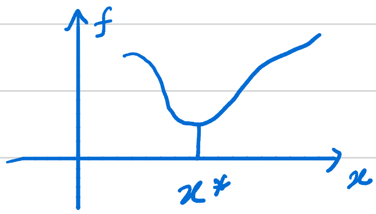
UNCONSTRAINED OPTIMIZATION ($S = \mathbb{R}^n$)

<u>Necessary Conditions for Optimality</u>: Let x^* be a local min. of f, and suppose f is <u>continuously differentiable</u> in a neighborhood around x^*



Consider arbitrary (direction) vector $z \in \mathbb{R}^n$.
For $\alpha > 0$ sufficiently small;

1) $f(x^*) \leq f(x^* + \alpha z)$

2) $g(\alpha) = f(x^* + \alpha z) - f(x^*) \geq 0$

3) $g(\beta)$ is continuously differentiable for $\beta \in [0, \alpha]$

By chain rule,

$$g'(\beta) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x^* + \beta z) . z_i$$

By Mean Value Theorem,

$$g(\alpha) = \underbrace{\cancel{g(0)}}_{=0} + g'(\beta)\alpha \quad \text{for some } \beta \in [0, \alpha]$$

Thus

$$g(\alpha) = \alpha \sum_{i=1}^{n} z_i \frac{\partial f}{\partial x_i}(x^* + \beta z) \geq 0$$

$$\Rightarrow \quad \sum_{i=1}^{n} z_i \frac{\partial f}{\partial x_i}(x^* + \beta z) \geq 0$$

Letting $\alpha \to 0$ and hence $\beta \to 0$, we get

$$\sum_{i=1}^{n} z_i \frac{\partial f}{\partial x_i}(x^*) \geq 0 \quad \text{for all } z \in \mathbb{R}^n$$

$$\sum_{i=1}^{n} z_i \frac{\partial f}{\partial x_i}(x^*) \geq 0 \qquad \forall z \in \mathbb{R}^n$$

Choosing $z = [1\ 0 \cdots 0]^T, [-1\ 0 \cdots 0]^T$

$$\Rightarrow \quad \frac{\partial f}{\partial x_1}(x^*) \geq 0 \quad \text{and} \quad -\frac{\partial f}{\partial x_1}(x^*) \geq 0$$

$$\Rightarrow \quad \frac{\partial f}{\partial x_1}(x^*) = 0$$

Similarly, $\frac{\partial f}{\partial x_i}(x^*) = 0 \quad \forall i$

Thus, $\nabla f(x^*) = \left[\frac{\partial f}{\partial x_1}(x^*) \cdots \frac{\partial f}{\partial x_n}(x^*)\right]^T = 0$

gradient

The same result holds if x^* is a local max.

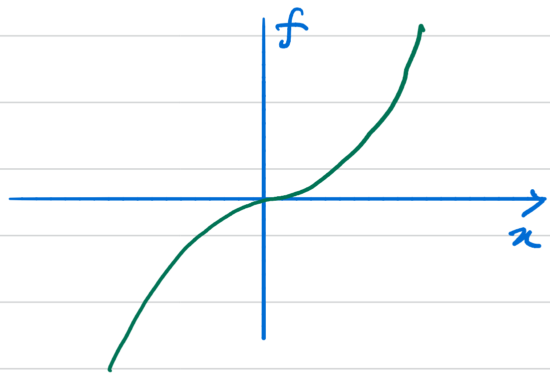
Theorem: If f is continuously differentiable and x^* is a local extremum then $\nabla f(x^*) = 0$

