Part VI: Optimization

VI.1: Minimum Problems:
Convexity and Newton's Method

VI.2: Lagrange Multipliers
= Derivatives of the Cost

VI.3: Linear Programming, Game Theory,
and Duality

VI.4: Gradient Descent Toward the Minimum

VI.5: Stochastic Gradient Descent and ADAM

The Expression "argmin"

Consider $f: \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = (x-1)^2.$$

Then $$\min_{x \in \mathbb{R}} f(x) = 0.$$

That is, the minimum value of $f(x)$ is 0.

Any $\bar{x} \in \mathbb{R}$ such that $f(\bar{x}) = \min_{x \in \mathbb{R}} f(x)$ is called an <u>optimal solution</u>.

The <u>set of optimal solutions</u> is denoted $\operatorname{argmin}_{x \in \mathbb{R}} f(x)$.

Since $\bar{x} = 1$ is the only optimal solution in this example, we have

$$\operatorname*{argmin}_{x \in \mathbb{R}} f(x) = \{1\}.$$

<u>Multivariable Calculus</u>

If $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable, then

$$\boxed{f(x + \Delta x) \approx f(x) + \nabla f(x)^T \Delta x + \frac{1}{2} \Delta x^T \nabla^2 f(x) \Delta x}$$

where $\nabla f(x)$ is the <u>gradient vector</u> of f at x and $\nabla^2 f(x)$ is the <u>Hessian matrix</u> of f at x.

If $f: \mathbb{R}^n \to \mathbb{R}^m$ is differentiable, then

$$\boxed{f(x+\Delta x) \approx f(x) + J(x)\Delta x}$$

where $J(x)$ is the <u>Jacobian matrix</u> of f at x.

Here, $f(x)$ is a vector,

$$f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix},$$

and $J(x)$ is the matrix of all first order partial derivatives,

$$J(x) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x) & \cdots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \cdots & \frac{\partial f_m}{\partial x_n}(x) \end{bmatrix}.$$

For example, let $f: \mathbb{R}^2 \to \mathbb{R}^3$ be given by

$$f(x) = \begin{bmatrix} x_1 x_2 \\ \sin x_1 + \cos x_2 \\ x_1^2 e^{x_2} \end{bmatrix}.$$

Then

$$J(x) = \begin{bmatrix} x_2 & x_1 \\ \cos x_1 & -\sin x_2 \\ 2x_1 e^{x_2} & x_1^2 e^{x_2} \end{bmatrix}.$$

VI.1 Minimum Problems: Convexity and Newton's Method

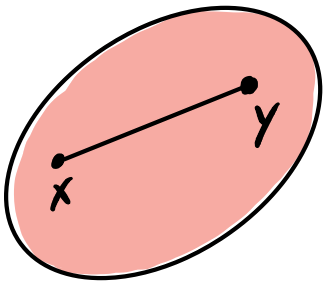
Let $f: \mathbb{R}^n \to \mathbb{R}$ and $K \subseteq \mathbb{R}^n$.

The constrained optimization problem

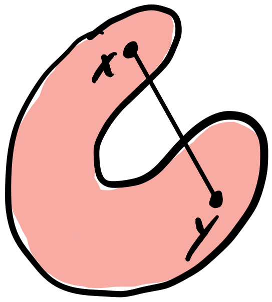
$$\boxed{\min f(x) \text{ subject to } x \in K}$$

is convex if f is a convex function and K is a convex set.

A set $K \subseteq \mathbb{R}^n$ is convex if the line segment between any two points $x, y \in K$ lies entirely in K.



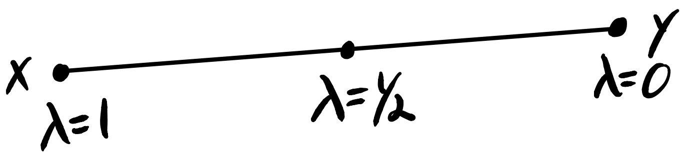
convex



not convex

The line segment between x and y can be written as

$$\{\lambda x + (1-\lambda) y \ : \ \lambda \in [0,1]\}$$



Thus, a set $K \subseteq \mathbb{R}^n$ is <u>convex</u> if $\lambda x + (1-\lambda) y \in K$ for all $x, y \in K$ and $\lambda \in [0,1]$.

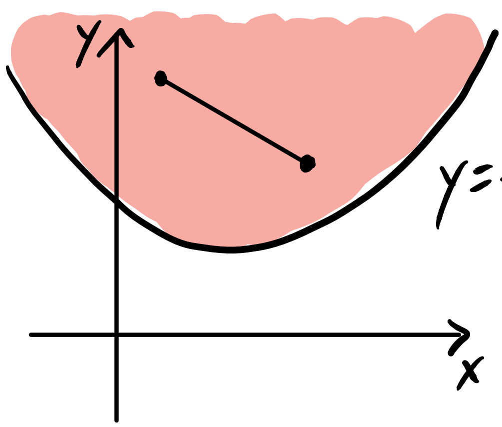
<u>Examples</u>:

(1) $K = \{x \in \mathbb{R}^n : Ax = b\}$.

