Optimization

Fall semester 2022/23

Kristóf Bérczi

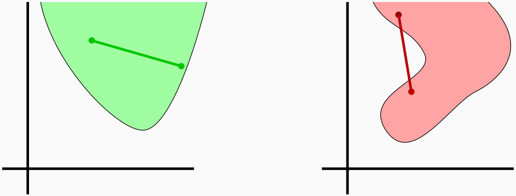
Eötvös Loránd University
Institute of Mathematics
Department of Operations Research

Lecture 3: Convexity

Convex sets

A set $K \subseteq \mathbb{R}^n$ is **convex** if for all $x, y \in K$ and $\theta \in [0,1]$, we have

$$\theta x + (1-\theta) y \in K.$$



Examples:

- **Polytopes:** $K = \{x \in \mathbb{R}^n : \langle a_i, x \rangle \leqslant b_i \text{ for } i = 1, \ldots, m\}$, where $a_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$ for $i = 1, \ldots, m$.
- **Ellipsoids:** $K = \{x \in \mathbb{R}^n : x^T A x \leqslant 1$ where $A \in \mathbb{R}^{n \times n}$ is a positive definite matrix.
- **Balls (in ℓ_p norms for $p \geqslant 1$):** $K = \{x \in \mathbb{R}^n : \sqrt[p]{\sum_{i=1}^n |x_i - a_i|^p} \leqslant 1\}$, where $a \in \mathbb{R}^n$ is a vector.

Convex functions

A function $f : \mathbb{R}^n \to \mathbb{R}$ is **convex** if its domain is a convex set and for all $x, y \in K$ and $\theta \in [0, 1]$, we have

$$f(\theta x + (1-\theta)y) \leqslant \theta f(x) + (1-\theta)f(y).$$

If the inequality always holds as strict inequality, the function is **strictly convex.**

The function f is **concave** or **strictly concave** if $-f$ is convex or strictly convex, respectively.

Remark: If $f : K \to \mathbb{R}^n$ is a convex function, then setting $f(x) = +\infty$ for $x \notin K$ results in a convex function when the arithmetic operations on $\mathbb{R} \cup \{+\infty\}$ are interpreted in the reasonable way.

Semidefinite matrices

A matrix $M \in \mathbb{R}^{n \times n}$ is **symmetric** if $M^T = M$.

The **identity matrix** of size $n \times n$ is denoted by I_n.

A symmetric matrix M is **positive semidefinite (PSD)** if $x^T M x \geqslant 0$ holds for all $x \in \mathbb{R}^n$, and this is denoted by $M \succeq 0$.

M is **positive definite (PD)** if $x^T M X > 0$ holds for all non-zero $x \in \mathbb{R}^n$, and this is denoted by $M \succ 0$.

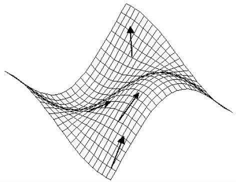
We define $M \succeq N \Leftrightarrow M - N \succeq 0$ and $M \succ N \Leftrightarrow M - N \succ 0$.

Calculus I

We are working with 'sufficiently smooth' functions $f : \mathbb{R}^n \to \mathbb{R}$.

The derivative of $f(x_1, \ldots, x_n)$ is called the **gradient**, and is defined as

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



The **directional derivative** of f in the direction d is $\langle \nabla f(x), d \rangle$.

The second derivatives of f can be summerized in the **Hessian** matrix

$$\begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \cdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Remark: The Hessian is symmetric if f is sufficiently differentiable.

Calculus II

Taylor expansion

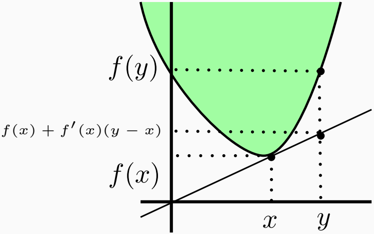
The Taylor series expansion of f around $x = a$ is

$$f(x) = \underbrace{\underbrace{f(a) + \langle \nabla f(a), x - a \rangle}_{\text{first order approximation}} + \frac{1}{2}(x-a)^T \nabla^2 f(a)(x-a)}_{\text{second order approximation}} + \dots$$

Consider a function in one dimension, i.e. $f : \mathbb{R} \to \mathbb{R}$.

