Modern Optimization Techniques

1. Theory

Lars Schmidt-Thieme

Information Systems and Machine Learning Lab (ISMLL)
Institute for Computer Science
University of Hildesheim, Germany

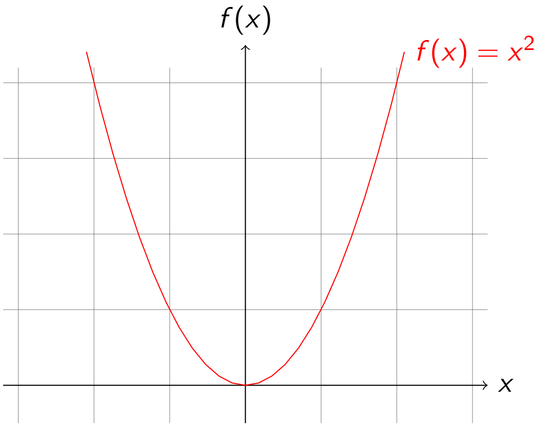
Syllabus

Mon. 07.11.	(1)	0. Overview
		1. Theory
Mon. 14.11.	(2)	1. Convex Sets and Functions
		2. Unconstrained Optimization
Mon. 21.11.	(3)	2.1 Gradient Descent
Mon. 28.11.	(4)	2.2 Stochastic Gradient Descent
Mon. 05.12.	(5)	2.3 Newton's Method
Mon. 12.12.	(6)	2.4 Quasi-Newton Methods
Mon. 19.12.	(7)	*canceled*
	—	— *Christmas Break* —
Mon. 09.01.	(8)	2.5 Subgradient Methods
Fri.. 13.01.	(9)	2.6 Coordinate Descent
		3. Equality Constrained Optimization
Mon. 23.01.	(10)	3.1 Duality
Mon. 30.01.	(11)	3.2 Methods
		4. Inequality Constrained Optimization
Mon. 06.01.	(12)	4.1 Primal Methods
Mon. 13.02.	(13)	4.2 Barrier and Penalty Methods

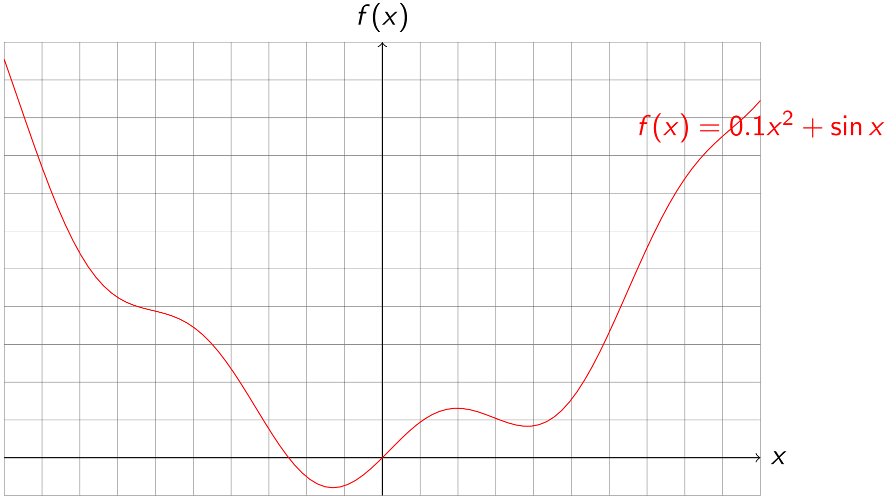
Outline

Outline

A convex function



A non-convex function



Convex Optimization Problem

An **optimization problem**

$$
\begin{aligned}
\text{minimize} \quad & f(x) \\
\text{subject to} \quad & h_q(x) \leq 0, \quad q = 1, \ldots, Q \\
& Ax = b
\end{aligned}
$$

is said to be convex if $f, h_1 \ldots h_Q$ are convex.

Note: The equality constraints also are convex, even linear.

Convex Optimization Problem

An **optimization problem**

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & h_q(x) \leq 0, \quad q = 1, \ldots, Q \\ & Ax = b \end{aligned}$$

is said to be convex if $f, h_1 \ldots h_Q$ are convex.

How do we know if a function is convex or not?

Note: The equality constraints also are convex, even linear.

Outline

Affine Sets

For any two points x_1, x_2 we can define the line through them as:

Affine Sets

For any two points x_1, x_2 we can define the line through them as:

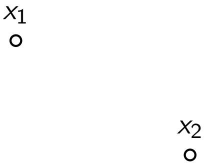
$$x = \theta x_1 + (1 - \theta) x_2, \quad \theta \in \mathbb{R}$$

Affine Sets

For any two points x_1, x_2 we can define the line through them as:

$$x = \theta x_1 + (1 - \theta)x_2, \quad \theta \in \mathbb{R}$$

Example:

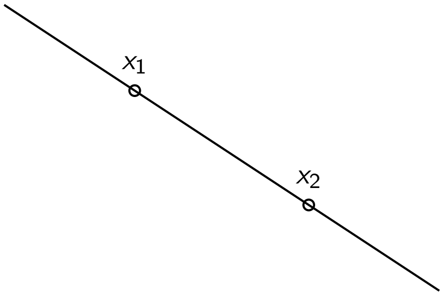


Affine Sets

For any two points x_1, x_2 we can define the line through them as:

$$x = \theta x_1 + (1 - \theta) x_2, \quad \theta \in \mathbb{R}$$

Example:

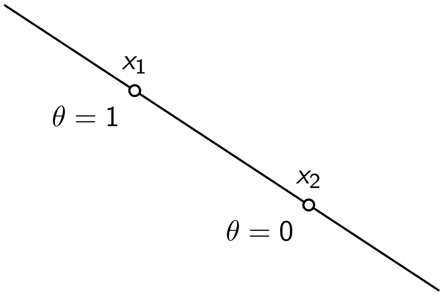


Affine Sets

For any two points x_1, x_2 we can define the line through them as:

$$x = \theta x_1 + (1 - \theta) x_2, \quad \theta \in \mathbb{R}$$

Example:

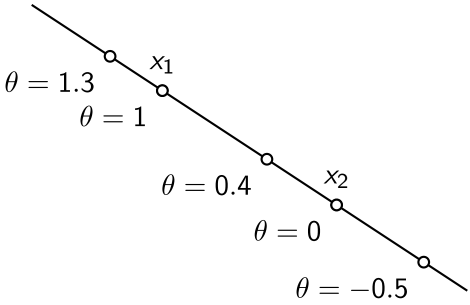


Affine Sets

For any two points x_1, x_2 we can define the line through them as:

$$x = \theta x_1 + (1 - \theta) x_2, \quad \theta \in \mathbb{R}$$

Example:



Affine Sets - Definition

An **affine set** is a set containing the line through any two distinct points in it.

