Course: Continuous Optimization

Contents.

1. **Introduction**
2. **Convex Analysis**
3. **Constrained Convex Optimization**
4. **Duality in Convex Optimization**
5. **Constrained nonlinear (nonconvex) Optimization**

———————————- *7 lectures by Georg Still*

6. **Conic programming**

——————————— *5 lectures by Peter Dickinson*

Version: 21-09-2015 (Georg Still)

We make use of

- **Script:** ***de Klerk/Roos/Terlaky, Optimization***

on: http://wwwhome.math.utwente.nl/ ∼stillgj/conopt/

We refer to this script *e.g.* by [KRT, Th.4.3]

- **Book: *Faigle/Kern/Still, Algorithmic principles of Mathematical Programming.***

on: *http://wwwhome.math.utwente.nl/∼stillgj/priv/*

We refer to this book *e.g.* by [FKS, Th.4.3]

- **Lecture sheets (on the home-page above)**

Material for the lectures:

For Chapter 1-4 of the course: Chapter 0-3 of [KRT]

For Chapter 5 of the course: Chapter 12 of [FKS]

General optimization problem

$$P: \quad \left.\begin{array}{r}\inf \\ \min\end{array}\right\} f(x) \quad \text{s.t.} \quad x \in \mathcal{F}$$

Notation:

- $\mathcal{F} \subset \mathbb{R}^n$, feasible set
- $f : \mathcal{F} \to \mathbb{R}$, objective function

A point $\overline{x} \in \mathcal{F}$ is called:

- *global minimizer* of f on $\mathcal{F}$ if:

$$f(x) \geq f(\overline{x}) \quad \forall x \in \mathcal{F}$$

- *local minimizer* of f on $\mathcal{F}$ if with some $\varepsilon > 0$:

$$f(x) \geq f(\overline{x}) \quad \forall x \in \mathcal{F},\ \|x - \overline{x}\| < \varepsilon$$

- *strict* local minimizer if:

$$f(x) > f(\overline{x}) \quad \forall \underline{\overline{x} \neq x} \in \mathcal{F}, \|x - \overline{x}\| < \varepsilon$$

Unconstrained/Constrained Optimization

$$P: \quad \min\ f(x) \quad \text{s.t.} \quad x \in \mathcal{F}$$

- P is an *unconstrained* problem if $\mathcal{F}$ is open (in particular if $\mathcal{F} = \mathbb{R}^n$)
- P is a *constrained* problem if $\mathcal{F} \neq \mathbb{R}^n$ is closed. Often, $\mathcal{F}$ is given by *equality- and inequality constraints*:
  $$\mathcal{F} = \{x \in \mathbb{R}^n \mid h_i(x) = 0,\ i \in I,\ \ g_j(x) \leq 0,\ j \in J\}$$
  with $h_i, g_j(x) \in C(\mathbb{R}^n, \mathbb{R})$

rough Classification: P is called

- *linear* if f, h_i, g_j are (affine) linear
- *convex* if f is a convex function and $\mathcal{F}$ is a convex set.
- *nonlinear* if the problem functions are (nonconvex) nonlinear.

As a rule: Special subclasses, e.g., linear or (some) convex problems allow *efficient solution methods*.

In general: To compute a *global minimizer* in nonlinear optimization is *"very difficult"*

---

min $\longleftrightarrow$ max: $\overline{x}$ is max of f on $\mathcal{F}$ iff $\overline{x}$ is min of $-f$ on $\mathcal{F}$ and

$$\max_{x \in \mathcal{F}} f(x) = -\min_{x \in \mathcal{F}} \{-f(x)\}$$

<u>**Ex.1.1**</u> *[KRT, Ex0.1]* (*from Euclid's book ≈300 BC*)
In a given triangle ABC find an inscribed parallelogram ADEF of max area.

Show that Euclid's problem is: $P: \quad \max_{0<x<b} \frac{H}{b}(b-x)x$
See [KRT] for a sketch and the definition of ***H*****,*****b****.*

<u>**Tartaglia's problem**</u> (*Niccolo Tartaglia, 1500-1557)*
How to divide the number **8** ***into two parts such that the result of multiplying the product of the parts by their difference is maximal?***

This leads to (check):

$$P: \quad \max \; x_1 x_2 (x_1 - x_2) \quad \text{s.t.} \quad x_1 + x_2 = 8, \; 0 \le x_2 < x_1$$

Answer: $x_{1,2} = 4 \pm \frac{4}{\sqrt{3}}$ **(see also [KRT, Ex2.4]).**

Keplers problem

(*in "New solid geometry of wine barrals" (1615)*)

Given a sphere (of radius R), inscribe a cilinder of maximal volume.