When f is convex, the tangent is 'below' the graph, i.e.

$$f(y) \geqslant f(x) + f'(x)(y - x).$$



First order condition

First order condition

Let f be a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ over a convex set K. Then f is convex if and only if for all $x, y \in K$

$$f(y) \geqslant f(x) + \langle \nabla f(x), y - x \rangle.$$

Proof of the one-dimensional case.

$\Rightarrow$ For any $\theta \in [0, 1]$, we have

$$(1-\theta)f(x) + \theta f(y) \geqslant f(\theta y + (1-\theta)x) = f(x + \theta(y - x)).$$

Subtracting $(1-\theta)f(x)$ and dividing by θ yields

$$f(y) \geqslant f(x) + \frac{f(x + \theta(y-x)) - f(x)}{\theta}.$$

Taking limit $\theta \to 0$ gives $f(y) \geqslant f(x) + \langle \nabla f(x), y - x \rangle$. $\square$

First order condition

First order condition

Let f be a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ over a convex set K. Then f is convex if and only if for all $x, y \in K$

$$f(y) \geqslant f(x) + \langle \nabla f(x), y - x \rangle.$$

Proof of the one-dimensional case.

$\Leftarrow$ Let $z = \theta x + (1-\theta)y$. The first order approximation underestimates both $f(x)$ and $f(y)$, hence

$$f(x) \geqslant f(z) + \nabla(z)^T(x - z),$$
$$f(y) \geqslant f(z) + \nabla(z)^T(y - z).$$

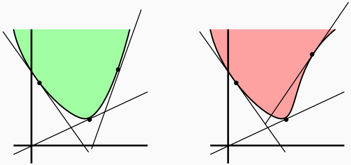
Therefore

$$(1-\theta)f(x) + \theta f(y) \geqslant f(z) + \nabla f(z)^T(\theta x + (1-\theta)y - z) = f(\theta(y) + (1-\theta)x).$$

$\square$

Second order condition

In the one-dimensional case, $f''(x) \geqslant 0$ when f is convex, that is, the slope of the tangent is non-decreasing, as otherwise when the slope decreases the function becomes non-convex.



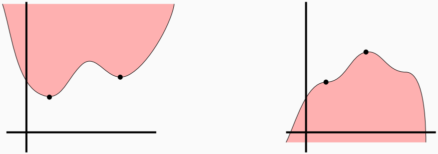
Second order condition

Let f be twice differentiable such that $\operatorname{dom} f$ is open. Then f is convex if and only if $\nabla^2 f(x) \succeq 0$ for all $x \in \operatorname{dom} f$.

Local vs. global optimum I

Convex optimization problem:

$$\inf_{x\in K} f(x) \quad \overset{\text{usually}}{\longrightarrow} \quad \min_{x\in K} f(x)$$



Intuition: $\nabla f(x) = 0$ when x is optimal.

Problem: $\nabla f(x) = 0$ may correspond to a local optimum/maximum.

Global optimum

If the domain of a convex differentiable function f is $\mathbb{R}^n$, then x is an optimal solution to $\inf_{x\in\mathbb{R}^n} f(x)$ if and only if $\nabla f(x) = 0$.

Local vs. global optimum II

Proof of the 'if' direction.
Assume that $\nabla f(x_0) = 0$. Since f is convex, we know that for all $y \in \mathbb{R}^n$ we have

$$\begin{aligned} f(y) &\geqslant f(x_0) + \langle \nabla f(x_0), y - x_0 \rangle \\ &= f(x_0) + \langle 0, y - x_0 \rangle \\ &= f(x_0). \end{aligned}$$

□

Remark: In the constrained setting, i.e. when $K \neq \mathbb{R}^n$, the following holds.

Global optimum

If f is a convex differentiable function, then x is an optimal solution to $\inf_{x \in \mathbb{R}^n} f(x)$ if and only if $\langle \nabla f(x), y - x \rangle \geqslant 0$ for all $y \in \mathbb{R}^n$.

Convex programs

A **convex program** can be written as follows.