Affine Sets - Definition

An **affine set** is a set containing the line through any two distinct points in it.

Examples:

- $\mathbb{R}^N$ for $N \in \mathbb{N}^+$
- solution set of linear equations $X := \{x \in \mathbb{R}^N \mid Ax = b\}$

Convex Sets

The **line segment** between any two points x_1, x_2 is the set of all points:

Convex Sets

The **line segment** between any two points x_1, x_2 is the set of all points:

$$x = \theta x_1 + (1 - \theta) x_2, \quad 0 \leq \theta \leq 1$$

Convex Sets

The **line segment** between any two points x_1, x_2 is the set of all points:

$$x = \theta x_1 + (1 - \theta) x_2, \quad 0 \leq \theta \leq 1$$

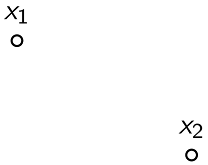
Example:

Convex Sets

The **line segment** between any two points x_1, x_2 is the set of all points:

$$x = \theta x_1 + (1 - \theta)x_2, \quad 0 \le \theta \le 1$$

Example:

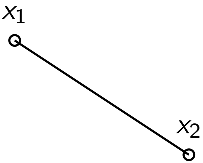


Convex Sets

The **line segment** between any two points x_1, x_2 is the set of all points:

$$x = \theta x_1 + (1-\theta)x_2, \quad 0 \le \theta \le 1$$

Example:

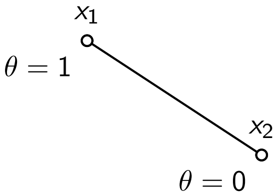


Convex Sets

The **line segment** between any two points x_1, x_2 is the set of all points:

$$x = \theta x_1 + (1-\theta)x_2, \quad 0 \le \theta \le 1$$

Example:

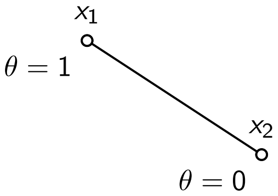


Convex Sets

The **line segment** between any two points x_1, x_2 is the set of all points:

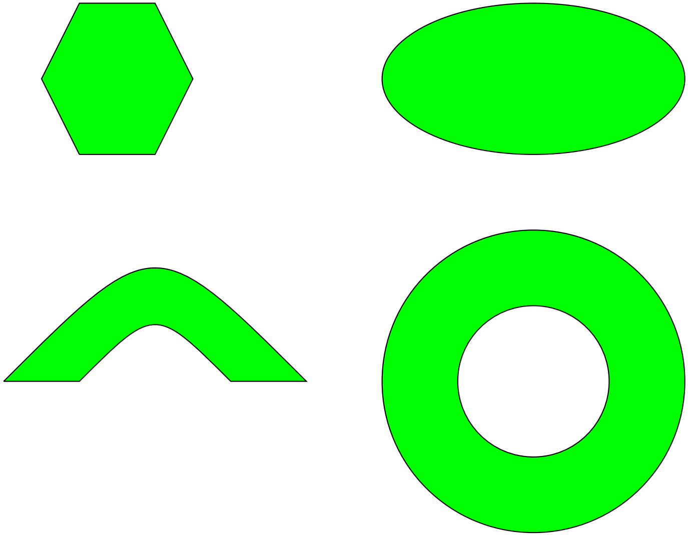
$$x = \theta x_1 + (1-\theta)x_2, \quad 0 \leq \theta \leq 1$$

Example:



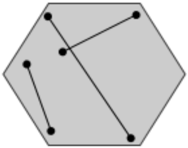
A **convex set** contains the line segment between any two points in the set.

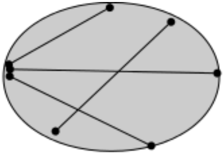
Convex Sets - Examples: Which ones are Convex?



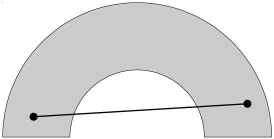
Convex Sets - Examples

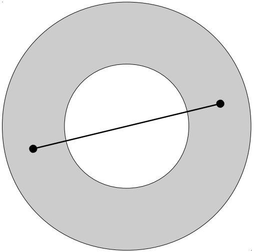
Convex Sets:





Non-convex Sets:





Convex Sets - Examples

All affine sets are also convex:

- $\mathbb{R}^N$ for $N \in \mathbb{N}^+$

- solution set of linear equations $X := \{x \in \mathbb{R}^N \mid Ax = b\}$

Convex sets (but in general not affine sets):

- solution set of linear inequalities $X := \{x \in \mathbb{R}^N \mid Ax \leq b\}$
  - half spaces, e.g. $X := \{x \in \mathbb{R}^N \mid a^T x \leq b\}$
    e.g., $X := \{x \in \mathbb{R}^N \mid x_1 \geq 0\}$

  - convex polygons (2d) / polyhedrons (3d) / polytopes (nd)

Convex Combination and Convex Hull

(standard) simplex:

$$\begin{aligned}\Delta^N :=& \{\theta \in \mathbb{R}^N \mid \theta_n \geq 0, n = 1, \ldots, N; \sum_{n=1}^{N} \theta_n = 1\} \\ =& \{\theta \in [0,1]^N \mid \mathbb{1}^T \theta = 1\}\end{aligned}$$

convex combination of some points $x_1, \ldots x_N \in \mathbb{R}^M$: any point x with

$$x = \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_N x_N, \quad \theta \in \Delta^N$$

convex hull of a set $X \subseteq \mathbb{R}^M$ of points:

$$\text{conv}(X) := \{\theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_N x_N \mid N \in \mathbb{N}, x_1, \ldots, x_N \in X, \theta \in \Delta^N\}$$

i.e., the set of all convex combinations of points in X.