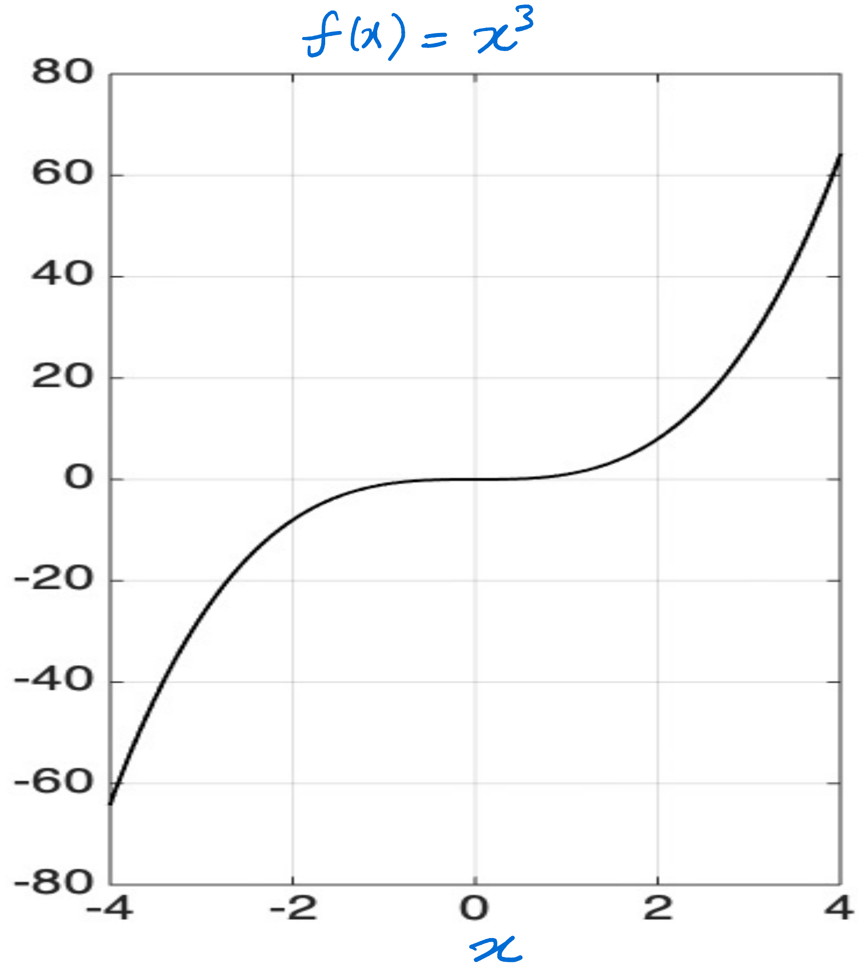
- All points x^* s.t. $\nabla f(x^*) = 0$ are called stationary points
- Thus, all extrema are stationary points
- But not all stationary points have to be extrema

