STK4400 – Week 1

K. R. Dahl & A. B. Huseby

Department of Mathematics
University of Oslo, Norway

STK4400: Convexity, optimization, and convex duality: Overview

Cover background theory in convexity, optimization, and convex duality.

Use this to characterise convex risk measures (via convex duality).

Compare two commonly used risk measures: Value-at-Risk and Conditional-Value-at-Risk.

Convexity and optimization: Overview

- Convexity theory: Convex sets, convex functions and properties of these
- Quasiconvexity
- Optimization theory
- Lagrange duality: A method for solving constrained optimization problems.
- Convex duality framework of Rockafellar: Very general (generalizes Lagrange duality). Can be used to rephrase and solve a large variety of optimization problems.

Basic convexity: Framework

Based on the presentation of convexity in Rockafellar, Hiriart-Urruty and Lemarèchal and Dahl.

Let X be a vector space.

An inner product $\langle \cdot, \cdot \rangle : X \times X \to \mathbb{R}$ is symmetric, linear in the first component and positive definite in the sense that $\langle x, x \rangle \geq 0$ for all $x \in X$, with equality if and only if $x = 0$.

Example: $X = \mathbb{R}^n$, $n \in \mathbb{N}$.

Essential definitions

Definition

(*i*) (Convex set) A set $C \subseteq X$ is called convex if $\lambda x_1 + (1-\lambda)x_2 \in C$ for all $x_1, x_2 \in C$ and $0 \leq \lambda \leq 1$.

(*ii*) (Convex combination) A convex combination of elements $x_1, x_2, \ldots, x_k$ in X is an element of the form $\sum_{i=1}^{k} \lambda_i x_i$ where $\sum_{i=1}^{k} \lambda_i = 1$ and $\lambda_i \geq 0$ for all $i = 1, \ldots, k$.

(*iii*) (Convex hull, $\mathrm{conv}(\cdot)$) Let $A \subseteq X$ be a set. The convex hull of A, denoted $\mathrm{conv}(A)$ is the set of all convex combinations of elements of A.

Core definitions

K. R. Dahl & A. B. Huseby

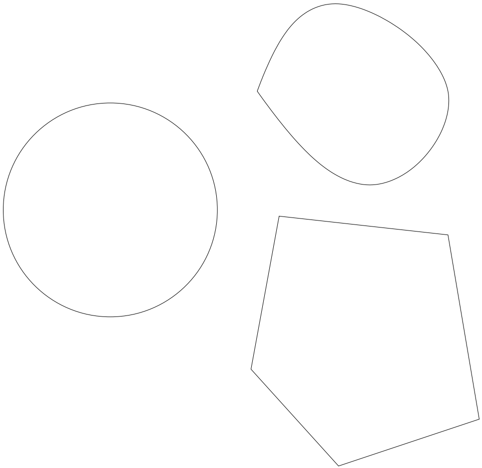
Definition

(*iv*) (Extreme points) Let $C \subseteq X$ be a convex set. An extreme point of C is a point that cannot be written as a convex combination of any other points than itself. That is: $e \in C$ is an extreme point for C if $\lambda x + (1-\lambda)y = e$ for some $x, y \in C$ implies $x = y = e$.

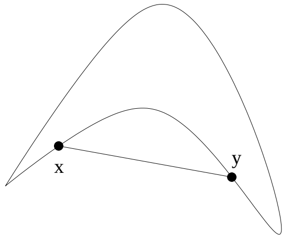
(*v*) (Hyperplane) $H \subset X$ is called a hyperplane if it is of the form $H = \{x \in X : \langle a, x \rangle = \alpha\}$ for some nonzero vector $a \in X$ and some real number α.

(*vi*) (Halfspace) A hyperplane H divides X into two sets $H^+ = \{x \in X : \langle a, x \rangle \geq \alpha\}$ and $H^- = \{x \in X : \langle a, x \rangle \leq \alpha\}$, these sets intersect in H. These sets are called halfspaces.

Some convex sets in the plane



A non-convex set



Hyperplane theorems in $\mathbb{R}^n$

Purpose: Will be used later in connection to environmental contours.

Only consider $\mathbb{R}^n$ for these theorems: Most generalize to an arbitrary real inner product space X.

