Modern Optimization Techniques

Lucas Rego Drumond

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

Theory Background

Outline

1. Introduction

2. Convex Sets

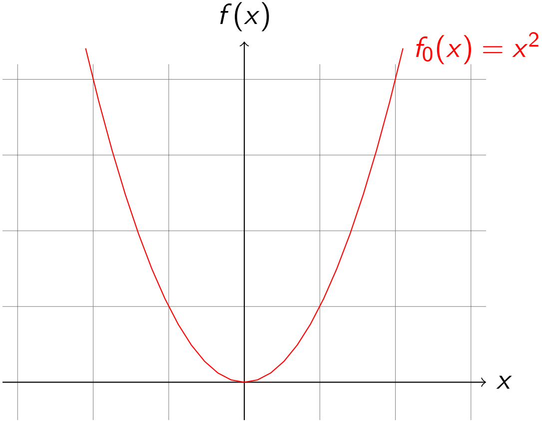
3. Convex Functions

4. Optimization Problems

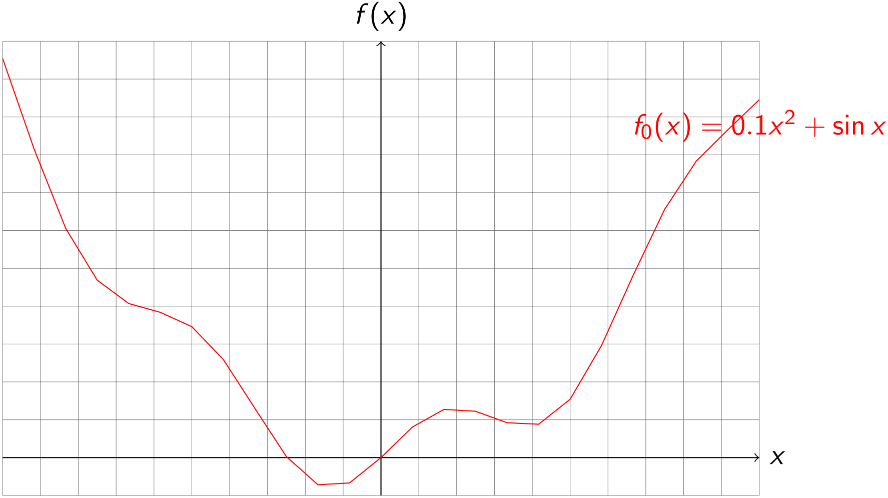
Outline

1. Introduction
2. Convex Sets
3. Convex Functions
4. Optimization Problems

A convex function



A non-convex function



Convex Optimization Problem

An **optimization problem**

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{aligned}$$

is said to be convex if $f_0, \ldots f_m$ are convex

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

Outline

1. Introduction

2. Convex Sets

3. Convex Functions

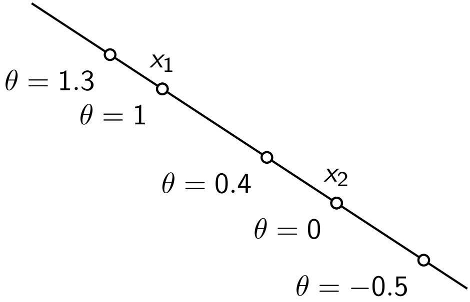
4. Optimization Problems

Affine Sets

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

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

Example:



Affine Sets - Definition

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

Examples:

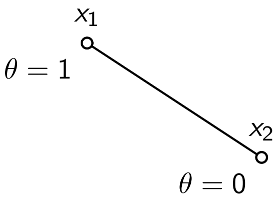
- $\mathbb{R}^n$ for $n \in \mathbb{N}^+$
- Solution set of linear equations $\{x | Ax = b\}$

Convex Sets

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

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

Example:

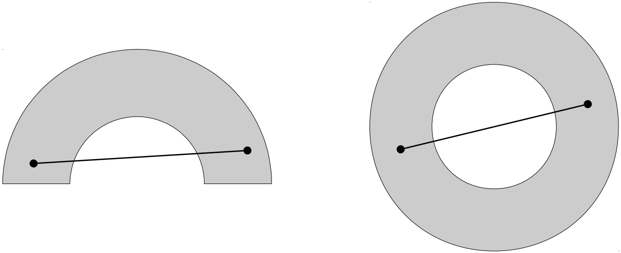


A **convex set** contains the line segment between any two points in the set

Convex Sets - Examples

Convex Sets:

Non-convex Sets:



Convex Combination and Convex Hull

Given a set of points $x_1, \ldots x_n$, a **convex combination** is the set of points x such that

$$x = \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_n x_n \qquad \theta_i \geq 0 \text{ and } \sum_{i=1}^{n} \theta_i = 1$$

A **convex hull** of a set is the set of all convex combinations of points in the set

