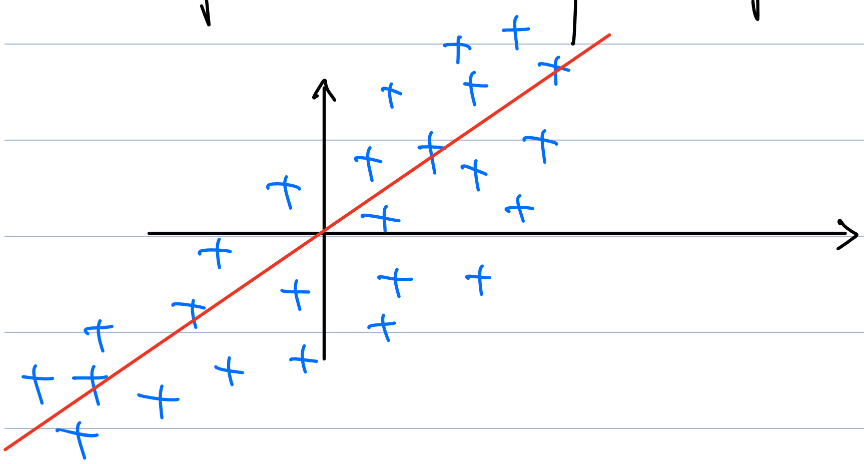
Fundamentals of Unconstrained Optimization

Unconstrained Problem:

$$\min_{x \in \mathbb{R}^n} f(x)$$

Example: Least Squares Problem



Suppose there are m data pts: $(a_i, b_i) \in \mathbb{R}^{n+1}$

Find an affine function $f_x(a) = x^T a$ to fit these pts

For the i-th data pt (a_i, b_i), there is an error by using $f_x(a)$ to fit it. Denote the error function as:

$$d(f_x(a_i), b_i)$$

① If $d(b, c) = (b-c)^2$

then the total error is:

$$\sum_{i=1}^{m} (f_x(a_i) - b_i)^2 = \sum_{i=1}^{m} (x^T a_i - b_i)^2$$

Let $A = \begin{bmatrix} — & a_1^T & — \\ — & a_2^T & — \\ & \vdots & \\ — & a_m^T & — \end{bmatrix}$, $\vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$

Then

$$\sum_{i=1}^{m} (x^T a_i - b_i)^2 = \sum_{i=1}^{m} (a_i^T x - b_i)^2$$

$$= \|Ax - \vec{b}\|^2$$

$(Ax - \vec{b})_i = a_i^T x - b_i$

Therefore, to find the best f_x or x, we can

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|Ax - \vec{b}\|^2$$

② $d(b, c) = |b - c|$, $\forall\, b, c \in \mathbb{R}$

Then the total error is:

$$\sum_{i=1}^{m} d(f_x(a_i), b_i) = \sum_{i=1}^{m} |f_x(a_i) - b_i|$$

$$= \sum_{i=1}^{m} |a_i^T x - b_i|$$

$$= \|Ax - b\|_1$$

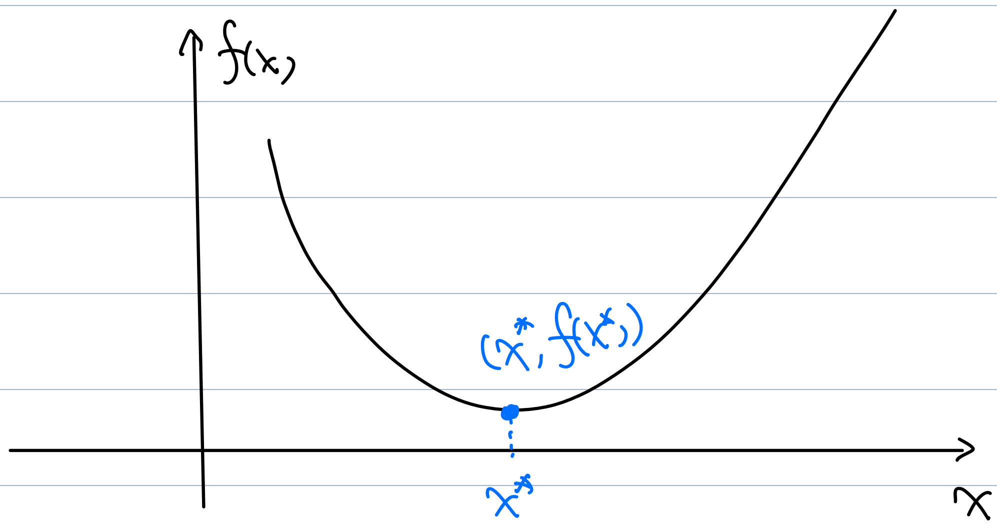
To find the best x, we can solve:

