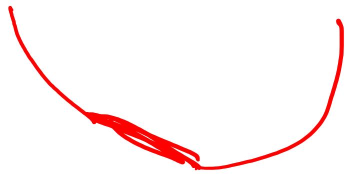
Convex Functions

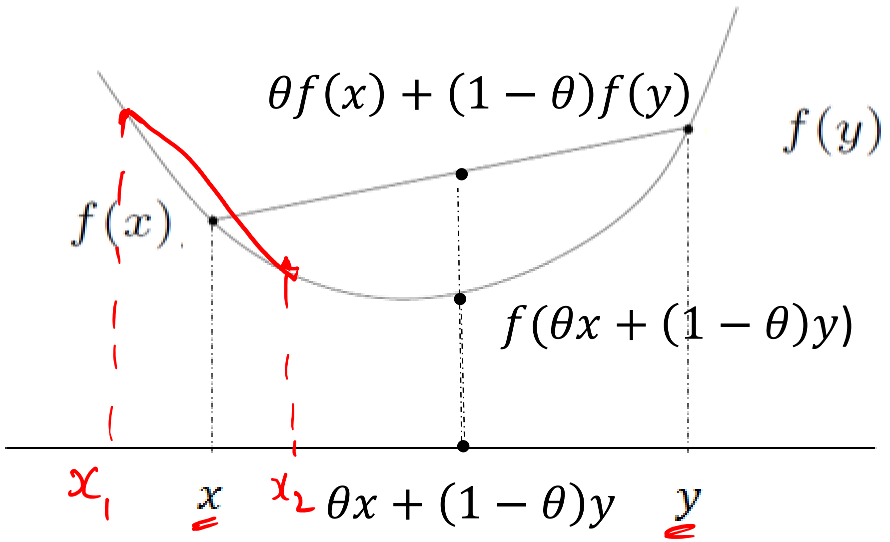


$f : \mathbf{R}^n \to \mathbf{R}$ is convex if $\mathbf{dom}\, f$ is a convex set and

$$\underbrace{f(\theta x + (1-\theta)y)} \leq \theta f(x) + (1-\theta) f(y) \quad \text{for all } x, y \in \mathbf{dom}\, f,\ 0 \leq \theta \leq 1$$

$f : R \to \mathbb{R}$





f is strictly convex if $\mathbf{dom}\, f$ is convex and $f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta) f(y)$

First Order Condition for Convexity

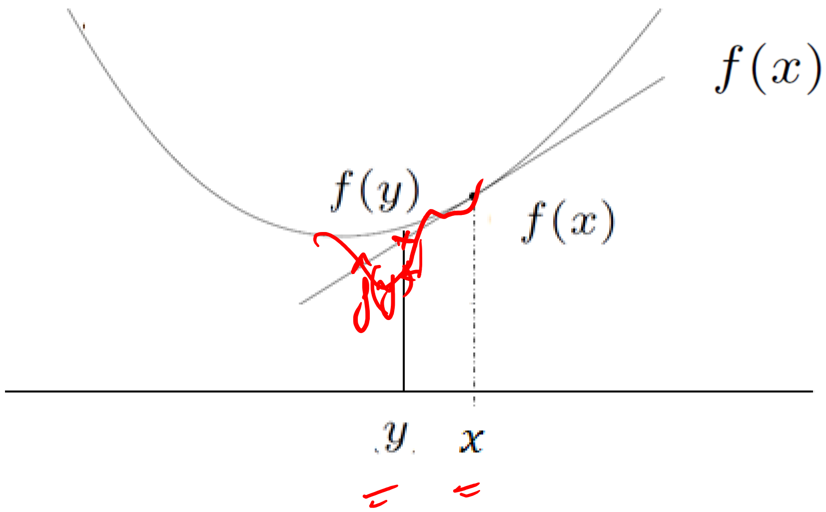
$$f(y) = f(x) + \nabla f^T(x)(y - x)$$

Assume $\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \ldots, \frac{\partial f(x)}{\partial x_n}\right)$ exists at each x ∈ **dom** f

1st-order condition: differentiable f with convex domain is convex if

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad \text{for all } x, y \in \textbf{dom}\, f$$

f



$f(x) + \nabla f(x)^T (y - x)$ → Taylor's first order approximation in variable y

Tangent at x is the global under-estimator of if function $f(x)$ is convex

Second Order Condition for Convexity

$\Delta x = y - x$

Taylor series expansion of a multi-variable scalar function $f(x)$ with $x \in \mathbb{R}^n$:

Hessian Matrix

$x^T A x$

$$f(x + \Delta x) = f(x) + \nabla f^T \Delta x + {}^1/_2\, \Delta x^T\, \nabla^2 f\, \Delta x + \; \ldots.$$

For $f(x)$ with $x \in \mathbb{R}^n$ Hessian matrix is defined as :

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix} \quad \text{and} \quad \nabla^2 f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