Note: $\mathbb{1} := (1, 1, \ldots, 1)^T$ vector of all ones.

Outline

Convex Functions

A function $f : X \to \mathbb{R}, X \subseteq \mathbb{R}^N$ is **convex** iff:

Convex Functions

A function $f : X \to \mathbb{R}, X \subseteq \mathbb{R}^N$ is **convex** iff:

- $\operatorname{dom} f := X$ is a convex set

Convex Functions

A function $f : X \to \mathbb{R}, X \subseteq \mathbb{R}^N$ is **convex** iff:

- $\operatorname{dom} f := X$ is a convex set
- for all $x_1, x_2 \in \operatorname{dom} f$ and $0 \leq \theta \leq 1$ it satistfies

$$f(\theta x_1 + (1-\theta)x_2) \leq \theta f(x_1) + (1-\theta) f(x_2)$$

(the function is below of its secant segments/chords.)

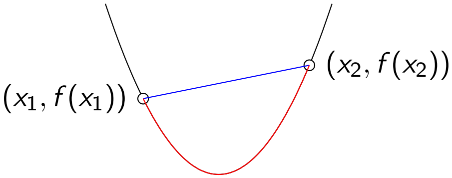
Convex Functions

A function $f : X \to \mathbb{R}, X \subseteq \mathbb{R}^N$ is **convex** iff:

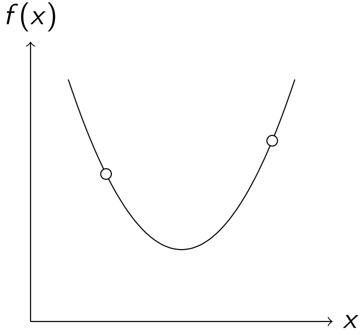
- $\operatorname{dom} f := X$ is a convex set
- for all $x_1, x_2 \in \operatorname{dom} f$ and $0 \leq \theta \leq 1$ it satistfies

$$f(\theta x_1 + (1-\theta)x_2) \leq \theta f(x_1) + (1-\theta) f(x_2)$$

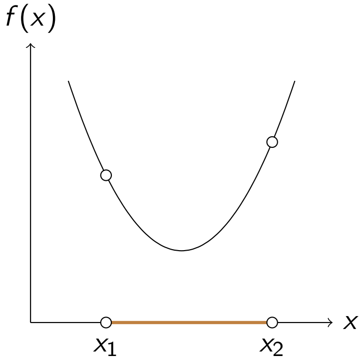
(the function is below of its secant segments/chords.)



Convex functions

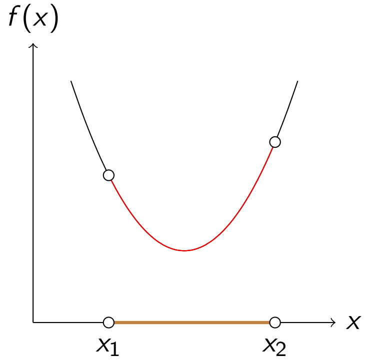


Convex functions



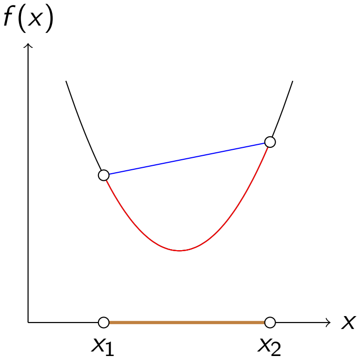
- $\theta x_1 + (1-\theta)x_2$

Convex functions



- $\theta x_1 + (1-\theta)x_2$
- $(\theta x_1 + (1-\theta)x_2, f(\theta x_1 + (1-\theta)x_2))$

Convex functions



- $\theta x_1 + (1-\theta)x_2$
- $(\theta x_1 + (1-\theta)x_2, f(\theta x_1 + (1-\theta)x_2))$
- $(\theta x_1 + (1-\theta)x_2, \theta f(x_1) + (1-\theta)f(x_2))$

How are Convex Functions Related to Convex Sets?

epigraph of a function $f : X \to \mathbb{R}, X \subseteq \mathbb{R}^N$:

$$\text{epi}(f) := \{(x, y) \in X \times \mathbb{R} \mid y \geq f(x)\}$$

How are Convex Functions Related to Convex Sets?

epigraph of a function $f : X \rightarrow \mathbb{R}, X \subseteq \mathbb{R}^N$:

$$\text{epi}(f) := \{(x, y) \in X \times \mathbb{R} \mid y \geq f(x)\}$$

f is convex (as function) $\iff$ epi(f) is convex (as set).

proof is straight-forward (try it!)

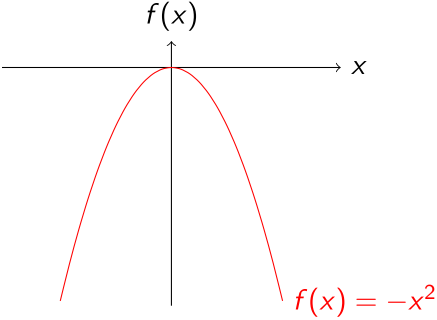
Concave Functions

A function f is called **concave** if $-f$ is convex

Concave Functions

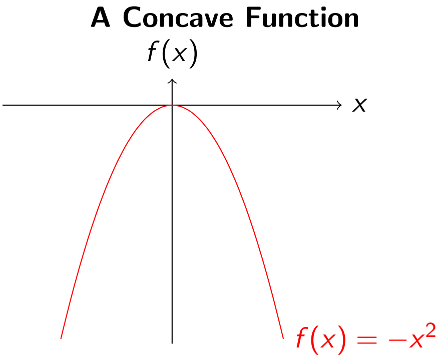
A function f is called **concave** if $-f$ is convex