Any hyperplane, Π in $\mathbb{R}^n$ can be written:

$$\Pi = \{\boldsymbol{x} : \boldsymbol{c}'\boldsymbol{x} = d\},$$

where $\boldsymbol{c} \in \mathbb{R}^n$ is a normal vector to Π and $d \in \mathbb{R}$.

Supporting hyperplane

Let

$$\Pi^- = \{\boldsymbol{x} : \boldsymbol{c}'\boldsymbol{x} \leq d\}$$

and

$$\Pi^+ = \{\boldsymbol{x} : \boldsymbol{c}'\boldsymbol{x} \geq d\}$$

denote the two half-spaces bounded by Π.

Let $\mathcal{S} \subseteq \mathbb{R}^n$.

A supporting hyperplane of $\mathcal{S}$, is a hyperplane Π such that we either have $\mathcal{S} \subseteq \Pi^-$ or $\mathcal{S} \subseteq \Pi^+$, and such that $\Pi \cap \partial\mathcal{S} \neq \emptyset$.

Supporting half-space

If Π is a supporting hyperplane of the set $\mathcal{S}$, and $\mathcal{S} \subseteq \Pi^-$, we say that Π^+ is a supporting half-space of $\mathcal{S}$.

Observe: If Π^+ is a supporting half-space of $\mathcal{S}$,

$$\Pi^+ \cap \mathcal{S} \subseteq \partial\mathcal{S}.$$

Introduce the notation:

$$\mathcal{P}(\mathcal{S}) = \text{The family of supporting half-spaces of } \mathcal{S}.$$

Projections

For a given nonempty set $\mathcal{S} \subseteq \mathbb{R}^n$ and a vector $\boldsymbol{x}_0 \notin \mathcal{S}$, the vector $\boldsymbol{x}^* \in \mathcal{S}$ is said to be the projection of $\boldsymbol{x}_0$ onto $\mathcal{S}$ if $\boldsymbol{x}^*$ is the point in $\mathcal{S}$ which is closest to $\boldsymbol{x}_0$.

In general the projection $\boldsymbol{x}^*$ may neither exist nor be unique.

Theorem (Projection)

Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a closed convex set, and let $\boldsymbol{x}_0 \notin \mathcal{S}$. Then the following holds true:

- *There exists a unique solution to the projection problem*
- *A vector $\boldsymbol{x}^* \in \mathcal{S}$ is the projection of $\boldsymbol{x}_0$ onto $\mathcal{S}$ if and only if:*

$$(\boldsymbol{x}^* - \boldsymbol{x}_0)'(\boldsymbol{x} - \boldsymbol{x}^*) \geq 0 \text{ for all } \boldsymbol{x} \in \mathcal{S}.$$

Illustration of the projection theorem in $\mathbb{R}^2$

K. R. Dahl & A. B. Huseby

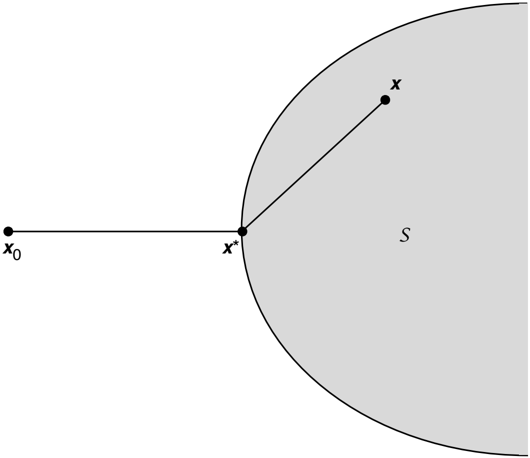


Figure: The point $\boldsymbol{x}^*$ is the projection of $\boldsymbol{x}_0$ onto the closed convex set $\mathcal{S}$.