Outline

1. Introduction

2. Convex Sets

3. Convex Functions

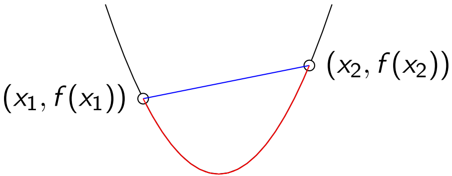
4. Optimization Problems

Convex Functions

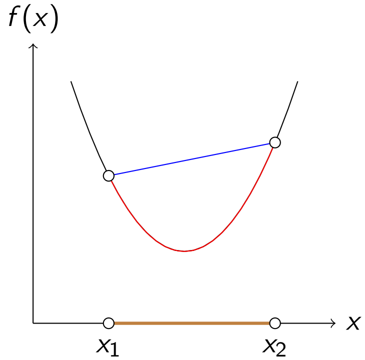
A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex iff:

- dom f is a convex set
- for all $x_1, x_2 \in$ dom f and $0 \leq \theta \leq 1$ it satistfies

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



Convex functions

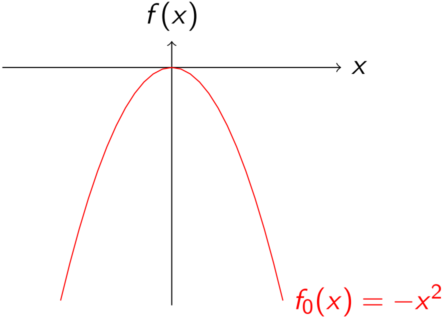


- $\theta x_1 + (1 - \theta) x_2$
- $(\theta x_1 + (1 - \theta) x_2, f(\theta x_1 + (1 - \theta) x_2))$
- $(\theta x_1 + (1 - \theta) x_2, \theta f(x_1) + (1 - \theta) f(x_2))$

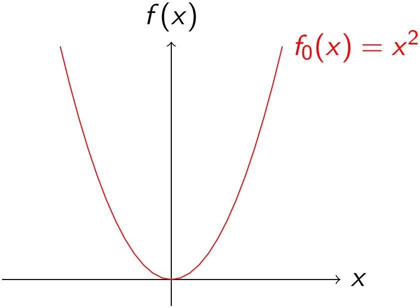
Convex Functions

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

A Concave Function



A Convex Function



Strictly Convex Functions

A function $f : \mathbb{R}^n \to \mathbb{R}$ is **strictly convex** if:

- dom f is a convex set
- for all $x_1, x_2 \in \text{dom } f$, $x \neq y$ 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:

- 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 $\operatorname{dom} f = \mathbb{R}^{++}$ 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: $f(x) = x^a$, with $\operatorname{dom} f = \mathbb{R}^{++}$ and $0 \leq a \leq 1$
- logarithm: $f(x) = \log x$, with $\operatorname{dom} f = \mathbb{R}^{++}$

Examples

Examples of Convex functions:

All norms are convex!

- For $\mathbf{x} \in \mathbb{R}^n$:
- p-norms: $||\mathbf{x}||_p = (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}$,
- for $p \geq 1$
- $||\mathbf{x}||_\infty = \max_k |x_k|$
- Affine functions on vectors are also convex: $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} + b$

1st-order condition

f is differentiable if dom f is open and the gradient

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

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

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

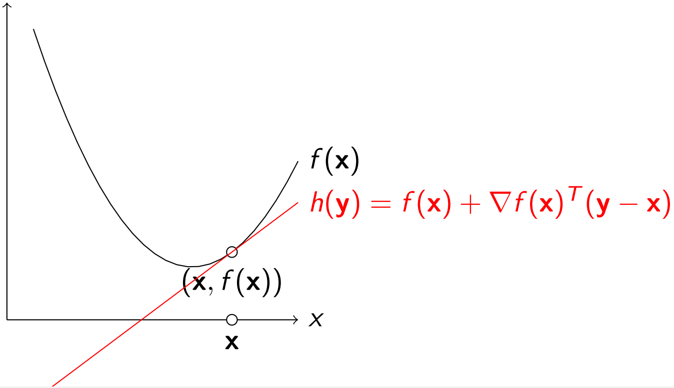
$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x})$$

1st-order condition

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

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

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



2nd-order condition

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

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

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

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

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

- if $\nabla^2 f(\mathbf{x}) \succ 0$ for all $\mathbf{x} \in \operatorname{dom} f$, then f is strictly 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 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 f_1 and f_2 are convex functions then $f_1 + f_2$ is convex
- Example: $f(x) = e^{3x} + x \log x$ with $\operatorname{dom} f = \mathbb{R}^{++}$ is convex since e^{3x} and $x \log x$ are convex

Recognizing Convex Functions

Composition with the affine function:

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

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: $f(\mathbf{x}) = \max_{i=1,\ldots,m}(a_i^T\mathbf{x} + b_i)$ is convex

Recognizing Convex Functions

Composition with scalar 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 h is nondecreasing *or*
  - g is concave, h is convex and h is nonincreasing
- 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

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

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

Outline

1. Introduction

2. Convex Sets

3. Convex Functions

4. Optimization Problems

Optimization Problem

$$
\begin{aligned}
\text{minimize} \quad & f_0(\mathbf{x}) \\
\text{subject to} \quad & f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, p \\
& h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, q
\end{aligned}
$$