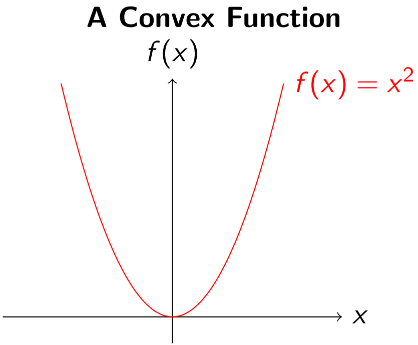
A Concave Function



Concave Functions

A function f is called **concave** if $-f$ is convex





Strictly Convex Functions

A function $f : X \to \mathbb{R}, X \subseteq \mathbb{R}^N$ is **strictly convex** if:

Strictly Convex Functions

A function $f : X \to \mathbb{R}, X \subseteq \mathbb{R}^N$ is **strictly convex** if:

- $\operatorname{dom} f$ is a convex set

Strictly Convex Functions

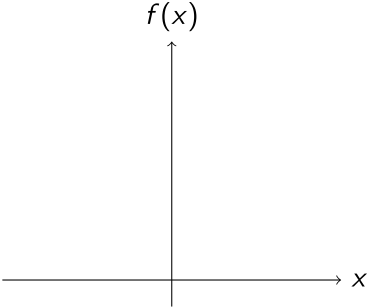
A function $f : X \to \mathbb{R}, X \subseteq \mathbb{R}^N$ is **strictly convex** if:

- $\operatorname{dom} f$ is a convex set
- for all $x_1, x_2 \in \operatorname{dom} f$, $x_1 \neq x_2$ and $0 < \theta < 1$ it satistfies

$$f(\theta x_1 + (1-\theta)x_2) < \theta f(x_1) + (1-\theta)f(x_2)$$

Examples

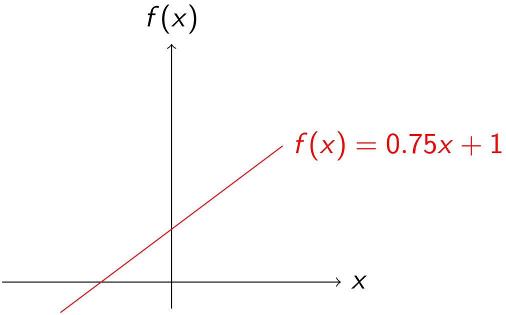
Examples of Convex functions:



Examples

Examples of Convex functions:

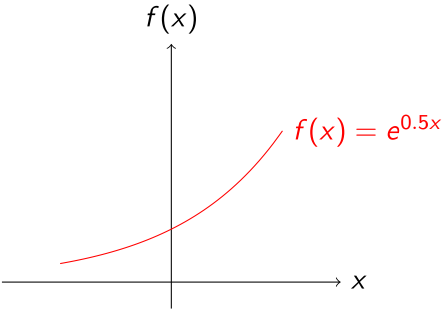
- affine: $f(x) := ax + b$, with $\operatorname{dom} f := \mathbb{R}$ and $a, b \in \mathbb{R}$



Examples

Examples of Convex functions:

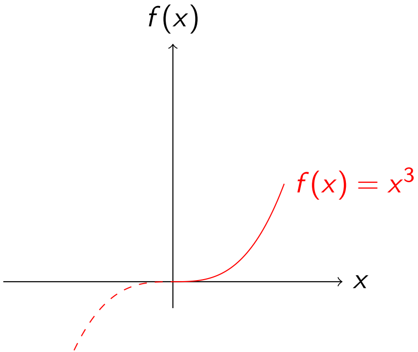
- affine: $f(x) := ax + b$, with $\operatorname{dom} f := \mathbb{R}$ and $a, b \in \mathbb{R}$
- exponential: $f(x) := e^{ax}$, with $a \in \mathbb{R}$



Examples

Examples of Convex functions:

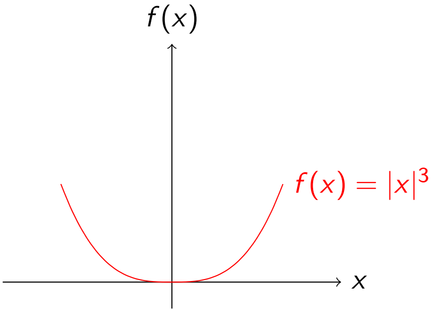
- affine: $f(x) := ax + b$, with $\operatorname{dom} f := \mathbb{R}$ and $a, b \in \mathbb{R}$
- exponential: $f(x) := e^{ax}$, with $a \in \mathbb{R}$
- powers: $f(x) := x^a$, with $\operatorname{dom} f := \mathbb{R}_0^+$ and $a \geq 1$ or $a \leq 0$



Examples

Examples of Convex functions:

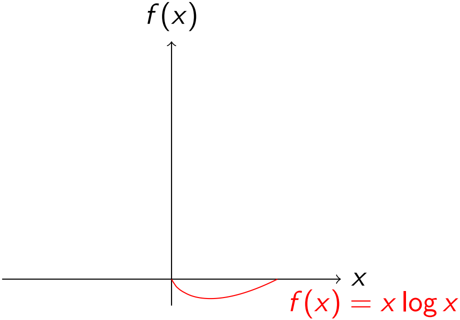
- affine: $f(x) := ax + b$, with $\operatorname{dom} f := \mathbb{R}$ and $a, b \in \mathbb{R}$
- exponential: $f(x) := e^{ax}$, with $a \in \mathbb{R}$
- powers: $f(x) := x^a$, with $\operatorname{dom} f := \mathbb{R}_0^+$ and $a \geq 1$ or $a \leq 0$
- powers of absolute value: $f(x) := |x|^a$, with $\operatorname{dom} f := \mathbb{R}$ and $a \geq 1$



Examples

Examples of Convex functions:

- affine: $f(x) := ax + b$, with $\operatorname{dom} f := \mathbb{R}$ and $a, b \in \mathbb{R}$
- exponential: $f(x) := e^{ax}$, with $a \in \mathbb{R}$
- powers: $f(x) := x^a$, with $\operatorname{dom} f := \mathbb{R}_0^+$ and $a \geq 1$ or $a \leq 0$
- powers of absolute value: $f(x) := |x|^a$, with $\operatorname{dom} f := \mathbb{R}$ and $a \geq 1$
- negative entropy: $f(x) := x \log x$, with $\operatorname{dom} f := \mathbb{R}^+$