If $\boldsymbol{x} \in \mathcal{S}$, and θ is the angle between $(\boldsymbol{x}^* - \boldsymbol{x}_0)$ and $(\boldsymbol{x} - \boldsymbol{x}^*)$, then we must have $\theta \in [-\pi/2, \pi/2]$. This holds if and only if:

$$(\boldsymbol{x}^* - \boldsymbol{x}_0)'(\boldsymbol{x} - \boldsymbol{x}^*) \geq 0 \text{ for all } \boldsymbol{x} \in \mathcal{S}.$$

Theorem (Projection hyperplane)

Let $\mathcal{S} \in \mathbb{R}^n$ be a closed convex set, and assume that $\mathbf{x}_0 \notin \mathcal{S}$. Then there exists a supporting hyperplane $\Pi = \{\boldsymbol{x} : \mathbf{c}'\boldsymbol{x} = d\}$ of $\mathcal{S}$ such that:

$$\mathbf{c}'\boldsymbol{x} \leq d \text{ for all } \boldsymbol{x} \in \mathcal{S}, \quad \text{and} \quad \mathbf{c}'\mathbf{x}_0 > d.$$

Proof of the projection hyperplane theorem

Proof: Since $\mathcal{S}$ is a closed convex set, it follows by the projection theorem that the projection of $\boldsymbol{x}_0$ onto $\mathcal{S}$, denoted $\boldsymbol{x}^*$ exists and satisfies:

$$(\boldsymbol{x}^* - \boldsymbol{x}_0)'(\boldsymbol{x} - \boldsymbol{x}^*) \geq 0 \quad \text{for all } \boldsymbol{x} \in \mathcal{S}. \tag{1}$$

Now, we let $\boldsymbol{c} = (\boldsymbol{x}_0 - \boldsymbol{x}^*)$, and $d = \boldsymbol{c}'\boldsymbol{x}^*$. Then (1) can be written as:

$$\boldsymbol{c}'(\boldsymbol{x} - \boldsymbol{x}^*) \leq 0 \quad \text{for all } \boldsymbol{x} \in \mathcal{S}. \tag{2}$$

Proof of projection hyperplane theorem, ctd.

Hence, by (2) we have:

$$\mathbf{c}'\mathbf{x} \leq \mathbf{c}'\mathbf{x}^* = d \quad \text{for all } \mathbf{x} \in \mathcal{S}.$$

Thus, $\mathcal{S} \subseteq \Pi^- = \{\mathbf{x} : \mathbf{c}'\mathbf{x} \leq d\}$, and since $\mathbf{x}^* \in \mathcal{S} \cap \Pi$, Π is a supporting hyperplane of $\mathcal{S}$. Furthermore, we have:

$$\mathbf{c}'(\mathbf{x}_0 - \mathbf{x}^*) = (\mathbf{x}_0 - \mathbf{x}^*)'(\mathbf{x}_0 - \mathbf{x}^*) > 0.$$

Hence, it follows that:

$$\mathbf{c}'\mathbf{x}_0 > \mathbf{c}'\mathbf{x}^* = d.$$

This concludes the proof.

Supporting hyperplane theorem

K. R. Dahl & A. B. Huseby

Theorem (Supporting hyperplane)

Let $\mathcal{S} \in \mathbb{R}^n$ be a convex set, and assume that either $\boldsymbol{x}_0 \notin \mathcal{S}$ or $\boldsymbol{x}_0 \in \partial\mathcal{S}$. Then there exists a hyperplane Π such that $\mathcal{S} \subseteq \Pi^-$ and such that $\boldsymbol{x}_0 \in \Pi$. If $\boldsymbol{x}_0 \in \partial\mathcal{S}$, Π is a supporting hyperplane of $\mathcal{S}$.

Proof: The result follows by a similar argument as for the projection hyperplane theorem and is left as an exercise to the reader.

Separating hyperplanes

Let $\mathcal{S}, \mathcal{T} \subseteq \mathbb{R}^n$. A hyperplane Π separates $\mathcal{S}$ and $\mathcal{T}$ if either $\mathcal{S} \subseteq \Pi^-$ and $\mathcal{T} \subseteq \Pi^+$ or $\mathcal{S} \subseteq \Pi^+$ and $\mathcal{T} \subseteq \Pi^-$.

Theorem (Separating hyperplane)

Assume that $\mathcal{S}, \mathcal{T} \subseteq \mathbb{R}^n$ are convex, and that $\mathcal{S} \cap \mathcal{T} \subseteq \partial\mathcal{S}$. Then there exists a hyperplane Π separating $\mathcal{S}$ and $\mathcal{T}$ such that $\mathcal{S} \subseteq \Pi^-$ and $\mathcal{T} \subseteq \Pi^+$.