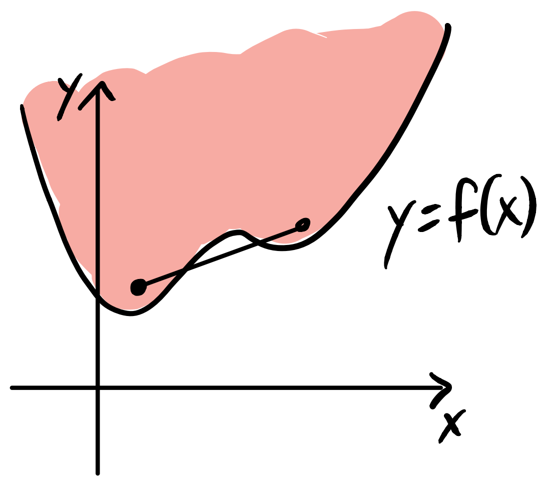
(2) $K = \{x \in \mathbb{R}^n : a^T x \leq b\}$.

(3) $K = K_1 \cap K_2$, K_1 and K_2 convex.

A function $f: \mathbb{R}^n \to \mathbb{R}$ is <u>convex</u> if the set of points on or above the graph of f is a convex set.



f convex



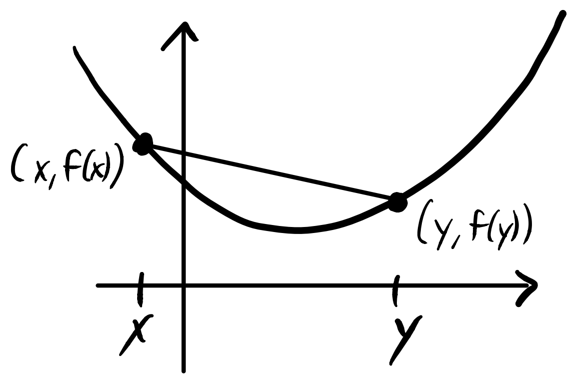
f <u>not</u> convex

The *epigraph* of $f:\mathbb{R}^n \to \mathbb{R}$ is

$$\text{epi}(f) = \{(x,y) \in \mathbb{R}^n \times \mathbb{R} : f(x) \le y\}.$$

Thus, f is a convex function *iff* $\text{epi}(f)$ is a convex set.

Also, a function f is convex if the line segment between $(x, f(x))$ and $(y, f(y))$ lies on or above the graph of f.



That line segment is

$$\{(\lambda x + (1-\lambda) y,\ \lambda f(x) + (1-\lambda) f(y)) : \lambda \in [0,1]\}.$$

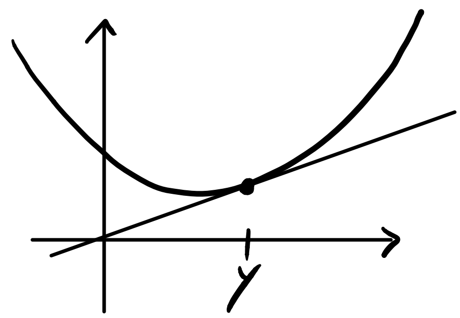
Thus, a function $f:\mathbb{R}^n \to \mathbb{R}$ is *convex* if $f(\lambda x + (1-\lambda) y) \le \lambda f(x) + (1-\lambda) f(y)$ for all $x, y \in \mathbb{R}^n$ and $\lambda \in [0,1]$.

The function is <u>strictly convex</u> if

$$f(\lambda x + (1-\lambda) y) < \lambda f(x) + (1-\lambda) f(y)$$

for all $x, y \in \mathbb{R}^n$, with $x \neq y$, and all $\lambda \in (0,1)$.

A differentiable convex function must be on or above all of its tangent lines.



Thus, $f(x) \geq f(y) + \nabla f(y)^T (x-y)$ for all $x, y \in \mathbb{R}^n$.

A twice differentiable function is convex <u>iff</u> its Hessian $\nabla^2 f(x)$ is positive semidefinite for all x.

If $\nabla^2 f(x)$ is always positive definite, then f is strictly convex.

Examples:

(1) $f(x) = c^T x$

(2) $f(x) = \frac{1}{2} x^T S x$, S sym. pos. def.

(3) $f(x) = x_1^2 + 2x_1 x_2 + x_2^2$

(4) $f(x) = x_1^2 - x_2^2$

(5) $f(x) = \|x\|$ (any vector norm).