Examples

Examples of Convex functions:

- affine: $f(x) := ax + b$, with $\operatorname{dom} f := \mathbb{R}$ and $a, b \in \mathbb{R}$
- exponential: $f(x) := e^{ax}$, with $a \in \mathbb{R}$
- powers: $f(x) := x^a$, with $\operatorname{dom} f := \mathbb{R}_0^+$ and $a \geq 1$ or $a \leq 0$
- powers of absolute value: $f(x) := |x|^a$, with $\operatorname{dom} f := \mathbb{R}$ and $a \geq 1$
- negative entropy: $f(x) := x \log x$, with $\operatorname{dom} f := \mathbb{R}^+$

Examples of Concave Functions:

Examples

Examples of Convex functions:

- affine: $f(x) := ax + b$, with $\operatorname{dom} f := \mathbb{R}$ and $a, b \in \mathbb{R}$
- exponential: $f(x) := e^{ax}$, with $a \in \mathbb{R}$
- powers: $f(x) := x^a$, with $\operatorname{dom} f := \mathbb{R}_0^+$ and $a \geq 1$ or $a \leq 0$
- powers of absolute value: $f(x) := |x|^a$, with $\operatorname{dom} f := \mathbb{R}$ and $a \geq 1$
- negative entropy: $f(x) := x \log x$, with $\operatorname{dom} f := \mathbb{R}^+$

Examples of Concave Functions:

- affine: $f(x) := ax + b$, with $\operatorname{dom} f := \mathbb{R}$ and $a, b \in \mathbb{R}$

Examples

Examples of Convex functions:

- affine: $f(x) := ax + b$, with $\operatorname{dom} f := \mathbb{R}$ and $a, b \in \mathbb{R}$
- exponential: $f(x) := e^{ax}$, with $a \in \mathbb{R}$
- powers: $f(x) := x^a$, with $\operatorname{dom} f := \mathbb{R}_0^+$ and $a \geq 1$ or $a \leq 0$
- powers of absolute value: $f(x) := |x|^a$, with $\operatorname{dom} f := \mathbb{R}$ and $a \geq 1$
- negative entropy: $f(x) := x \log x$, with $\operatorname{dom} f := \mathbb{R}^+$

Examples of Concave Functions:

- affine: $f(x) := ax + b$, with $\operatorname{dom} f := \mathbb{R}$ and $a, b \in \mathbb{R}$
- powers/roots: $f(x) := x^a$, with $\operatorname{dom} f := \mathbb{R}_0^+$ and $0 \leq a \leq 1$

Examples

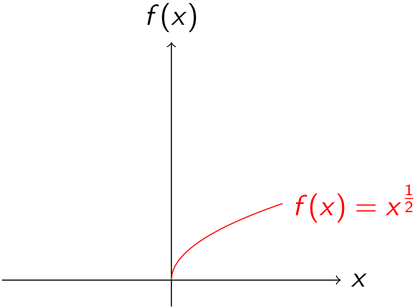
Examples of Convex functions:

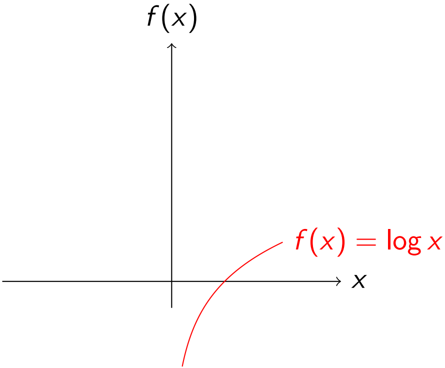
- affine: $f(x) := ax + b$, with $\operatorname{dom} f := \mathbb{R}$ and $a, b \in \mathbb{R}$
- exponential: $f(x) := e^{ax}$, with $a \in \mathbb{R}$
- powers: $f(x) := x^a$, with $\operatorname{dom} f := \mathbb{R}_0^+$ and $a \geq 1$ or $a \leq 0$
- powers of absolute value: $f(x) := |x|^a$, with $\operatorname{dom} f := \mathbb{R}$ and $a \geq 1$
- negative entropy: $f(x) := x \log x$, with $\operatorname{dom} f := \mathbb{R}^+$

Examples of Concave Functions:

- affine: $f(x) := ax + b$, with $\operatorname{dom} f := \mathbb{R}$ and $a, b \in \mathbb{R}$
- powers/roots: $f(x) := x^a$, with $\operatorname{dom} f := \mathbb{R}_0^+$ and $0 \leq a \leq 1$
- logarithm: $f(x) := \log x$, with $\operatorname{dom} f := \mathbb{R}^+$

Examples





Examples

Examples of Convex functions:

All norms are convex!

- Immediate consequence of the triangle inequality and absolute homogeneity.

$$||\theta x + (1-\theta)y|| \leq ||\theta x|| + ||(1-\theta)y|| = \theta||x|| + (1-\theta)||y||$$

Examples

Examples of Convex functions:

All norms are convex!

- Immediate consequence of the triangle inequality and absolute homogeneity.

$$||\theta x + (1-\theta)y|| \leq ||\theta x|| + ||(1-\theta)y|| = \theta||x|| + (1-\theta)||y||$$

- For $\mathbf{x} \in \mathbb{R}^N$, $p \geq 1$:
  p-norm: $||\mathbf{x}||_p := (\sum_{n=1}^{N} |x_n|^p)^{\frac{1}{p}}$,

Examples

Examples of Convex functions:

All norms are convex!

- Immediate consequence of the triangle inequality and absolute homogeneity.

$$||\theta x + (1-\theta)y|| \leq ||\theta x|| + ||(1-\theta)y|| = \theta||x|| + (1-\theta)||y||$$

- For $\mathbf{x} \in \mathbb{R}^N$, $p \geq 1$:
  p-norm: $||\mathbf{x}||_p := (\sum_{n=1}^N |x_n|^p)^{\frac{1}{p}}$,

- $||\mathbf{x}||_\infty := \max_{n=1:N} |x_n|$

Examples

Examples of Convex functions:

All norms are convex!

