Course: Continuous Optimization

Contents.

Version: 15-09-2011 (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 0-3 of the course: Chapter 0-3 of [KRT]

For Chapter 4-6 of the course: Chapter 11,12 of [FKS]

In the Chapter0,1 we follow the exposition in [KRT].

0.1 general optimization problem

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

Not.:

- $\mathcal{F} \subset \mathbb{R}^n$, 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 $\mathcal{F} = \mathbb{R}^n$)
- 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 $h_i, g_j(x) \in C(\mathbb{R}^n, \mathbb{R})$

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*.

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

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

Ex. (*from Euclid's book ≈300 BC*) *In a given triangle ABC find an inscribed parallelogram ADEF of max area.*

Ex.0.1 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.

0.3.1 Tartaglia's problem (*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.

Read also Steiner's problem and the examples in [KRT, Sect.0.4.].

1.1 Convex sets, convex functions

Def.1.1 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.1.4 $f : \mathcal{C} \to \mathbb{R}$, 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

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

Def.1.5 Let $f : \mathcal{C} \to \mathbb{R}$ with $\mathcal{C} \subset \mathbb{R}^n$. The set in $\mathbb{R}^{n+1}$

$$\text{epi}\,(f) := \{(x, \tau) \mid f(x) \le \tau,\ x \in \mathcal{C},\ \tau \in \mathbb{R}\}$$

is called the *epigraph of f*.

Ex.1.5 Show: A function is convex if and only if its epigraph is a convex set.

Ex.1.6 Let be given the quadratic function on $\mathbb{R}^n$,

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

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

Recall A symmetric matrix $Q \in \mathbb{R}^{n \times n}$ (i.e., $Q = Q^T$) is called *positive semidefinite (psd)* (not. $Q \succeq 0$) if

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

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

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

Def.1.8 Let $\mathcal{S} \subset \mathbb{R}^n$ be an arbitrary set. The set $\mathrm{conv}\,(\mathcal{S})$,

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

is called the ***convex hull*** of the set $\mathcal{S}$ (*set of convex combinations of* $\mathcal{S}$). The set $\mathrm{aff}\,(\mathcal{S})$ defined by

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

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

Def. 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. 1.15' Show that for $\mathcal{S} \subset \mathbb{R}^n$ the set $\mathrm{aff}\,(\mathcal{S})$ is the smallest affine space containing $\mathcal{S}$.

Def. 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

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

L.1.18 Let $\mathcal{C} \subset \mathbb{R}^n$ be a nonempty compact (closed and bounded) convex set. Then $\mathcal{C}$ has at least one extreme point.

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

Proof By using

- L.1.18 and the following
- **Th.2.23** Let $\emptyset \neq U \subset \mathbb{R}^n$ be convex and $w \notin U$. Then there exists a separating hyperplane $H = \{x \mid a^T x = \alpha\}, 0 \neq a \in \mathbb{R}^n, \alpha \in \mathbb{R}$ such that

$$a^T w \geq \alpha \geq a^T x \quad \forall x \in U$$

and $\alpha > a^T u_0$ for some $u_0 \in U$.

1.2.3 Relative interior of a convex set

Def. 1.32 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 all $\overline{x} \in \mathcal{C}$ there exists $\tilde{x} \in \mathcal{C}$ and $0 < \lambda < 1$ such that $x = \lambda\overline{x} + (1 - \lambda)\tilde{x}$. The set of relative interior points of the set $\mathcal{C}$ will be denoted by $\mathcal{C}^0$.

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

L.1.34 Let $\mathcal{C} \subset \mathbb{R}^n$ be a convex set. If $x \in \mathcal{C}^0$, $y \in \mathcal{C}$ and $0 < \lambda \leq 1$, then we have

$$z = \lambda x + (1 - \lambda)y \in \mathcal{C}^0.$$

Cor. 1.35 The relative interior $\mathcal{C}^0$ of a convex set $\mathcal{C} \subset \mathbb{R}^n$ is convex.

L. 1.36 Let $\mathcal{C} \subset \mathbb{R}^n$ be convex. Then $(\mathcal{C}^0)^0 = \mathcal{C}^0$. Moreover, if $\mathcal{C}$ is nonempty then its relative interior $\mathcal{C}^0$ is nonempty.

1.3.1 Basic properties of convex functions

L. 1.37 (*main theorem*) **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 $\mathcal{C}$.**

Example: **Find a function f which is convex but not continuous on $[-1, 1]$**
(but continuous on $\mathcal{C}^0 = (-1, 1)$).