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_1$$

Def (global minimizer): $x^* \in \mathbb{R}^n$ is called a global minimizer of a function $f: \mathbb{R}^n \to \mathbb{R}$ if:

$$f(x^*) \leq f(x), \quad \forall x \in \mathbb{R}^n \quad \text{- - -} \quad (1)$$

Example:



Remark: if (1) is changed to

$$f(x^*) < f(x), \quad \forall x \in \mathbb{R}^n \text{ and } x \neq x^*$$

then x^* is called a strict global minimizer.

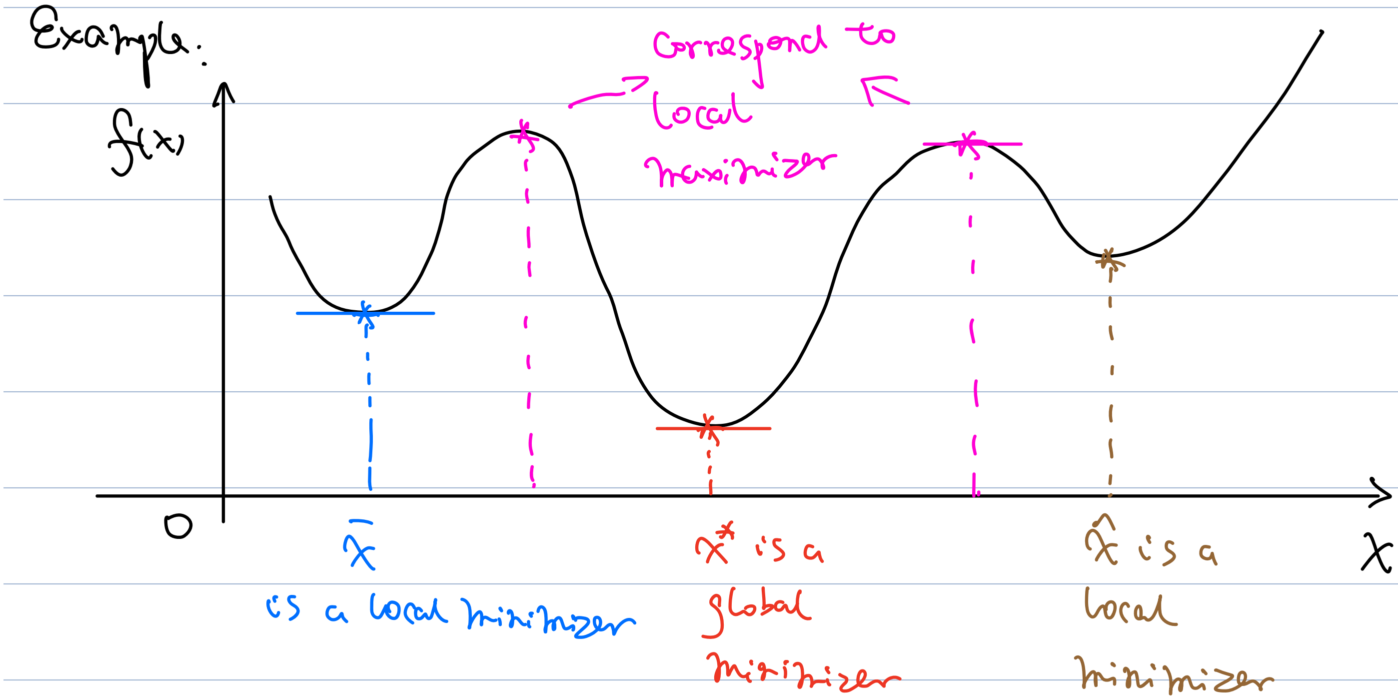
Def (global maximizer): $x^* \in \mathbb{R}^n$ is called a global maximizer of a function $f: \mathbb{R}^n \to \mathbb{R}$ if:

$$f(x^*) \geq f(x), \quad \forall x \in \mathbb{R}^n$$

Def (Local minimizer): $x^* \in \mathbb{R}^n$ is called a local minimizer of a function $f: \mathbb{R}^n \to \mathbb{R}$ if there exists $\delta > 0$, such that

$$f(x^*) \le f(x), \quad \forall x \in B_\delta(x^*) = \{x \in \mathbb{R}^n : \|x - x^*\| \le \delta\}$$

Example:



Remark: a global minimizer must be a local minimizer.