2.1 Convex sets, convex functions

Def.2.1 Given $x^1, x^2 \in \mathbb{R}^n$ and $0 \le \lambda \le 1$. Then the point

$$x = \lambda x^1 + (1-\lambda)x^2$$

is a *convex combination* of x^1, x^2.
The set $\mathcal{C} \subset \mathbb{R}^n$ is called convex, if all convex combinations of any two points $x^1, x^2 \in \mathcal{C}$ are in $\mathcal{C}$.

Def.2.2 $f : \mathcal{C} \to \mathbb{R}$, defined on a convex set $\mathcal{C} \subset \mathbb{R}^n$ is called convex if for all $x^1, x^2 \in \mathcal{C}$ and $0 \le \lambda \le 1$ one has

$$f(\lambda x^1 + (1-\lambda)x^2) \le \lambda f(x^1) + (1-\lambda)f(x^2)$$

and called strictly convex if for all $x^1 \neq x^2 \in \mathcal{C}$ and $0 < \lambda < 1$:

$$f(\lambda x^1 + (1-\lambda)x^2) < \lambda f(x^1) + (1-\lambda)f(x^2)$$

Examples: **The following functions are convex:**

- $f(x) = x^2,\ x^4,\ e^x, -\ln x$
- any norm $f(x) = \|x\|$ on $\mathbb{R}^n$
- linear functions: $f(x) = a^T x + b, \quad x = (x_1, \ldots, x_n)$

Ex.2.1 **Show the convexity of the functions**

$$f(x) = \|x\|, \text{ and } \quad f(x) = a^T x + b \quad \text{for } x \in \mathbb{R}^n.$$

Recall A matrix $Q \in \mathbb{R}^{n \times n}$ is called *positive semidefinite (psd)* (notation $Q \succeq 0$) if Q is symmetric (i.e., $Q = Q^T$) and

$$x^T Q x \geq 0 \quad \forall x \in \mathbb{R}^n$$

and *positive definite (pd)* (not. $Q \succ 0$) if $Q = Q^T$ and

$$x^T Q x > 0 \quad \forall x \in \mathbb{R}^n, x \neq 0 .$$

Quadratic functions

Ex.2.2 Let be given the quadratic function on $\mathbb{R}^n$,

$$f(x) = x^T Q x + c^T x + d,$$

with symmetric matrix Q, $c \in \mathbb{R}^n$, $d \in \mathbb{R}$. Show that f is convex iff Q is psd.

Def.2.3 Given $x^i \in \mathbb{R}^n$, $i = 1, \ldots, k$, a combination

- $x = \sum_{i=1}^k \lambda_i x^i$ with $\lambda_i \in \mathbb{R}$, $\sum_{i=1}^k \lambda_i = 1$ is called an affine combination and
- $x = \sum_{i=1}^k \lambda_i x^i$ with $\underline{\lambda_i \geq 0}$, $\sum_{i=1}^k \lambda_i = 1$ is called a convex combination of the elements x^i.

Let $\mathcal{S} \subset \mathbb{R}^n$ be an arbitrary set. The set $\mathrm{aff}\,(\mathcal{S})$ defined by

$$\begin{aligned} \mathrm{aff}\,(\mathcal{S}) \quad := \quad \{x = \sum_{i=1}^k \lambda_i x^i \mid \; & x^i \in \mathcal{S}, \\ & \underline{\lambda_i \in \mathbb{R}}, \; \sum_{i=1}^k \lambda_i = 1, \; k \geq 1\} \end{aligned}$$

is called the ***affine hull*** of the set $\mathcal{S}$ (*set of affine combinations of* $\mathcal{S}$).

The set conv $(\mathcal{S})$,

$$\operatorname{conv}(\mathcal{S}) := \{x = \sum_{i=1}^{k} \lambda_i x^i \mid x^i \in \mathcal{S}, \quad \lambda_i \geq 0, \ \sum_{i=1}^{k} \lambda_i = 1, \ k \geq 1\}$$

is called the *convex hull* of the set $\mathcal{S}$ (*set of convex combinations of $\mathcal{S}$*).