- $f_0 : \mathbb{R}^n \to \mathbb{R}$ is the **objective function**
- $\mathbf{x} \in \mathbb{R}^n$ is the optimization variable
- $(f_i)_{i=1,\ldots,m} : \mathbb{R}^n \to \mathbb{R}$ are the inequality constraint functions
- $(h_i)_{i=1,\ldots,q} : \mathbb{R}^n \to \mathbb{R}$ are the equality constraint functions

Convex Optimization Problem

An **optimization problem**

$$\begin{aligned} \text{minimize} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, p \\ & h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, q \end{aligned}$$

is said to be convex if $f_0, \ldots f_p$ are convex and $h_1, \ldots, h_q$ are **affine**:

$$\begin{aligned} \text{minimize} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, p \\ & A\mathbf{x} = \mathbf{b} \end{aligned}$$

Practical Example: Household Spending

Suppose we have the following data about different households:

- Number of workers in the household (a_1)
- Household composition (a_2)
- Region (a_3)
- Gross normal weekly household income (a_4)
- **Weekly household spending** (y)

We want to create a model of the weekly household spending

Practical Example: Household Spending

If we have data about m households, we can represent it as:

$$A_{m,n} = \begin{pmatrix} 1 & a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ 1 & a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & a_{m,1} & a_{m,2} & a_{m,3} & a_{m,4} \end{pmatrix} \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

We can model the household consumption is a linear combination of the household features with parameters β:

$$\hat{y}_i = \beta^T \mathbf{a_i} = \beta_0 1 + \beta_1 a_{i,1} + \beta_2 a_{i,2} + \beta_3 a_{i,3} + \beta_4 a_{i,4}$$

Practical Example: Household Spending

We have:

$$\begin{pmatrix} 1 & a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ 1 & a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & a_{m,1} & a_{m,2} & a_{m,3} & a_{m,4} \end{pmatrix} \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} \approx \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

We want to find parameters β such that the measured error of the predictions is minimal:

$$\sum_{i=1}^{m} (\beta^T \mathbf{a_i} - y_i)^2 = ||A\beta - \mathbf{y}||_2^2$$

The Least Squares Problem

$$\text{minimize} \quad ||A\beta - \mathbf{y}||_2^2$$

$$||A\beta - \mathbf{y}||_2^2 = (A\beta - \mathbf{y})^T(A\beta - \mathbf{y})$$

$$\frac{d}{d\beta}(A\beta - \mathbf{y})^T(A\beta - \mathbf{y}) = 2A^T(A\beta - \mathbf{y})$$

$$\begin{aligned} 2A^T(A\beta - \mathbf{y}) &= 0 \\ A^TA\beta - A^T\mathbf{y} &= 0 \\ A^TA\beta &= A^T\mathbf{y} \\ \beta &= (A^TA)^{-1}A^T\mathbf{y} \end{aligned}$$

The Least Squares Problem

$$\text{minimize} \quad ||A\beta - \mathbf{y}||_2^2$$

- Convex Problem!
- Analytical solution: $\beta^* = (A^T A)^{-1} A^T \mathbf{y}$
- Often applied for data fitting
- $A\beta - \mathbf{y}$ is usually called the residual or error
- Extensions like the regularized least squares

Practical Example: Household Location

Suppose we have the following data about different households:

- Number of workers in the household (a_1)
- Household composition (a_2)
- Weekly household spending (a_3)
- Gross normal weekly household income (a_4)
- **Region** (y): North $y = 1$ or south $y = 0$

We want to create a model of the location of the household

Practical Example: Household Location

If we have data about m households, we can represent it as:

$$A_{m,n} = \begin{pmatrix} 1 & a_{1,2} & \dots & a_{1,n} \\ 1 & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & a_{m,2} & \dots & a_{m,n} \end{pmatrix} \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

We can model the household location is a linear combination of the household features with parameters β:

$$\hat{y}_i = \sigma(\beta^T \mathbf{a_i}) = \sigma(\beta_0 1 + \beta_1 a_{i,1} + \beta_2 a_{i,2} + \beta_3 a_{i,3} + \beta_4 a_{i,4})$$

where: $\sigma(x) = \frac{1}{1+e^{(-x)}}$

The Logistic Regression

The logistic regression learning problem is

$$\text{maximize} \quad \sum_{i=1}^{m} y_i \log \sigma(\beta^T \mathbf{a_i}) + (1 - y_i) \log(1 - \sigma(\beta^T \mathbf{a_i}))$$

$$A_{m,n} = \begin{pmatrix} 1 & a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ 1 & a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & a_{m,1} & a_{m,2} & a_{m,3} & a_{m,4} \end{pmatrix} \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

Linear Programming

$$
\begin{array}{rll}
\text{minimize} & \mathbf{c}^T \beta & \\
\text{subject to} & \mathbf{a}_i^T \beta \leq b_i & i = 1, \ldots, m
\end{array}
$$

- No simple analytical solution
- There are reliable algorithms available:
  - Simplex
  - Interior Points Method