Def (Local maximizer): $x^* \in \mathbb{R}^n$ is called a local maximizer of a function $f: \mathbb{R}^n \to \mathbb{R}$ if there exists $\delta > 0$, such that

$$f(x^*) \ge f(x), \quad \forall x \in B_\delta(x^*) = \{x \in \mathbb{R}^n : \|x - x^*\| \le \delta\}$$

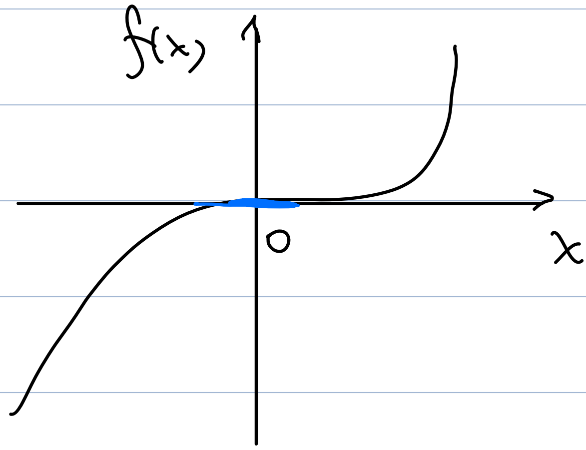
Def (Stationary point): a pt $\bar{x} \in \mathbb{R}^n$ is called a stationary pt of a function $f: \mathbb{R}^n \to \mathbb{R}$ if

$$\nabla f(\bar{x}) = \vec{0}$$

Remark. any local minimizer / maximizer is a stationary pt.

Def (Saddle pt): a pt $\bar{x} \in \mathbb{R}^n$ is called a saddle pt of a function $f: \mathbb{R}^n \to \mathbb{R}$ if $\bar{x}$ is a stationary pt but is not a local minimizer or a local maximizer.

Example:



Theorem (first-order necessary optimality condition):
Assume $f: \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function. If x^* is a local minimizer of f, then x^* must be a stationary pt of f, i.e. $\nabla f(x^*) = \vec{0}$.

Example: Let $f(x,y) = x^2 - 2xy + y^2 - 2$, for $x, y \in \mathbb{R}$

Q1: Is $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ a local minimizer?

$$\nabla f(x,y) = \begin{bmatrix} 2x-2y \\ -2x+2y \end{bmatrix}, \quad \text{so } \nabla f(1,0) = \begin{bmatrix} 2 \\ -2 \end{bmatrix} \neq \vec{0}$$

so $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ is not a local minimizer

Q2: Is $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ a local minimizer?

$$\nabla f(1,1) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$f(x,y) = x^2 - 2xy + y^2 - 2 = (x-y)^2 - 2$$

So $f(1,1) = 0 - 2 \leq f(x,y), \ \forall (x,y)$

Def (Convexity): A set $X \subseteq \mathbb{R}^n$ is convex if

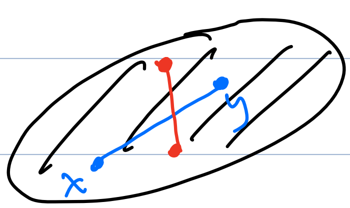
$$\lambda x + (1-\lambda) y \in X, \quad \forall x \in X, \forall y \in X, \forall \lambda \in [0,1]$$

A function $f: X \to \mathbb{R}$ is convex if X is convex

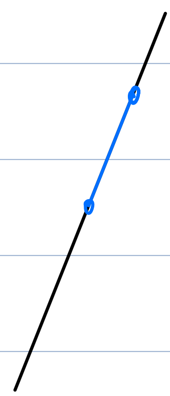
and $$f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y)$$

$$\forall x \in X, \forall y \in X, \forall \lambda \in [0,1]$$

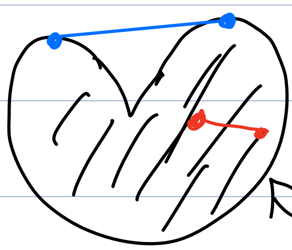
Example: convex or nonconvex sets.