Convex program

$$\begin{aligned} &\inf \;\; f_0(x) \\ \text{s.t. } & f_i(x) \leqslant 0 \text{ for } 1 \leqslant i \leqslant m \\ & h_j(x) = 0 \text{ for } 1 \leqslant j \leqslant p \end{aligned}$$

- f_i is convex for $i = 0, \dots, m$
- h_j is convex for $j = 1, \dots, p$

Remark: The domain of the problem is $D := \left(\bigcap_{i=0}^{m} \operatorname{dom} f_i\right) \cap \left(\bigcap_{j=1}^{p} \operatorname{dom} h_j\right)$, which is a convex set $\Rightarrow$ Roughly speaking, this makes the problem tractable.

Question: Can we define a dual program? How to give a lower bound?

Dual programs I

Idea: "move the constraints into the objective function"

The **Lagrangian** associated with the problem is

$$L(x, \lambda, \mu) := f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \mu_j h_j(x),$$

where the λ_is and μ_js are called **Lagrangian multipliers**, and $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^p$ are called the **dual variables**.

The **Lagrangian dual function** is the min value of the Lagrangian over x,

$$g(\lambda, \mu) := \inf_x L(x, \lambda, \mu).$$

Dual programs II

Let OPT_P denote the optimum value of the primal problem, and let $\hat{x}$ be an arbitrary feasible solution. Furthermore, assume that $\lambda \geqslant 0$. Then

$$\begin{aligned} g(\lambda, \mu) &\leqslant f_0(\hat{x}) + \sum_{i=1}^{m} \lambda_i f_i(\hat{x}) + \sum_{j=1}^{p} \mu_j h_j(\hat{x}) \\ &\leqslant f_0(\hat{x}), \end{aligned}$$

hence $g(\lambda, \mu) \leqslant \inf_{x \text{ feasible}} f_0(x) = OPT_P$.

Conclusion:

- This gives a lower bound when $\lambda \geqslant 0$ and $g(\lambda, \mu) > -\infty \Rightarrow$ Such a pair λ, μ is called **dual feasible**.

Weak duality

The goal is to get the best lower bound on OPT_P using the Lagrangian dual.

The **dual program** is thus defined as

Dual program

$$\begin{aligned} \max \;\; & g(\lambda, \mu) \\ \text{s.t.} \;\; & \lambda \geqslant 0 \end{aligned}$$

Let OPT_D denote the optimal value of the dual. Then **weak duality** holds by construction, that is, $OPT_D \leqslant OPT_P$.

Remarks:

- The dual program is always convex, regardeless of the primal.
- That is, for any primal program (even though non-convex), we can always write a convex program that gives a lower bound on the primal objective value.

Strong duality

Question: Does $OPT_D = OPT_P$ always holds?

Answer: Unfortunately **NOT**. But!

The **Slater's condition** requires that there is $x \in \operatorname{rel\,int}(D)$ such that $f_i(x) < 0$ for $1 \leqslant i \leqslant m$ and $h_j(x) = 0$ for $1 \leqslant j \leqslant p$.

(That is, the exists an **interior** point in the domain, which is a feasible solution, and satisfies the non-affine inequality constraints **strictly**.)

Strong duality

If Slater's condition holds, then $OPT_D = OPT_P$.

Complementary slackness

Assume that $OPT_D = OPT_P$. Let x^* be a primal, λ^*, μ^* be dual optimal solutions. Then

$$\begin{aligned} f_0(x^*) &= g(\lambda^*, \mu^*) \\ &= \inf_x L(x, \lambda^*, \mu^*) \\ &\leqslant L(x^*, \lambda^*, \mu^*) \\ &= f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* f_i(x^*) + \sum_{j=1}^{p} \mu_j^* h_j(x^*) \\ &\leqslant f_0(x^*), \end{aligned}$$

as $\lambda \geqslant 0$ (dual feasible) and $f_i(x^*) \leqslant 0$, $h_j(x^*) = 0$ (primal feasible).

Therefore

- x^* is a minimizer of $L(x, \lambda^*, \mu^*)$, and
- $\lambda_i^* f_i(x^*) = 0$ for $1 \leqslant i \leqslant m$, called the **complementary slackness condition**, meaning that the non-zero pattern of λ_i^* and $f_i(x^*)$ must be complementary.