L. 1.39 (*Jensen's inequality*)
Let f be a convex function defined on a convex set $\mathcal{C} \subseteq \mathbb{R}^n$. 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

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

L.1.40 and L.1.42 as Exercise

Level sets: Given $f : \mathcal{C} \to \mathbb{R}$ and $\alpha \in \mathbb{R}$, we define:

$$\begin{array}{rcll} \mathcal{D}_\alpha^= &=& \{x \in \mathcal{C} \mid f(x) = \alpha\} & : \quad \text{level set} \\ \mathcal{D}_\alpha &=& \{x \in \mathcal{C} \mid f(x) \leq \alpha\} & : \quad \text{lower level set} \end{array}$$

of f (with level α).

L. 1.44 If f is *convex*, then for each $\alpha \in \mathbb{R}$ the lower level set $\mathcal{D}_\alpha$ is a (possibly empty) convex set.

Level sets of convex quadratic functions

Let Q be positive definite, $c \in \mathbb{R}^n$. $f(x) := \frac{1}{2}x^T Qx$:

$$D_\alpha^= = \{x \in \mathbb{R}^n \;:\; \frac{1}{2}x^T Qx = \alpha\}$$

is an ellipsoid with center 0. Since

$$\begin{aligned} f(x) &:= \tfrac{1}{2}x^T Qx + c^T x \\ &= \tfrac{1}{2}(x + Q^{-1}c)^T Q(x + Q^{-1}c) - \tfrac{1}{2}c^T Q^{-1}c. \end{aligned}$$

the level set $\quad D_\alpha = \{x \in \mathbb{R}^n \;:\; f(x) = \alpha\}$

forms an ellipsoid with center $x_0 = -Q^{-1}c$

<ins>Recall:</ins> With positive definite Q and $\alpha > 0$

$$E = \{x \mid \frac{1}{2}(x - x_0)^T Q(x - x_0) = \alpha\}$$

represents an ellipsoid with center x_0.

1.3.2 On derivatives of convex functions $f : \mathbb{R}^n \to \mathbb{R}$

Def. 1.45 If the limit exists, the *directional derivative* $\delta f(x,s)$ of the function f, at a point x, and in the direction s, is defined as

$$\delta f(x,s) = \lim_{\lambda \to 0} \frac{f(x+\lambda s) - f(x)}{\lambda}.$$

If $s = e^i$, where e^i denotes the $i-$th unit vector, then $\delta f(x,e^i)$ is the partial derivative : $\frac{\partial f(x)}{\partial x_i} = \delta f(x,e^i)$.

The *gradient* of f is: $\nabla f(x) := \left(\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_n} \right)^T$

L. 1.46 If f is differentiable at the point x then for all $s \in \mathbb{R}^n$: $\quad \delta f(x,s) = \nabla f(x)^T s$.

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

Recall The ***Hessian matrix*** (or shortly ***Hessian***) $\nabla^2 f(x)$ of $f \in C^2$ at a point x:

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

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 $\alpha \in [0,1]$ or

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

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

L. 1.47 The function f is convex on $\mathcal{C} \subset \mathbb{R}^n$ iff for any $x \in \mathcal{C}$ and $x + s \in \mathcal{C}$

$$\phi(\lambda) := f(x + \lambda s) \quad \text{is convex on } [0, 1].$$

In words: A function is convex if and only if it is convex along any line segment in its domain.

Characterizations of convex functions

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
$$f(x) \geq f(\overline{x}) + \nabla f(\overline{x})^T (x - \overline{x}).$$
3. For any $x, x + s \in \mathcal{C}$ with $\phi(\lambda) := f(x + \lambda s)$ the derivative $\phi'(\lambda) = s^T \nabla f(x + \lambda s)$ is a monotonically non-decreasing function on $(0, 1)$.

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. 1.20 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 (PD) for all $x \in \mathcal{C}$.

Ex. 1.21 *The converse is not true:* $f(x) = x^4$ is strictly convex for all $x \in \mathbb{R}$, but $f''(0) = 0$.

Ex.1.22n The function $f(x) = e^x$ is convex on $\mathbb{R}$ and

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

Ex.1.23n The function $f(x) = -\log x$ is convex for $x > 0$. Show: For $a_i \geq 0, i = 1, \ldots, n$ we have

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