Proof of the separating hyperplane theorem

K. R. Dahl &
A. B. Huseby

Proof: We let $\boldsymbol{u}_0 = 0 \in \mathbb{R}^n$ and introduce the set:

$$\mathcal{U} = \{\boldsymbol{x} - \boldsymbol{y} : \boldsymbol{x} \in \mathcal{S}_o, \boldsymbol{y} \in \mathcal{T}\},$$

where $\mathcal{S}_o = \mathcal{S} \setminus \partial\mathcal{S}$ is the (convex) set of inner points in $\mathcal{S}$.

Proof of the separating hyperplane theorem, ctd.

K. R. Dahl & A. B. Huseby

We first argue that $\mathcal{U}$ is convex. To show this we must show that if $\boldsymbol{u}_1, \boldsymbol{u}_2 \in \mathcal{U}$, then $\alpha \boldsymbol{u}_1 + (1-\alpha)\boldsymbol{u}_2 \in \mathcal{U}$ for all $\alpha \in [0,1]$:

Since $\boldsymbol{u}_1, \boldsymbol{u}_2 \in \mathcal{U}$ there exists $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{S}_o$ and $\boldsymbol{y}_1, \boldsymbol{y}_2 \in \mathcal{T}$ such that:

$$\boldsymbol{u}_1 = \boldsymbol{x}_1 - \boldsymbol{y}_1 \quad \text{and} \quad \boldsymbol{u}_2 = \boldsymbol{x}_2 - \boldsymbol{y}_2.$$

Since $\mathcal{S}_o$ and $\mathcal{T}$ are convex, it follows that for any $\alpha \in [0,1]$, we have:

$$\alpha \boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2 \in \mathcal{S}_o \quad \text{and} \quad \alpha \boldsymbol{y}_1 + (1-\alpha)\boldsymbol{y}_2 \in \mathcal{T}$$

Proof of the separating hyperplane theorem, ctd.

Hence, we have:

$$\begin{aligned}\alpha \boldsymbol{u}_1 + (1-\alpha)\boldsymbol{u}_2 &= \alpha(\boldsymbol{x}_1 - \boldsymbol{y}_1) + (1-\alpha)(\boldsymbol{x}_2 - \boldsymbol{y}_2) \\ &= (\alpha \boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2) - (\alpha \boldsymbol{y}_1 + (1-\alpha)\boldsymbol{y}_2) \in \mathcal{U}.\end{aligned}$$

By the assumption that $\mathcal{S} \cap \mathcal{T} \subseteq \partial \mathcal{S}$ it follows that $\mathcal{S}_o$ and $\mathcal{T}$ do not have any element in common.
Hence, it follows that:

$$\boldsymbol{u} = \boldsymbol{x} - \boldsymbol{y} \neq 0, \text{ for all } \boldsymbol{x} \in \mathcal{S}_o \text{ and } \boldsymbol{y} \in \mathcal{T}.$$

Thus, we conclude that:

$$\boldsymbol{u}_0 = 0 \notin \mathcal{U}.$$

Proof of the separating hyperplane theorem, ctd.

K. R. Dahl & A. B. Huseby

Then, by the supporting hyperplane theorem there exists a hyperplane $\Pi_0 = \{\boldsymbol{x} : \boldsymbol{c}'\boldsymbol{x} = d_0\}$ such that $\mathcal{U} \subseteq \Pi_0^-$ and such that $\boldsymbol{u}_0 \in \Pi_0$.

In fact, $\boldsymbol{u}_0 \in \Pi_0$ implies that $\boldsymbol{c}'\boldsymbol{u}_0 = \boldsymbol{c}'0 = d_0$. Thus, $d_0 = 0$.

Since $\mathcal{U} \subseteq \Pi_0^-$, we have $\boldsymbol{c}'\boldsymbol{u} \leq d_0 = 0$ for all $\boldsymbol{u} \in \mathcal{U}$, implying that:

$$\boldsymbol{c}'(\boldsymbol{x} - \boldsymbol{y}) \leq 0 \text{ for all } \boldsymbol{x} \in \mathcal{S}_o \text{ and } \boldsymbol{y} \in \mathcal{T},$$

Proof of the separating hyperplane theorem, ctd.