- Immediate consequence of the triangle inequality and absolute homogeneity.

$$||\theta x + (1-\theta)y|| \leq ||\theta x|| + ||(1-\theta)y|| = \theta||x|| + (1-\theta)||y||$$

- For $\mathbf{x} \in \mathbb{R}^N$, $p \geq 1$:
  p-norm: $||\mathbf{x}||_p := (\sum_{n=1}^{N} |x_n|^p)^{\frac{1}{p}}$,

- $||\mathbf{x}||_\infty := \max_{n=1:N} |x_n|$

Affine functions on vectors are also convex: $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} + b$

Outline

1st-Order Condition

f is **differentiable** if $\mathrm{dom}\, f$ is open and the gradient

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \ldots, \frac{\partial f(\mathbf{x})}{\partial x_n} \right)^T$$

exists everywhere.

1st-order condition: a differentiable function f is convex iff

1st-Order Condition

f is **differentiable** if $\operatorname{dom} f$ is open and the gradient

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \ldots, \frac{\partial f(\mathbf{x})}{\partial x_n} \right)^T$$

exists everywhere.

1st-order condition: a differentiable function f is convex iff

- $\operatorname{dom} f$ is a convex set

1st-Order Condition

f is **differentiable** if $\operatorname{dom} f$ is open and the gradient

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \ldots, \frac{\partial f(\mathbf{x})}{\partial x_n}\right)^T$$

exists everywhere.

1st-order condition: a differentiable function f is convex iff

- $\operatorname{dom} f$ is a convex set
- for all $\mathbf{x}, \mathbf{y} \in \operatorname{dom} f$

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x})$$

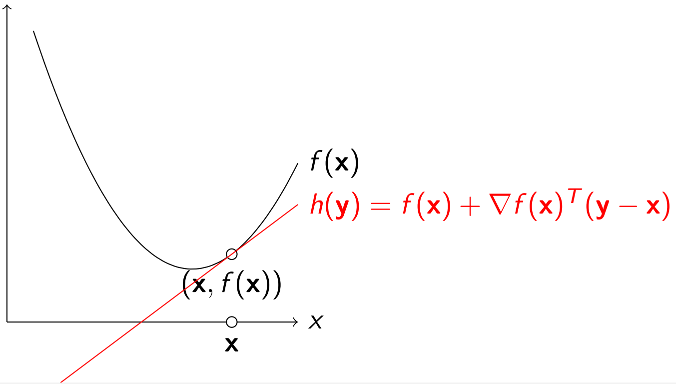
(the function is above (or on) any of its tangents.)

1st-Order Condition

1st-order condition: a differentiable function f is convex iff

- $\mathrm{dom}\, f$ is a convex set
- for all $\mathbf{x}, \mathbf{y} \in \mathrm{dom}\, f$

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x})$$



1st-Order Condition / Proof

Let $\operatorname{dom} f = X$ be convex.

$$f : X \to \mathbb{R} \text{ convex} \Leftrightarrow f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) \quad \forall \mathbf{x}, \mathbf{y}$$

1st-Order Condition / Proof

Let $\operatorname{dom} f = X$ be convex.

$$f : X \to \mathbb{R} \text{ convex} \Leftrightarrow f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) \quad \forall \mathbf{x}, \mathbf{y}$$

$$\text{“} \Rightarrow \text{”} : f(x + t(y-x)) \leq (1-t) f(x) + t f(y) \quad | : t$$

1st-Order Condition / Proof

Let $\operatorname{dom} f = X$ be convex.

$$f : X \to \mathbb{R} \text{ convex} \Leftrightarrow f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) \quad \forall \mathbf{x}, \mathbf{y}$$

$$\text{“} \Rightarrow \text{”} : f(x + t(y - x)) \leq (1 - t)f(x) + tf(y) \quad | : t$$

$$f(y) \geq \frac{f(x + t(y - x)) - f(x)}{t} + f(x) \stackrel{t \to 0^+}{\longrightarrow} \nabla f(x)^T(y - x) + f(x)$$

1st-Order Condition / Proof

Let $\operatorname{dom} f = X$ be convex.

$$f : X \to \mathbb{R} \text{ convex} \Leftrightarrow f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) \quad \forall \mathbf{x}, \mathbf{y}$$

$$\text{“} \Rightarrow \text{”} : f(x + t(y - x)) \leq (1 - t)f(x) + tf(y) \quad | : t$$

$$f(y) \geq \frac{f(x + t(y - x)) - f(x)}{t} + f(x) \overset{t \to 0^+}{\longrightarrow} \nabla f(x)^T(y - x) + f(x)$$

“ $\Leftarrow$ ” : Apply twice to $z := \theta x + (1 - \theta)y$

$$f(x) \geq f(z) + \nabla f(z)^T(x - z)$$

$$f(y) \geq f(z) + \nabla f(z)^T(y - z)$$

1st-Order Condition / Proof

Let $\operatorname{dom} f = X$ be convex.

$$f : X \to \mathbb{R} \text{ convex} \Leftrightarrow f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) \quad \forall \mathbf{x}, \mathbf{y}$$

$$\text{“} \Rightarrow \text{”} : f(x + t(y - x)) \leq (1 - t)f(x) + tf(y) \quad | : t$$

$$f(y) \geq \frac{f(x + t(y - x)) - f(x)}{t} + f(x) \overset{t \to 0^+}{\longrightarrow} \nabla f(x)^T(y - x) + f(x)$$

$$\text{“} \Leftarrow \text{”} : \text{Apply twice to } z := \theta x + (1 - \theta)y$$

$$f(x) \geq f(z) + \nabla f(z)^T(x - z)$$

$$f(y) \geq f(z) + \nabla f(z)^T(y - z)$$

$$\begin{aligned} \rightsquigarrow \theta f(x) + (1 - \theta)f(y) \quad &\geq f(z) + \nabla f(z)^T(\theta x + (1 - \theta)y) - \nabla f(z)^T z \\ &= f(z) + \nabla f(z)^T z - \nabla f(z)^T z \\ &= f(z) = f(\theta x + (1 - \theta)y) \end{aligned}$$

1st-Order Condition / Strict Variant

strict 1st-order condition: a differentiable function f is strictly convex iff

- $\mathrm{dom}\, f$ is a convex set
- for all $\mathbf{x}, \mathbf{y} \in \mathrm{dom}\, f, \mathbf{x} \neq \mathbf{y}$

$$f(\mathbf{y}) > f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x})$$

Let $\operatorname{dom} f = X$ be convex.

$$f : X \to \mathbb{R} \text{ convex} \Leftrightarrow f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) \quad \forall \mathbf{x}, \mathbf{y}$$

Q: What does this imply for points $\mathbf{x}$ with $\nabla f(\mathbf{x}) = 0$?