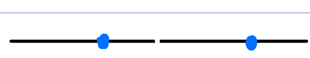


↑ is a convex set



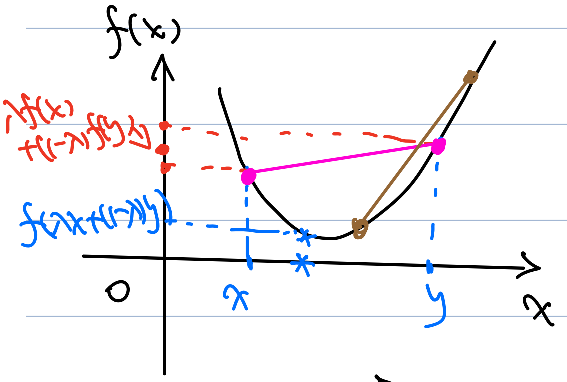
$\mathbb{R}^n$ is convex

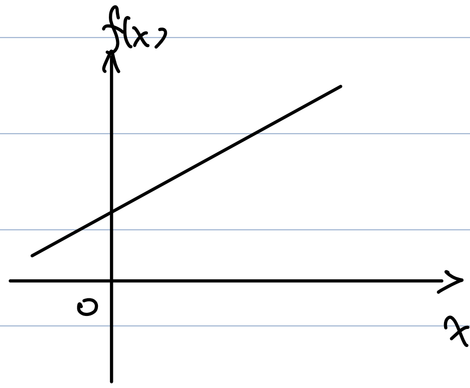


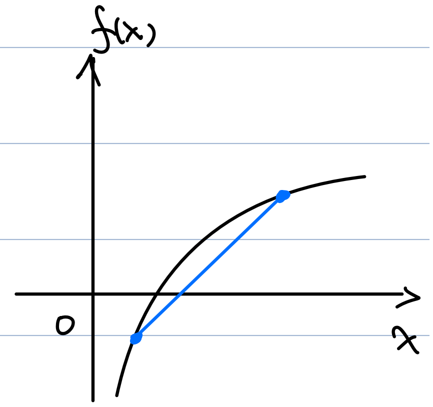


nonconvex

Example: convex or nonconvex functions

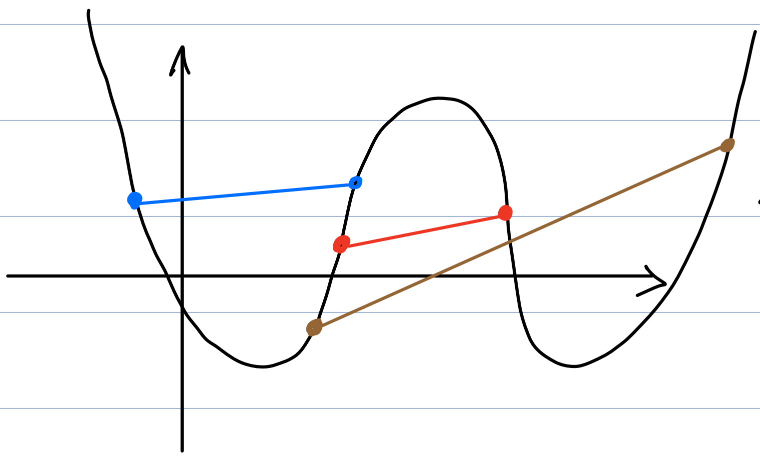






convex function

nonconvex but concave



nonconvex

non concave

Examples of convex functions:

(1) $f(x) = e^x, \quad \forall x \in \mathbb{R}$,

(2) $f(x) = \|Ax-b\|^2, \quad \forall x \in \mathbb{R}^n$

(3) $f(x) = -\ln x, \quad \forall x > 0$

(4): $f(x) = a^T x - b, \quad \forall x \in \mathbb{R}^n$, where $a \in \mathbb{R}^n$, $b \in \mathbb{R}$

(5): If $f : \mathbb{R}^n \to \mathbb{R}$ is convex,

then $\alpha \cdot f : \mathbb{R}^n \to \mathbb{R}$ is convex if $\alpha \geq 0$

(6): If $f, g: \mathbb{R}^n \to \mathbb{R}$ are convex,

then $f+g$ is convex

(7): If $f: \mathbb{R}^m \to \mathbb{R}$ is convex

then $f(Ax+b)$ is convex, where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$

Theorem: Let $f: \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function

① If f is convex, then $\nabla^2 f(x)$ is positive semidefinite for all x

② If $\nabla^2 f(x)$ is positive semidefinite for all x, then f is convex.

Theorem (Sufficient optimality condition for a convex function)

If $f: \mathbb{R}^n \to \mathbb{R}$ is a convex function, then any stationary pt $\bar{x}$ of f is a global minimizer of f.