K. R. Dahl & A. B. Huseby

or equivalently:

$$\boldsymbol{c}'\boldsymbol{x} \leq \boldsymbol{c}'\boldsymbol{y} \text{ for all } \boldsymbol{x} \in \mathcal{S}_0 \text{ and } \boldsymbol{y} \in \mathcal{T}. \tag{3}$$

We then let $d = \sup_{\boldsymbol{x} \in \mathcal{S}_o} \boldsymbol{c}'\boldsymbol{x}$. By the definition of d we have:

$$\boldsymbol{c}'\boldsymbol{x} \leq d, \text{ for all } \boldsymbol{x} \in \mathcal{S}_o. \tag{4}$$

If $\boldsymbol{x}_0 \in \partial\mathcal{S}$, there exists $\{\boldsymbol{x}_k\} \subseteq \mathcal{S}_o$ such that $\lim_{k\to\infty} \boldsymbol{x}_k = \boldsymbol{x}_0$, and so:

$$\boldsymbol{c}'\boldsymbol{x}_0 = \lim_{k\to\infty} \boldsymbol{c}'\boldsymbol{x}_k \leq d. \tag{5}$$

Proof of the separating hyperplane theorem, ctd.

By combining (4) and (5), we have:

$$\boldsymbol{c}'\boldsymbol{x} \leq d, \text{ for all } \boldsymbol{x} \in \mathcal{S}. \tag{6}$$

Moreover, by (3) and the definition of d it follows that we have:

$$\boldsymbol{c}'\boldsymbol{y} \geq d, \text{ for all } \boldsymbol{y} \in \mathcal{T}. \tag{7}$$

Hence, by letting $\Pi = \{\boldsymbol{x} : \boldsymbol{c}'\boldsymbol{x} = d\}$, we conclude that:

$$\mathcal{S} \subseteq \Pi^- \text{ and } \mathcal{T} \subseteq \Pi^+ \tag{8}$$

Hence, Π is a hyperplane separating the convex sets $\mathcal{S}$ and $\mathcal{T}$. This concludes the proof.

Separating hyperplane, supporting halfspace

K. R. Dahl & A. B. Huseby

In the proof of the separating hyperplane theorem, we defined $\Pi = \{\boldsymbol{x} : \boldsymbol{c}'\boldsymbol{x} = d\}$, where:

$$d = \sup_{\boldsymbol{x} \in \mathcal{S}_o} \boldsymbol{c}'\boldsymbol{x}.$$

This implies that there exists a sequence $\{\boldsymbol{x}_k\} \subseteq \mathcal{S}$ with limit $\boldsymbol{x}_0 = \lim_{k \to \infty} \boldsymbol{x}_k \in \partial\mathcal{S}$, such that $\boldsymbol{c}'\boldsymbol{x}_0 = d$.

Thus, $\boldsymbol{x}_0 \in \Pi \cap \partial\mathcal{S}$, implying that Π^+ is a supporting halfspace of $\mathcal{S}$.

Separating hyperplane, supporting halfspace theorem

Theorem (Separating hyperplane, supporting halfspace)

Assume that $\mathcal{S}, \mathcal{T} \subseteq \mathbb{R}^n$ are convex, and that $\mathcal{S} \cap \mathcal{T} \subseteq \partial \mathcal{S}$. Then there exists a hyperplane Π separating $\mathcal{S}$ and $\mathcal{T}$ such that $\mathcal{S} \subseteq \Pi^-$, $\mathcal{T} \subseteq \Pi^+$ and where $\Pi^+ \in \mathcal{P}(\mathcal{S})$.

Polyhedrons and polytopes

Polyhedrons and polytopes are subsets of $\mathbb{R}^n$.

They are useful because they are descriptions of the solution set of systems of linear inequalities. The following definitions are from Rockafellar.

Definition (Polyhedron)

A set $K \subseteq \mathbb{R}^n$ is called a polyhedron if it can be described as the intersection of finitely many closed half-spaces.

Hence, a polyhedron can be described as the solution set of a system of finitely many (non-strict) linear inequalities.

It is straightforward to show that a polyhedron is a convex set (do this!).

Polytopes

A (convex) polytope is a set of the following form:

Definition (Polytope)

A set $K \subseteq \mathbb{R}^n$ is called a (convex) polytope if it is the convex hull of finitely many points.

Clearly, all polytopes are convex since a convex hull is always convex.

Examples of (convex) polytopes in $\mathbb{R}^2$: Triangles, squares and hexagons.