Global Minima

Let $\operatorname{dom} f = X$ be convex.

$$f : X \to \mathbb{R} \text{ convex} \Leftrightarrow f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) \quad \forall \mathbf{x}, \mathbf{y}$$

Consequence: Points x with $\nabla f(x) = 0$ are (equivalent) global minima.

- minima form a convex set
- if f is strictly convex: there is exactly one global minimum x^*.

2nd-Order Condition

f is **twice differentiable** if $\operatorname{dom} f$ is open and the Hessian $\nabla^2 f(x)$

$$\nabla^2 f(\mathbf{x})_{n,m} = \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_m}$$

exists everywhere.

2nd-order condition: a twice differentiable function f is convex iff

2nd-Order Condition

f is **twice differentiable** if $\mathrm{dom}\, f$ is open and the Hessian $\nabla^2 f(x)$

$$\nabla^2 f(\mathbf{x})_{n,m} = \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_m}$$

exists everywhere.

2nd-order condition: a twice differentiable function f is convex iff

- $\mathrm{dom}\, f$ is a convex set

2nd-Order Condition

f is **twice differentiable** if $\mathrm{dom}\, f$ is open and the Hessian $\nabla^2 f(x)$

$$\nabla^2 f(\mathbf{x})_{n,m} = \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_m}$$

exists everywhere.

2nd-order condition: a twice differentiable function f is convex iff

- $\mathrm{dom}\, f$ is a convex set
- for all $\mathbf{x} \in \mathrm{dom}\, f$

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

2nd-Order Condition

f is **twice differentiable** if $\mathrm{dom}\, f$ is open and the Hessian $\nabla^2 f(x)$

$$\nabla^2 f(\mathbf{x})_{n,m} = \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_m}$$

exists everywhere.

2nd-order condition: a twice differentiable function f is convex iff

- $\mathrm{dom}\, f$ is a convex set
- for all $\mathbf{x} \in \mathrm{dom}\, f$

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

Furthermore:

- for functions f on $\mathrm{dom}\, f \subseteq \mathbb{R}$ simply $f''(x) \geq 0$ for all $x \in \mathrm{dom}\, f$
- if $\nabla^2 f(\mathbf{x}) \succ 0$ for all $\mathbf{x} \in \mathrm{dom}\, f$, then f is strictly convex
  - the converse is not true,
    e.g., $f(x) = x^4$ is strictly convex, but has 0 derivative at 0.

Positive Semidefinite Matrices (Review)

A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is **positive semidefinite ($A \succeq 0$)**:

$$x^T A x \geq 0, \quad \forall x \in \mathbb{R}^N$$

Equivalent:

(i) all eigenvalues of A are ≥ 0.

(ii) $A = B^T B$ for some matrix B

Positive Semidefinite Matrices (Review)

A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is **positive semidefinite ($A \succeq 0$)**:

$$x^T A x \geq 0, \quad \forall x \in \mathbb{R}^N$$

Equivalent:

(i) all eigenvalues of A are ≥ 0.

(ii) $A = B^T B$ for some matrix B

A symmetric matrix $A \in \mathbb{R}^{N \times N}$ is **positive definite ($A \succ 0$)**:

$$x^T A x > 0, \quad \forall x \in \mathbb{R}^N \setminus \{0\}$$

Equivalent:

(i) all eigenvalues of A are > 0.

(ii) $A = B^T B$ for some nonsingular matrix B

Example: Quadratic Functions

$$f(x) := ax^2 + bx + c, \quad x, a, b, c \in \mathbb{R}$$

Example: Quadratic Functions

univariate / onedimensional:

$$f(x) := ax^2 + bx + c, \quad x, a, b, c \in \mathbb{R}$$

multivariate / multidimensional:

$$f(x) := x^T A x + b^T x + c, \quad x, b, c \in \mathbb{R}^N, \quad A \in \mathbb{R}^{N \times N}$$

Example: Quadratic Functions

univariate / onedimensional:

$$f(x) := ax^2 + bx + c, \quad x, a, b, c \in \mathbb{R}$$

multivariate / multidimensional:

$$f(x) := x^T Ax + b^T x + c, \quad x, b, c \in \mathbb{R}^N, \quad A \in \mathbb{R}^{N \times N}$$

Q: Are quadratic functions convex?

Example: Quadratic Functions

univariate / onedimensional:

$$f(x) := ax^2 + bx + c, \quad x, a, b, c \in \mathbb{R}$$

multivariate / multidimensional:

$$f(x) := x^T A x + b^T x + c, \quad x, b, c \in \mathbb{R}^N, \quad A \in \mathbb{R}^{N \times N}$$

A quadratic function f is convex $\iff$ A is pos. semidef.

Recognizing Convex Functions

- There are a number of operations that preserve the convexity of a function.

- If f can be obtained by applying those operations to a convex function, f is also convex.

Recognizing Convex Functions

- There are a number of operations that preserve the convexity of a function.

- If f can be obtained by applying those operations to a convex function, f is also convex.

Nonnegative multiple:

- if f is convex and $a \geq 0$ then af is convex.

- Example: $5x^2$ is convex since x^2 is convex

Recognizing Convex Functions

- There are a number of operations that preserve the convexity of a function.