Example: $f(x) = x^3$

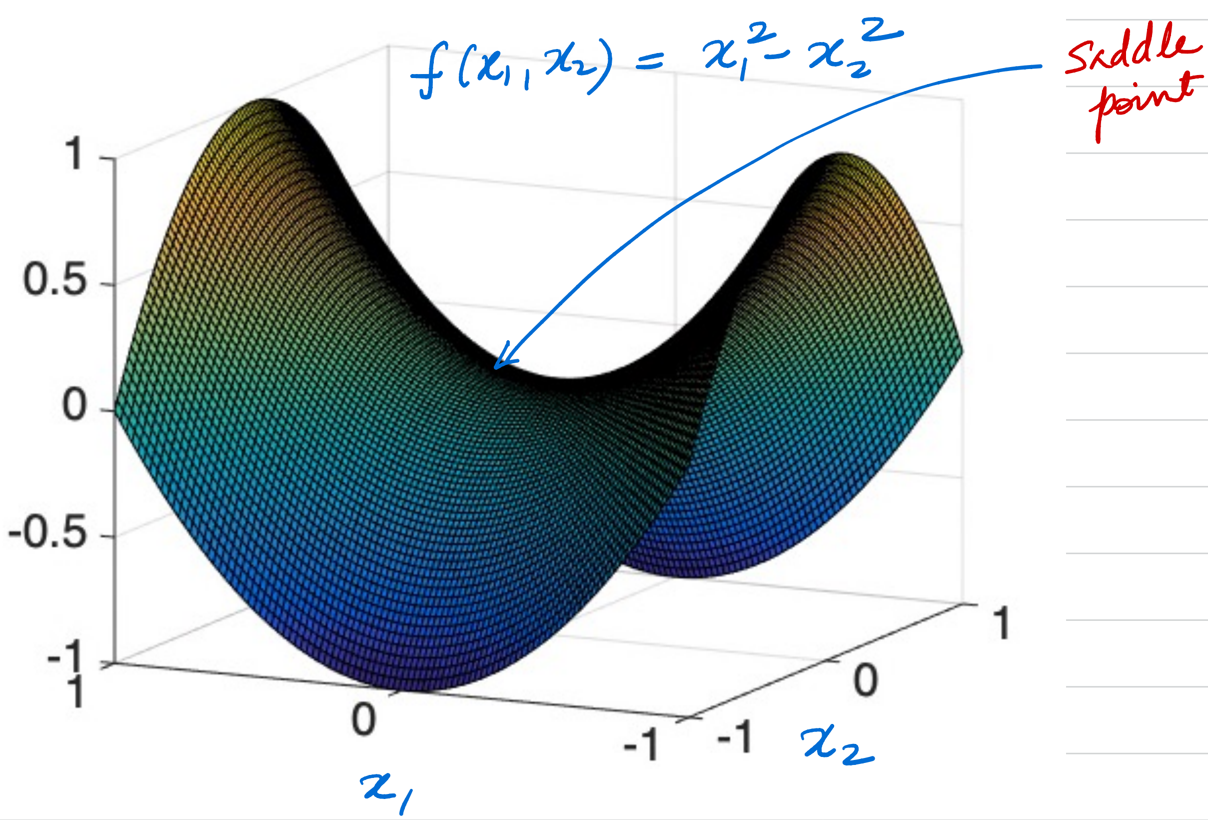
$$\frac{d}{dx} f(x) = 3x^2$$

$\frac{d}{dx} f(0) = 0$ but 0 is not extremum





$f(x) = x^3$
80
60
40
20
0
-20
-40
-60
-80
-4
-2
0
2
4
x



$f(x_1, x_2) = x_1^2 - x_2^2$
Saddle point
1
0.5
0
-0.5
-1
1
0
-1
x_1
1
0
-1
x_2

Second Order Necessary Condition

twice cont. differentiable

Definition The Hessian of f at point x is an $n \times n$ symmetric matrix denoted by $\nabla^2 f(x)$ with:

$$[\nabla^2 f(x)]_{ij} = \frac{\partial^2}{\partial x_i \partial x_j} f(x)$$

Theorem Suppose f is twice continuously differentiable and x^* is local minimum. Then

$$\nabla f(x^*) = 0 \quad \text{and} \quad \nabla^2 f(x^*) \succeq 0$$

Proof We have already established that $\nabla f(x^*) = 0$.

As before, for $z \in \mathbb{R}^n$, let α be small enough so that

$$g(\alpha) = f(x^* + \alpha z) - f(x^*) \geq 0$$

By Taylor series expansion,

$$g(\alpha) = \underbrace{\cancel{g(0)}}_{=0} + \alpha g'(0) + \frac{\alpha^2}{2} g''(0) + o(\alpha^2)$$

$$g'(\alpha) = \sum_{i=1}^{n} z_i \frac{\partial}{\partial x_i} f(x^* + \alpha z) \qquad = \nabla f(x^* + \alpha z)^T z$$

$$g''(\alpha) = \sum_{i=1}^{n} \sum_{j=1}^{n} z_i z_j \frac{\partial^2}{\partial x_i \partial x_j} f(x^* + \alpha z)$$

$$= z^T \nabla^2 f(x^* + \alpha z) z$$

$$g'(0) = \nabla f(x^*)^T z \quad \text{and} \quad g''(0) = z^T \nabla^2 f(x^*) z$$

$$0 \leq g(\alpha) = \alpha \underbrace{\cancel{\nabla f(x^*)^T z}}_{=0 \text{ as shown before}} + \frac{\alpha^2}{2} z^T \nabla^2 f(x^*) z + o(\alpha^2)$$

Thus for all $z \in \mathbb{R}^n$, $\frac{\alpha^2}{2} z^T \nabla^2 f(x^*) z + o(\alpha^2) \geq 0$

Dividing by α^2 and taking $\alpha \to 0$, we get

$$z^T \nabla^2 f(x^*) z \geq 0 \quad \forall z \in \mathbb{R}^n$$

$$\Rightarrow \quad \nabla^2 f(x^*) \succeq 0.$$

Sufficient Conditions for Optimality

Theorem Suppose f is twice continuously diff. in a neighborhood of x^* and (i) $\nabla f(x^*) = 0$, and (ii) $\nabla^2 f(x^*) \succ 0$. Then x^* is (strict) local min.