Def.2.4 **A set of the form $x_0 + V$ with $x_0 \in \mathbb{R}^n$ and V a linear subspace of $\mathbb{R}^n$ is called an *affine space* with dimension given by dim(V).**

Ex. 2.3 **Show that for $\mathcal{S} \subset \mathbb{R}^n$ the set aff $(\mathcal{S})$ is the smallest affine space containing $\mathcal{S}$.**

Lemma 2.5 (*Jensen's inequality*, **[KRT,L.1.39]**)
Let $\mathcal{C} \subseteq \mathbb{R}^n$ be a convex set and let $f : \mathcal{C} \to \mathbb{R}$ be a convex function. Let the points $x^1, \cdots, x^k \in \mathcal{C}$ be given and let $\lambda^1, \cdots, \lambda^k \geq 0$ be such that $\sum_{i=1}^k \lambda^i = 1$, $k \geq 2$. Then

$$\sum_{i=1}^k \lambda^i x^i \in \mathcal{C} \quad \text{and} \quad f\Big(\sum_{i=1}^k \lambda^i x^i\Big) \leq \sum_{i=1}^k \lambda^i f(x^i).$$

Proof by induction wrt. k as an exercise

Ex. 2.4 **Show that for $\mathcal{S} \subset \mathbb{R}^n$ the set $\mathrm{conv}\,(\mathcal{S})$ is the smallest convex set containing $\mathcal{S}$.**
Ex. 2.5 **Prove [KRT, L1.40].**
Ex. 2.6 **Prove [KRT, L1.42].**

Def. 2.6 **The point $\overline{x} \in \mathcal{C}$ is an *extreme point* of the convex set $\mathcal{C}$ if there cannot exist $x^1 \neq x^2, x^1, x^2 \in \mathcal{C}$ and $0 < \lambda < 1$ such that**

$$\overline{x} = \lambda x^1 + (1 - \lambda) x^2$$

Th.2.7 [*Krein–Milman Theorem*], [KRT,Th.1.19]
Let $\mathcal{C} \subset \mathbb{R}^n$ be a compact convex set. Then $\mathcal{C}$ is the convex hull of its extreme points.

Proof. *See file Extraproofs on CO-site*

Ex.2.7 *(Max-value of a convex function is attained at an extreme point)*
Consider with compact, convex $\mathcal{F} \subset \mathbb{R}^n$ and convex continuous function f the max problem:
$(P) \quad \max f(x) \quad \text{s.t.} \quad x \in \mathcal{F}$.
Show that the maximum value of (P) is attained (also) at an extreme point of $\mathcal{F}$.

Def. 2.8 Let $\mathcal{C} \subset \mathbb{R}^n$ be a convex set. The point $x \in \mathcal{C}$ is in the *relative interior* of $\mathcal{C}$ if for any $\overline{x} \in \mathcal{C}$ there exists $\tilde{x} \in \mathcal{C}$ and $0 < \lambda < 1$ such that $x = \lambda \overline{x} + (1 - \lambda)\tilde{x}$.
The set of points in the relative interior of the set $\mathcal{C}$ will be denoted by $\mathcal{C}^0$.

By definition: $\mathcal{C}^0 \subseteq \mathcal{C}$.

L.2.9 *[KRT,L.1.34,1.36]*

Let $\mathcal{C} \subset \mathbb{R}^n$ be convex and nonempty. Then $\mathcal{C}^0$ is nonempty, convex and $(\mathcal{C}^0)^0 = \mathcal{C}^0$.

Properties of convex functions

Th.2.10, *[KRT,L.1.37]*

Let f be a convex function defined on the convex set $\mathcal{C}$. Then f is continuous on the relative interior $\mathcal{C}^0$ of $\mathcal{C}$.

