

Part III: Optimization

III.1: Minimum Problems:

Convexity and Newton's Method

III.2: Lagrange Multipliers

= Derivatives of the Cost

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

III.4: Gradient Descent Toward the Minimum

III.5: Stochastic Gradient Descent and ADAM

The Expression "argmin"

Consider $f: \mathbb{R} \rightarrow \mathbb{R}$ 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 optimal solution.

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

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

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

Multivariable Calculus

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

where $\nabla f(x) + \Delta x$ is the gradient vector of f
at x and $\nabla^2 f(x)$ is the Hessian matrix
of f at x .

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

$$f(x + \Delta x) \approx f$$

where $J(x)$ is the Jacobian matrix of f at x .

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

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

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

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

For example, let $f: \mathbb{R}^2 \rightarrow \mathbb{R}^3$ 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. I] Minimum Problems:

Convexity and Newton's Method

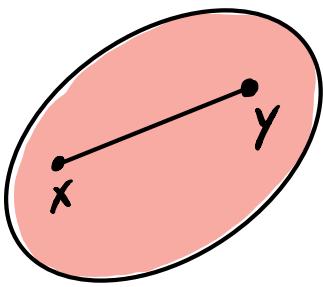
Let $f: \mathbb{R}^n \rightarrow \mathbb{R}$ and $K \subseteq \mathbb{R}^n$.

The constrained optimization problem

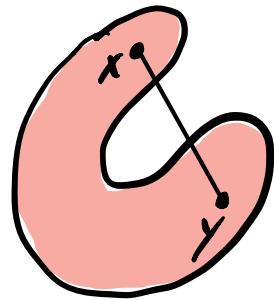
$\min f(x)$ subject to $x \in K$

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 .



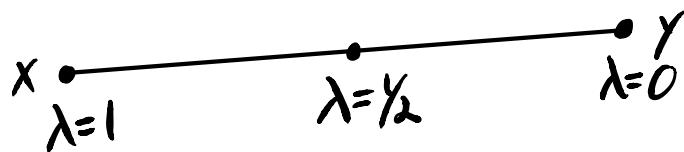
convex



not convex

The line segment between x and y
can be written as

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



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

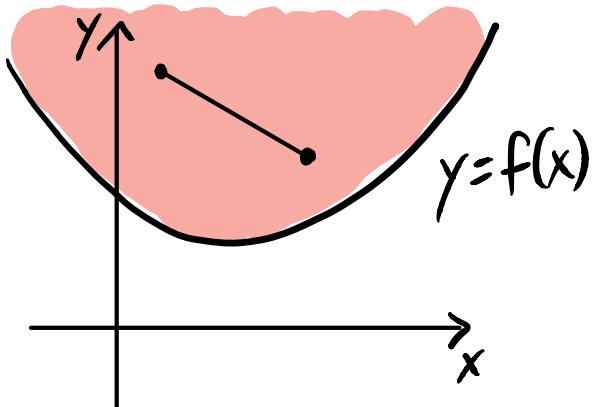
Examples:

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

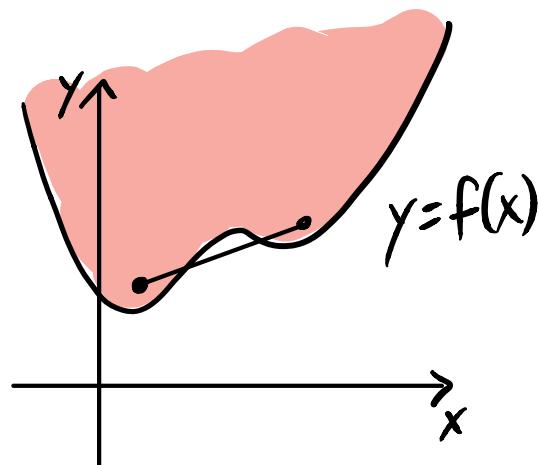
$$(2) \quad K = \{x \in \mathbb{R}^n : a^T x \leq b\}.$$

$$(3) \quad K = K_1 \cap K_2, \quad K_1 \text{ and } K_2 \text{ convex.}$$

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



f convex



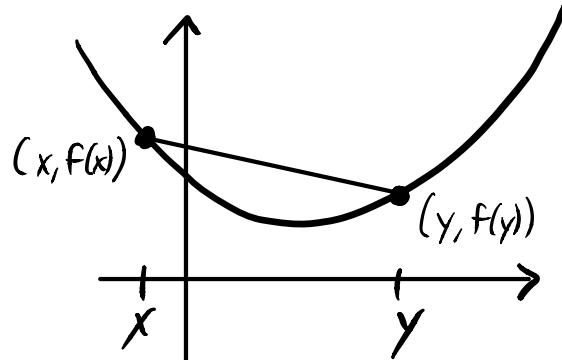
f not convex

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

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

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

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

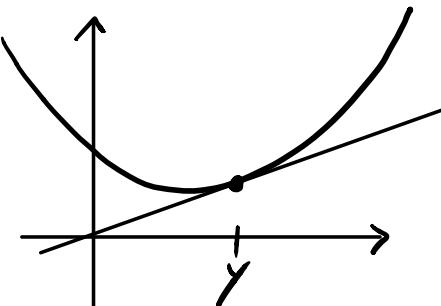
for all $x, y \in \mathbb{R}^n$ and $\lambda \in [0,1]$.

The function is strictly convex 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 iff 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:

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

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

$$(3) \quad f(x) = x_1^2 + 2x_1x_2 + x_2^2$$

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

$$(5) \quad f(x) = \|x\| \quad (\text{any vector norm}).$$