Theorem (Second-order necessary optimality condition): assume $f: \mathbb{R}^n \to \mathbb{R}$ to be twice differentiable. If x^* is a local minimizer of f, then $\nabla f(x^*) = \vec{0}$, $\nabla^2 f(x^*) \succeq 0$

Example: $f(x, y) = \cos(x+y)$

Is $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ a local minimizer of f? X

Sol: $\nabla f(x,y) = \begin{bmatrix} -\sin(x+y) \\ -\sin(x+y) \end{bmatrix}$, so $\nabla f(0,0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$$\nabla^2 f(x,y) = \begin{bmatrix} -\cos(x+y) & -\cos(x+y) \\ -\cos(x+y) & -\cos(x+y) \end{bmatrix}, \text{ so } \nabla^2 f(0,0) = \begin{bmatrix} -1 & -1 \\ -1 & -1 \end{bmatrix}$$

$$= -\begin{bmatrix} 1 \\ 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix}^T$$

Remark: even if $\nabla f(x^*) = \vec{0}$, $\nabla^2 f(x^*) \succeq 0$

x^* may not be a local minimizer of f

e.g. $f(x) = x^3$, $f'(0) = 0$, $f''(0) = 0$

Theorem (second-order sufficient condition). Let $f: \mathbb{R}^n \to \mathbb{R}$ be twice differentiable. If $\nabla f(x^*) = \vec{0}$, and $\nabla^2 f(x^*) \succ 0$, then x^* is a local minimizer.

Example: Let $f(x) = \dfrac{1+(2-x)^2}{1+x^2}$, $\forall x \in \mathbb{R}$

Find all local minimizer(s) of f.

Sol: solve $f'(x) = 0$

Alternatively, note $f(x) = \dfrac{1+4-4x+x^2}{1+x^2} = 1 + 4\dfrac{1-x}{1+x^2}$

Let $g(x) = \frac{1-x}{1+x^2}$. Local minimizers of g are the same as those of f.

Solve $g'(x) = 0$

We have $g'(x) = \frac{-1\cdot(1+x^2) - 2x\cdot(1-x)}{(1+x^2)^2} = \frac{x^2-2x-1}{(1+x^2)^2}$

So $g'(x) = 0 \iff -(1+x^2) - 2x\cdot(1-x) = 0$

$= (x-(1+\sqrt{2}))(x-(1-\sqrt{2}))$

$\Rightarrow x = \underbrace{1+\sqrt{2}}_{x_1}$ or $\underbrace{1-\sqrt{2}}_{x_2}$

Check: $g''(x)$ at $1 \pm \sqrt{2}$

$$g''(x) = \frac{(2x-2)(1+x^2)^2 - (x^2-2x-1)(4x(1+x^2))}{(1+x^2)^4}$$

$$= \frac{(2x-2)(1+x^2) - (x^2-2x-1)(4x)}{(1+x^2)^3}$$

So $g''(x_1) = \frac{(2x_1-2)(1+x_1^2)}{(1+x_1^2)^3} = \frac{2x_1-2}{(1+x_1^2)^2} > 0$

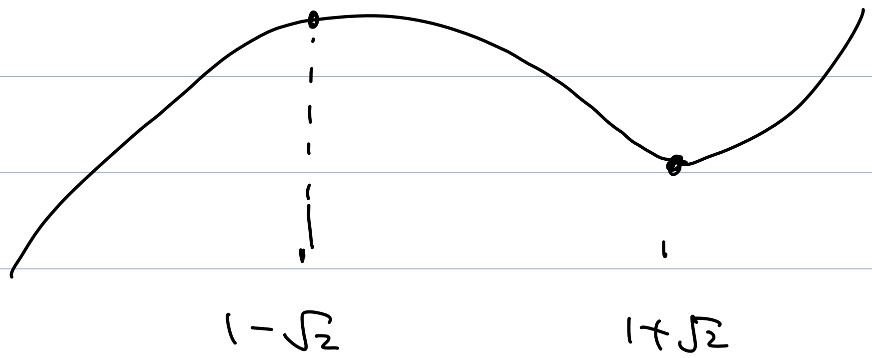
$x_1 = 1+\sqrt{2}$

So x_1 is a local minimizer

$g''(x_2) = \frac{2x_2-2}{(1+x_2^2)^2} < 0$

$x_2 = 1-\sqrt{2}$

So x_2 is a local maximizer.



Corollary: Let f be twice differentiable. If $\nabla f(x^*) = \vec{0}$, but $\nabla^2 f(x^*)$ is indefinite, then x^* is a saddle pt.

$\nabla^2 f(x^*)$ has positive eigenvalues and negative eigenvalues