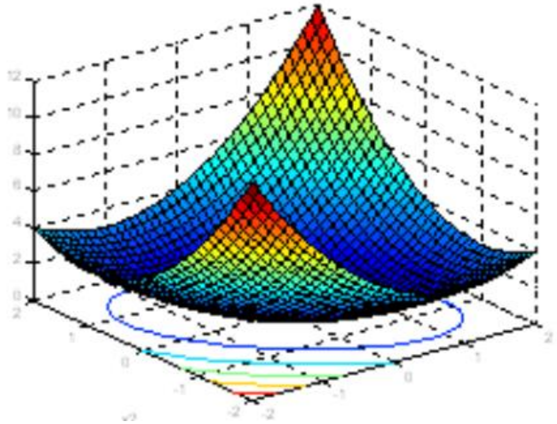
For twice differentiable functions,
$f(x)$ is convex if and only If :

$$\boxed{\nabla^2 f(x) \succeq 0 \quad \text{for all } x \in \mathbf{dom}\, f}$$

Convexity : Examples

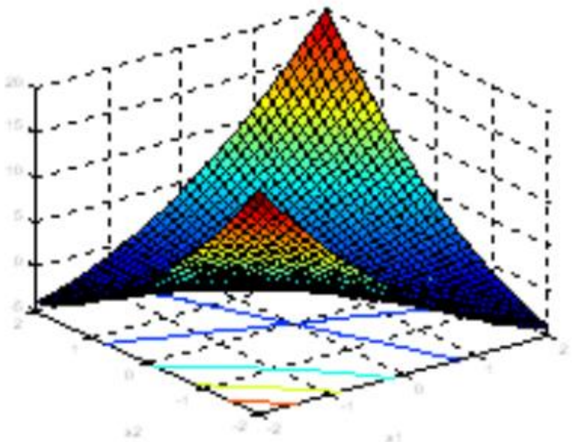
$x^T \nabla^2 f \, x$

$f = x_1^2 + x_1 x_2 + x_2^2$ $\quad \Delta f = \begin{Bmatrix} 2x_1 + x_2 \\ x_1 + 2x_2 \end{Bmatrix}$



$$\nabla^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial f}{\partial x_1 \partial x_2} \\ \frac{\partial f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Eigen Values, $\lambda_{1,2} = 1\,,3$



Change f to $f = x_1^2 + 3x_1 x_2 + x_2^2$ $\quad$ Eigen Values, $\lambda_{1,2} = -1\,,5$

Convexity : Examples

Quadratic Function:

$f : \mathbb{R}^n \rightarrow \mathbb{R}$

$P \in S^{n \times n}$ $q \in \mathbb{R}^n$

$r \in \mathbb{R}$

$$f(x) = (1/2)x^T P x + q^T x + r \qquad (\text{with } P \in \mathbf{S}^n)$$

$$\nabla f(x) = Px + q \qquad \nabla^2 f(x) = P$$

$$\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}^\top$$

$$\frac{\partial \mathbf{x}^\top \mathbf{A}}{\partial \mathbf{x}} = \mathbf{A}$$

$$\frac{\partial \mathbf{x}^\top \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}^\top)\,\mathbf{x}$$

$f(x) = x^T x$

$= x^T I x$

$\|Ax - b\|^2$

$\|e\|^2$

convex if $P \succeq 0$

Poll : Is the function $f(x) = \|x\|^2$ convex ? ($x \in \mathbb{R}^n$)

Convexity : Examples

$x^T A^T \underbrace{Ax}_{y}$

$y^T y$

$$\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}^\top$$

$$\frac{\partial \mathbf{x}^\top \mathbf{A}}{\partial \mathbf{x}} = \mathbf{A}$$

$$\frac{\partial \mathbf{x}^\top \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}^\top)\,\mathbf{x}$$

Least Squares Objective:

$$f(x) = \|Ax - b\|_2^2$$

$$\nabla f(x) = 2A^T(Ax - b) \qquad \nabla^2 f(x) = 2A^T A$$

Convex for any **A,** hence will have global minimum

Operations that preserve Convexity

- **Non-Negative Multiple:** αf is convex if f is convex, $\alpha \geq 0$

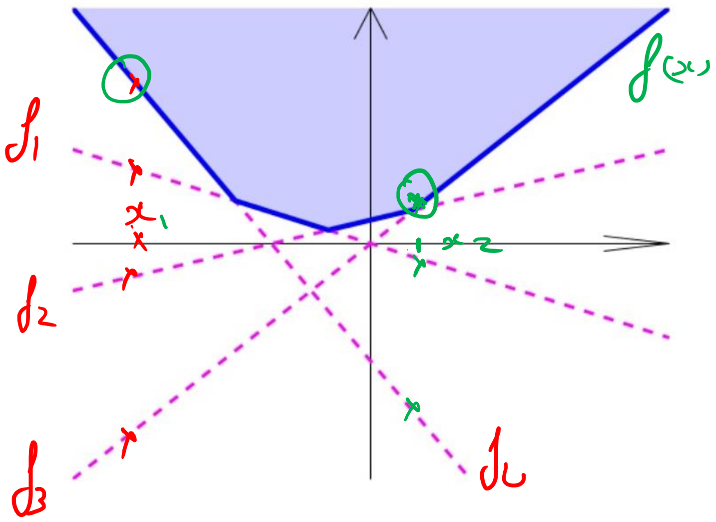
- **Sum:** $f_1 + f_2$ convex if f_1, f_2 convex