- If f can be obtained by applying those operations to a convex function, f is also convex.

Nonnegative multiple:

- if f is convex and $a \geq 0$ then af is convex.

- Example: $5x^2$ is convex since x^2 is convex

Q: Is the sum of two conex functions g and h convex again?

Recognizing Convex Functions

- There are a number of operations that preserve the convexity of a function.

- If f can be obtained by applying those operations to a convex function, f is also convex.

Nonnegative multiple:

- if f is convex and $a \geq 0$ then af is convex.

- Example: $5x^2$ is convex since x^2 is convex

Sum:

- if g and h are convex functions, then $g + h$ is convex.

- Example: $f(x) = e^x + x \log x$ with $\operatorname{dom} f = \mathbb{R}^+$ is convex since e^x and $x \log x$ are convex

Recognizing Convex Functions / Composition

Composition of two convex functions:

- let $g : \mathbb{R}^N \to \mathbb{R}$, $h : \mathbb{R} \to \mathbb{R}$ be both convex and

$$f(x) := h(g(x))$$

- in general f is **not** convex
- counter example $N = 1, \quad g(x) = h(x) = e^{-x}$:

$$\begin{aligned}
\left(e^{-e^{-x}}\right)'' &= \left(e^{-e^{-x}}(-e^{-x})(-1)\right)' \\
&= \left(e^{-e^{-x}} e^{-x}\right)' \\
&= e^{-e^{-x}} e^{-x} e^{-x} + e^{-e^{-x}} e^{-x}(-1) \\
&= e^{-e^{-x}} e^{-x}(e^{-x} - 1) < 0 \quad \text{for } x > 0
\end{aligned}$$

Recognizing Convex Functions / Composition

Composition with affine functions:

- if f is convex then $f(A\mathbf{x} + \mathbf{b})$ is convex.

Recognizing Convex Functions / Composition

Composition with affine functions:

- if f is convex then $f(A\mathbf{x} + \mathbf{b})$ is convex.
- Example: norm of an affine function $||A\mathbf{x} + \mathbf{b}||$

Recognizing Convex Functions / Composition

Composition with nondecreasing functions:

- if $g : \mathbb{R}^N \to \mathbb{R}$, $h : \mathbb{R} \to \mathbb{R}$ and

$$f(\mathbf{x}) = h(g(\mathbf{x}))$$

Recognizing Convex Functions / Composition

Composition with nondecreasing functions:

- if $g : \mathbb{R}^N \to \mathbb{R}$, $h : \mathbb{R} \to \mathbb{R}$ and

$$f(\mathbf{x}) = h(g(\mathbf{x}))$$

- f is convex if:
  - g is convex, h is convex and nondecreasing *or*
  - g is concave, h is convex and nonincreasing

Recognizing Convex Functions / Composition

Composition with nondecreasing functions:

- if $g : \mathbb{R}^N \to \mathbb{R}$, $h : \mathbb{R} \to \mathbb{R}$ and

$$f(\mathbf{x}) = h(g(\mathbf{x}))$$

- f is convex if:
  - g is convex, h is convex and nondecreasing *or*
  - g is concave, h is convex and nonincreasing
- proof:

$$\begin{aligned}\nabla^2 h(g(\mathbf{x})) &= \nabla \left(h'(g(\mathbf{x}))\nabla g(\mathbf{x})\right) \\ &= h''(g(\mathbf{x}))\nabla g(\mathbf{x})\nabla g(\mathbf{x})^T + h'(g(\mathbf{x}))\nabla^2 g(\mathbf{x})\end{aligned}$$

- Examples:
  - $e^{g(\mathbf{x})}$ is convex if g is convex
  - $\frac{1}{g(\mathbf{x})}$ is convex if g is concave and positive

Recognizing Convex Functions

Pointwise Maximum:

- if $f_1, \ldots, f_M$ are convex functions then
  $f(\mathbf{x}) := \max\{f_1(\mathbf{x}), \ldots, f_M(\mathbf{x})\}$ is convex.

Recognizing Convex Functions

Pointwise Maximum:

- if $f_1, \ldots, f_M$ are convex functions then
  $f(\mathbf{x}) := \max\{f_1(\mathbf{x}), \ldots, f_M(\mathbf{x})\}$ is convex.

- Example: $f(\mathbf{x}) := \max_{m=1,\ldots,M}(a_m^T \mathbf{x} + b_m)$ is convex

Recognizing Convex Functions

There are many different ways to establish the convexity of a function:

1. Check the definition.

Recognizing Convex Functions

There are many different ways to establish the convexity of a function:

1. Check the definition.
2. Show that $\nabla^2 f(\mathbf{x}) \succeq 0$ for twice differentiable functions.

Recognizing Convex Functions

There are many different ways to establish the convexity of a function:

1. Check the definition.
2. Show that $\nabla^2 f(\mathbf{x}) \succeq 0$ for twice differentiable functions.
3. Show that f can be obtained from other convex functions by operations that preserve convexity.

Summary

- **Convex sets** are closed under line segments (convex combinations).
- **Convex functions** are defined on a convex domain and
  - are below any of their secant segments / chords (definition),
  - are globally above (or on) their tangents (1st-order condition),
  - have a positive semidefinite Hessian (2nd-order condition).
- For convex functions, points with vanishing gradients are (equivalent) **global minima**.
- Operations that preserve convexity:
  - scaling with a nonnegative constant
  - sums
  - pointwise maximum
  - composition with an affine function
  - composition with a nondecreasing convex scalar function
  - composition of a noninc. convex scalar function with a concave funct.
    - esp. $-g$ for a concave g

Further Readings

- Convex sets:
  - Boyd and Vandenberghe, 2004, chapter 2, esp. 2.1
  - see also ch. 2.2 and 2.3
- Convex functions:
  - Boyd and Vandenberghe, 2004, chapter 3, esp. 3.1.1–7, 3.2.1–5
- Convex optimization:
  - Boyd and Vandenberghe, 2004, chapter 4, esp. 4.1–3
  - see also ch. 4.4

References

Boyd, Stephen and Lieven Vandenberghe (2004). *Convex Optimization*. Cambridge University Press.