<u>Ex.2.8</u> (*A convex function may have “jumps” at boundary points***)**
Find a function f which is convex on $[-1, 1]$ but not continuous on $[-1, 1]$

(*It must however be continuous on* $\mathcal{C}^0 = (-1, 1)$**).**

Recall: Gradient and Hessian of $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^1$ or C^2

- The *gradient* of f is: $\nabla f(x) := \left(\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_n}\right)^T$
- The *Hessian matrix* (or shortly *Hessian*) $\nabla^2 f(x)$ of $f \in C^2$ at a point x:

$$\nabla^2 f(x) = \left(\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right)_{i,j=1,\cdots,n}$$

Taylor expansion around $\overline{x}$ of order 2

Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be C^2 near $\overline{x}$. Then for $h \in \mathbb{R}^n$ ($\|h\|$ *small*):

$$f(\overline{x} + h) = f(\overline{x}) + \nabla f(\overline{x})^T h + \frac{1}{2} h^T \nabla^2 f(\overline{x} + \alpha h) h$$

for some $\alpha = \alpha(h) \in [0, 1]$. **Alternative form:**

$$f(\overline{x} + h) = f(\overline{x}) + \nabla f(\overline{x})^T h + \frac{1}{2} h^T \nabla^2 f(\overline{x}) h + o(\|h\|^2)$$

Ex.2.9 For $f(x) = \frac{1}{2}x^T Qx + c^T x$ with symmetric Q and $c \in \mathbb{R}^n$ show: $\nabla f(x) = Qx + c$ and $\nabla^2 f(x) = Q$

Characterizations of convex C^1, C^2-functions

L.2.11 *[KRT,L. 1.49]*

Let f be a C^1 function on the open convex set $\mathcal{C} \subseteq \mathbb{R}^n$.
Equivalent statements:

1. The function f is convex on $\mathcal{C}$.
2. For any $x, \overline{x} \in \mathcal{C}$ one has

$$f(x) \geq f(\overline{x}) + \nabla f(\overline{x})^T (x - \overline{x}).$$

L.2.12, *[KRT,L.1.50]*

Let f be a C^2 function on the open convex set $\mathcal{C} \subseteq \mathbb{R}^n$. The function f is convex iff its Hessian $\nabla^2 f(x)$ is psd for all $x \in \mathcal{C}$.

Ex. 2.10 Let f be a C^2 function on the open convex set $\mathcal{C}$. Then f is strictly convex if its Hessian $\nabla^2 f(x)$ is positive definite for all $x \in \mathcal{C}$.

Remark *The converse is not true:* $f(x) = x^4$ is strictly convex on $\mathbb{R}$, but $f''(0) = 0$.

Ex.2.11 The function $f(x) = e^x$ is convex on $\mathbb{R}$ and

$$e^x \geq 1 + x \quad \forall x \in \mathbb{R}$$

Ex.2.12 Show that the function $f(x) = -\log x$ is convex for $x > 0$. Moreover for $a_i \geq 0$, $i = 1, \ldots, n$ we have

$$(a_1 \cdots a_n)^{1/n} \leq \frac{a_1 + \ldots + a_n}{n}$$

The following is easy to prove *(Exercise!)*: *If $g : \mathbb{R}^n \to \mathbb{R}$ is convex then for any $\alpha \in \mathbb{R}$ the lower level set*

$$\mathcal{C}_\alpha := \{x \in \mathbb{R}^n \mid g(x) \leq \alpha\} \text{ is (closed) convex (possibly empty)}$$

The next exercise shows the converse: Any convex (bounded) set is given as a lower level set of a convex function.

Ex. 2.13 ***[gauge-function g]***

Let $\mathcal{C} \subset \mathbb{R}^n$ be a bounded convex set with $0 \in \text{int}(\mathcal{C})$. We define the *gauge function* g for $\mathcal{C}$ by:

$$\boldsymbol{g(x) := \inf\{\lambda \geq 0 \mid x \in \lambda\mathcal{C}\}} = \inf\{\lambda > 0 \mid \frac{x}{\lambda} \in \mathcal{C}\} \text{ for } \boldsymbol{x \in \mathbb{R}^n}$$

1. **Show** $g(x) \geq 0, \ \forall x \in \mathbb{R}^n$ **and** $g(x) = 0$ **iff** $x = 0$**.**
2. **For all** $\sigma \geq 0$**,** $0 \leq \rho \leq 1$ **and** $x, y \in \mathbb{R}^n$ **we have**

$$\boldsymbol{g(\sigma x) = \sigma g(x)}\ , \quad \boldsymbol{g(\rho x + (1-\rho)y) \leq \rho g(x) + (1-\rho)g(y)}\ .$$

3. **We have:** $x \in \text{cl}(\mathcal{C})$ **iff** $g(x) \leq 1$ **and** $x \in \text{bd}(\mathcal{C})$ **iff** $g(x) = 1$

Note that $g(x)$ defines a norm. So each compact convex set is defined as the unit ball of some norm g.