- **Pointwise Maximum :** if $f_1, \ldots, f_m$ are convex,

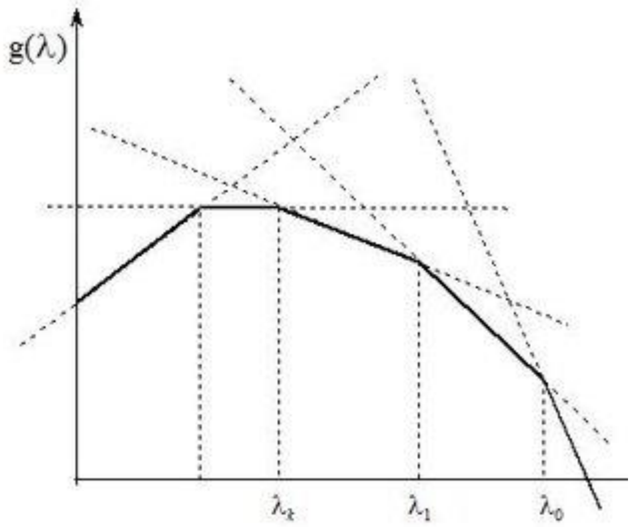
  then $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ is convex

Proof: $f(x) = \max\{f_1(x), f_2(x)\}$

$$\begin{aligned} f(\theta x + (1-\theta)y) &= \max\{f_1(\theta x + (1-\theta)y), f_2(\theta x + (1-\theta)y)\} \\ &\leq \max\{\theta f_1(x) + (1-\theta)f_1(y), \theta f_2(x) + (1-\theta)f_2(y)\} \\ &\leq \theta \max\{f_1(x), f_2(x)\} + (1-\theta)\max\{f_1(y), f_2(y)\} \\ &= \theta f(x) + (1-\theta)f(y) \end{aligned}$$



Pointwise Maximum - Convex

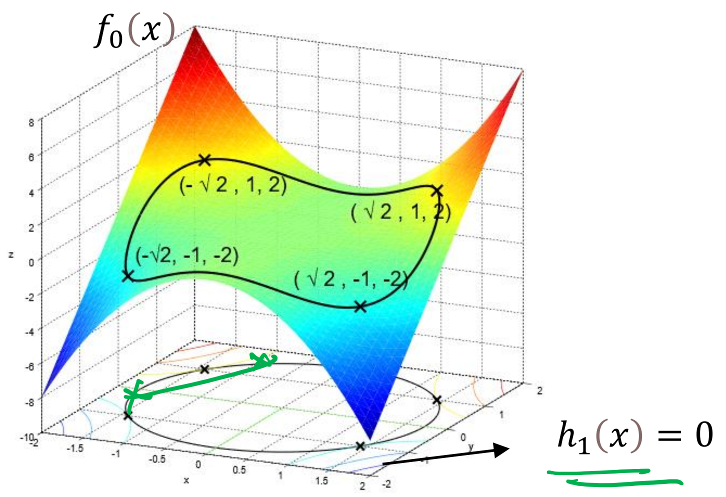


Pointwise Minimum - Concave

Convex Optimization Formulation

General Optimization Formulation:

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le 0, \quad i = 1, \dots, m \\ & h_i(x) = 0, \quad i = 1, \dots, p \end{array}$$



Convex Optimization Formulation:

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le 0, \quad i = 1, \dots, m \\ & a_i^T x = b_i, \quad i = 1, \dots, p \end{array}$$

Poll: Suppose we consider $f_0(x)$ that is convex, will the above formulation convex ? Yes/No

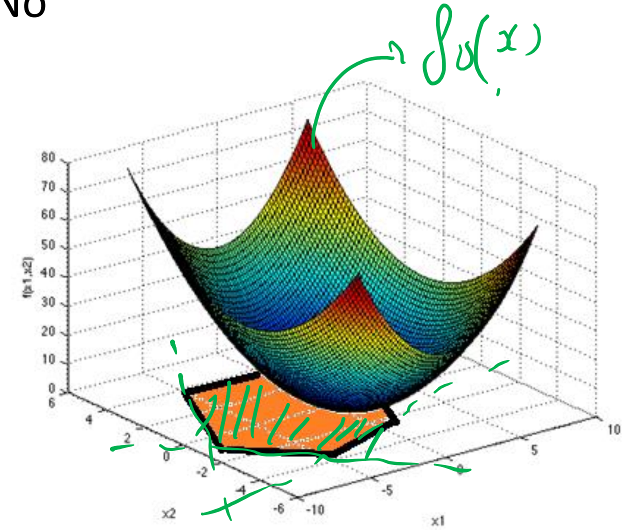
$f_0, f_1, \dots, f_m$ are convex; equality constraints are affine

Feasible region of a convex optimization problem is convex

Alternative way of writing:

$f_i(x) < 0$

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le 0, \quad i = 1, \dots, m \\ & Ax = b \end{array}$$



Convex formulation