Proof Consider $z \in \mathbb{R}^n$, $\alpha > 0$ and let

$$g(\alpha) = f(x^* + \alpha z) - f(x^*)$$

$$g(\alpha) = \underbrace{\cancel{g(0)}}_{=0} + \alpha \underbrace{\cancel{g'(0)}}_{=0} + \frac{\alpha^2}{2} g''(0) + o(\alpha^2)$$

$$= \frac{\alpha^2}{2} z^T \nabla^2 f(x^*) z + o(\alpha^2)$$

$$= \frac{\alpha^2}{2} \Big(\underbrace{z^T \nabla^2 f(x^*) z}_{>0 \ \forall z \neq 0 \text{ by PD}} + 2 \underbrace{\frac{o(\alpha^2)}{\alpha^2}}_{\to 0 \text{ as } \alpha^2 \to 0} \Big)$$

$\Rightarrow g(\alpha) > 0$ for α suff. small for all $z \neq 0$

$\Rightarrow$ x^* is strict local min.

Using Optimality Conditions to find Minimum

- Find all points satisfying necessary condition $\nabla f(x) = 0$, i.e. all stationary points.
- Filter out points that don't satisfy $\nabla^2 f(x) \succeq 0$
- Points with $\nabla^2 f(x) \succ 0$ are strict local min.
- Among all points with $\nabla^2 f(x) \succeq 0$, declare as global min. one with smallest value of f, assuming that global min. exists.

Example $f(x) = 2x^2 - x^4$

$$\nabla f(x) = \frac{d}{dx} f(x) = 4x - 4x^3 = 0$$

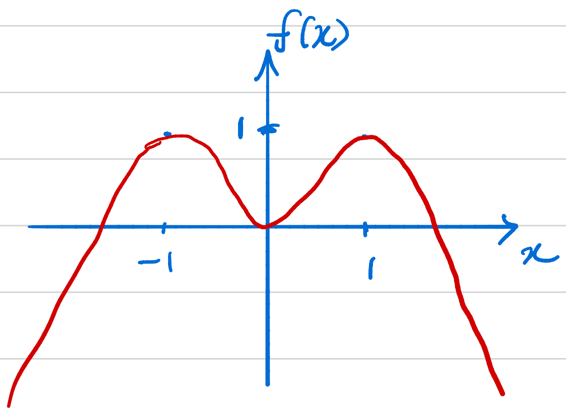
$\Rightarrow$ $x = 0$, $x = 1$, $x = -1$ are stationary points.

$$\nabla^2 f(x) = 4 - 12x^2 = \begin{cases} -8 & \text{if } x = 1, \text{ or } -1 \\ 4 & \text{if } x = 0 \end{cases}$$

$\Rightarrow$ $x = 0$ is the only local min. and it is strict

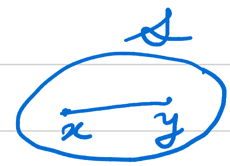
But $-f(x) \to \infty$ as $|x| \to \infty$ $\Rightarrow$ no global min., but global max exists.

Corollary to WT



- $x = 1$, and $x = -1$ are strict local max.
- $f(1) = f(-1) = 1$ and So they are both global max.

Convex Sets and Functions



Definition $\mathcal{S} \subseteq \mathbb{R}^n$ is a convex set if

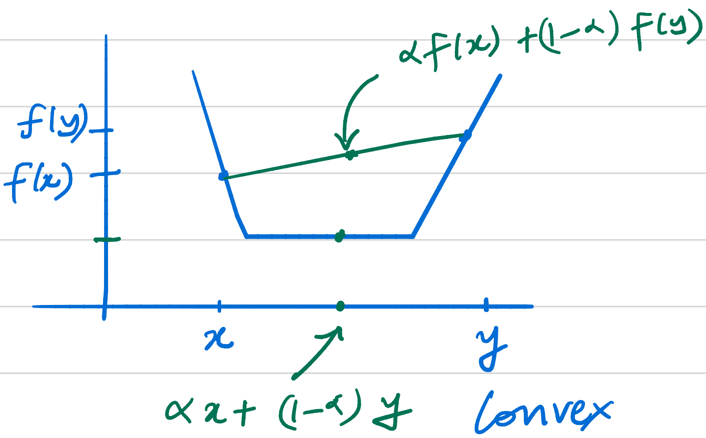
$$\forall\ x, y \in \mathcal{S},\ \alpha \in [0,1], \quad \alpha x + (1-\alpha) y \in \mathcal{S}$$