Karush-Kuhn-Tucker (KKT) conditions I

Assume that $f_0, f_1, \ldots, f_m, h_1, \ldots, h_p$ are all differentiable.

Since x^* minimize $L(x, \lambda^*, \mu^*)$ by the above, the gradient of L at x^* must be zero, that is,

$$\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{j=1}^{p} \mu_j^* \nabla h_j(x^*) = 0.$$

To sum up, the following are some necessary conditions for any pair of primal and dual optimal solutions.

Primal feasibility: $f_i(x^*) \leqslant 0$ for $1 \leqslant 1 \leqslant m$, $h_j(x^*) = 0$ for $1 \leqslant j \leqslant p$.

Dual feasibility: $\lambda_i^* \geqslant 0$ for $1 \leqslant i \leqslant m$.

Complementary slackness: $\lambda_i^* f_i(x^*) = 0$ for $1 \leqslant i \leqslant m$.

Lagrangian optimality: $\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{j=1}^{p} \mu_j^* \nabla h_j(x^*) = 0$.

This set of conditions is called the **KKT conditions**.

Karush-Kuhn-Tucker (KKT) conditions II

When the primal problem is convex, the KKT conditions are also **sufficient**!

$\Rightarrow$ Any x^*, λ^*, μ^* satisfying KKT must be primal and dual optimal solutions.

Reason: If the primal is convex, then$L(x, \lambda, \mu)$ is convex in x when λ, μ are fixed. Hence a local optimal solution is also a global optimal solution. More precisely:

$$\begin{aligned} g(\lambda^*, \mu^*) &= \inf_x L(x, \lambda^*, \mu^*) \\ &= L(x^*, \lambda^*, \mu^*) \\ &= f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) + \sum_{j=1}^p \mu_j^* h_j(x^*) \\ &= f_0(x^*). \end{aligned}$$

Summary: For a convex problem with differentiable functions, if Slater's condition is satisfied, then the KKT conditions are **necessary** and **sufficient** for optimality.

Reading assignment

N. Vishnoi. Algorithms for convex optimization.

- Chapter 1
- Chapter 2
- Chapter 3
- Chapter 4
- Chapter 5

L.C. Lau. Convexity and optimization.

- Lecture 1
- Lecture 2
- Lecture 3
- Lectures 4-5

Exercises

1. Is it true, that a set $K \subseteq \mathbb{R}^n$ is convex if and only if for any $x, y \in K$ we have $(x + y)/2 \in K$? (1pt)
2. Prove that for an arbitrary function $f : \mathbb{R}^n \to \mathbb{R}$, the conjugate function $f^*(y) = \sup\{y^T x - f(x) \mid x \in \operatorname{dom}(f)\}$ is convex. (1pt)
3. Verify the following statements. (5pts)
   - a. e^{ax} is convex on $\mathbb{R}$ for any $a \in \mathbb{R}$.
   - b. x^a is convex on $\mathbb{R}_{>0}$ when $a \geqslant 1$ or $a \leqslant 0$, otherwise it is concave.
   - c. $\log x$ is concave on $\mathbb{R}_{>0}$.
   - d. $x \log x$ is convex on $\mathbb{R}_{>0}$.
   - e. $\max\{x_1, \ldots, x_n\}$ is convex on $\mathbb{R}^n$.

Exercises

4. Consider the optimization problem

$$\begin{aligned} \textit{minimize} \quad & x^2 + 1 \\ \textit{subject to} \quad & (x-2)(x-4) \leqslant 0 \end{aligned}$$

with variable $x \in \mathbb{R}$.

a. *Analysis of primal problem.* Give the feasible set, the optimal value, and the optimal solution. (1pt)
b. *Lagrangian and dual function.* Plot the objective $x^2 + 1$ versus x On the same plot, show the feasible set, optimal point and value, and plot the Lagrangian $L(x, \lambda)$ versus x for a few positive values of λ. Verify the lower bound property, that is, $y^* \geqslant \inf_x L(x, \lambda)$ for $\lambda \geqslant 0$. Derive and sketch the Lagrange dual function g. (2pts)
c. *Lagrange dual problem.* State the dual problem, and verify that it is a concave maximization problem. Find the dual optimal value and dual optimal solution λ^*. Does strong duality hold? (2pts)