Definition $f : \mathcal{S} \to \mathbb{R}$, where $\mathcal{S}$ is a convex set is called a convex function if

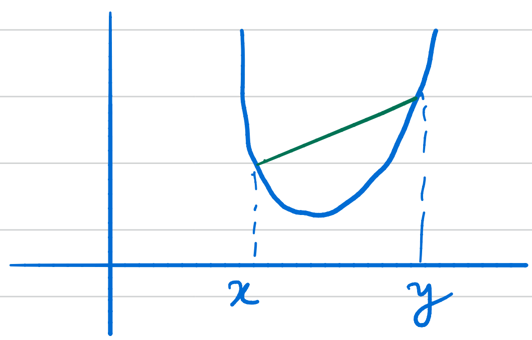
$$f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha) f(y)$$

$$\forall \alpha \in [0,1],\ x, y \in \mathcal{S}$$

f is said to be strictly convex if inequality is strict for all $\alpha \in (0,1)$, $x \neq y$.

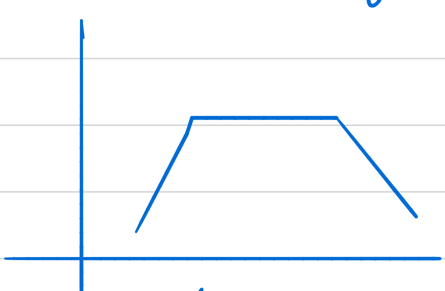


Convex

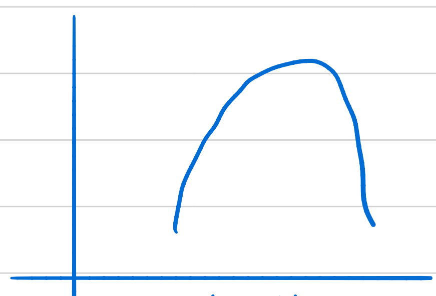


strictly convex

f is concave (strictly concave) if $-f$ is convex (strictly convex)



concave



strictly concave

Some Results Regarding Convex Sets and Functions

1. Suppose f is a convex function (over $\mathbb{R}^n$) and define the set

$$\mathcal{S} = \{ x \in \mathbb{R}^n : f(x) \leq a \}, \quad a \in \mathbb{R}$$

Then $\mathcal{S}$ is a convex set

Proof Let $x, y \in \mathcal{S}$, i.e., $f(x) \leq a$, $f(y) \leq a$

Then for all $\alpha \in [0,1]$,

$$f(\alpha x + (1-\alpha) y) \underset{\text{convexity of } f}{\leq} \alpha f(x) + (1-\alpha) f(y)$$

$$\leq \alpha a + (1-\alpha) a = a$$

$\Rightarrow \alpha x + (1-\alpha) y \in \mathcal{S} \Rightarrow \mathcal{S}$ is convex

2) If $f_1, f_2, \ldots, f_K$ are convex functions over convex set $\mathcal{S}$,

(i) $f_{sum}(x) = \sum_{i=1}^{K} f_i(x)$ is convex over $\mathcal{S}$

(ii) $f_{max}(x) = \max_{i=1,\ldots,K} f_i(x)$ is convex over $\mathcal{S}$

Proof: (i) is easy to see.

(ii) For $x, y \in \mathcal{S}$, $\alpha \in [0,1]$,

$$f_{max}(\alpha x + (1-\alpha) y) = \max_{i=1,\ldots,K} f_i(\alpha x + (1-\alpha) y)$$

$$\leq \max_{i=1,\ldots,K} [\alpha f_i(x) + (1-\alpha) f_i(y)]$$

$$\leq \max_{i=1,\ldots,K} \alpha f_i(x) + \max_{i=1,\ldots,K} (1-\alpha) f_i(y)$$

$$= \alpha f_{max}(x) + (1-\alpha) f_{max}(y)$$