Polytopes are compact

Actually, all polytopes in $\mathbb{R}^n$ are compact sets.

Lemma

Let $K \subseteq \mathbb{R}^n$ be a polytope. Then K is a compact set.

Proof: Since K is a polytope, it is the convex hull of finitely many points, say $K = \mathrm{conv}(\{k_1\}, ..., \{k_m\})$, so

$$K = \{\sum_{i=1}^{m} \lambda_i k_i : \sum_{i=1}^{m} \lambda_i = 1, \lambda_i \geq 0 \text{ for all } i = 1, ..., m\}.$$

Proof, ctd.

K. R. Dahl & A. B. Huseby

Consider the continuous function $f : \mathbb{R}^m \to \mathbb{R}^n$, $f(x_1, ..., x_m) = \sum_{i=1}^m x_i k_i$, and the compact set

$$S = \{(\lambda_1, ..., \lambda_m) : \sum_{i=1}^m \lambda_i = 1, \lambda_i \geq 0 \text{ for all } i = 1, ..., m\} \subseteq \mathbb{R}^m$$

(S is closed and bounded, hence compact in $\mathbb{R}^m$)
Then, since f is continuous and S is compact, $f(S) := \{x : x = f(s) \text{ for some } s \in S\} \subseteq \mathbb{R}^n$ is a compact set (see for example Munkres). But $f(S) = K$ from the definitions, and hence K is compact. This concludes the proof.

Polytopes are bounded polyhedrons

From Lemma 10, any polytope is a closed and bounded set, since compactness is equivalent to being closed and bounded in $\mathbb{R}^n$.

The following theorem connects the notion of polytope and polyhedron.

Theorem

A set $K \subseteq \mathbb{R}^n$ is a polytope if and only if it is a bounded polyhedron.

For a proof of this, see Ziegler.

Relative interior

K. R. Dahl & A. B. Huseby

Sometimes, one needs to consider what is called the relative interior of a set.

Definition (Relative interior, $\mathrm{rint}(\cdot)$)

Let $S \subseteq X$. $x \in S$ is a relative interior point of S if it is contained in some open set whose intersection with $\mathrm{aff}(S)$ is contained in S. $\mathrm{rint}(S)$ is the set of all relative interior points of S.

Here, $\mathrm{aff}(S)$ is the smallest affine set that contains S (where a set is affine if it contains any affine combination of its points; an affine combination is like a convex combination except the coefficients are allowed to be negative).

Convex cone

K. R. Dahl & A. B. Huseby

Another useful notion is that of a convex cone.

Definition (Convex cone)

$C \subseteq X$ is called a convex cone if for all $x, y \in C$ and all $\alpha, \beta \geq 0$:

$$\alpha x + \beta y \in C.$$

Properties of convex sets

From these definitions, we can derive some properties of convex sets.

Theorem (Properties of convex sets)

(i) *If* $\{C_j\}_{j \in J} \subseteq X$ *is an arbitrary family of convex sets, then the intersection* $\cap_{j \in J} C_j$ *is also a convex set.*

(ii) $\text{conv}(A)$ *is a convex set, and it is the smallest (set inclusion-wise) convex set containing* A*.*

(iii) *If* $C_1, C_2, \ldots, C_m \subseteq X$ *are convex sets, then the Cartesian product* $C_1 \times C_2 \times \ldots \times C_m$ *is also a convex set.*

(iv) *If* $C \subseteq X$ *is a convex set, then the interior of* C*,* $\text{int}(C)$*, the relative interior* $\text{rint}(C)$ *and the closure of* C*,* $\text{cl}(C)$*, are convex sets as well.*

The proof is left as an exercise.

The extended real numbers

K. R. Dahl & A. B. Huseby

Sometimes, one considers not just $\mathbb{R}$, but $\bar{\mathbb{R}}$, the extended real numbers.

Definition (The extended real numbers, $\bar{\mathbb{R}}$)

Let $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ denote the extended real numbers.

When working with the extended real numbers the following computational rules apply: $a - \infty = -\infty$, $a + \infty = \infty$, $\infty + \infty = \infty$, $-\infty - \infty = -\infty$ and $\infty - \infty$ is not defined.

Indicator function

The following function is often useful, in particular in optimization.

Definition (The indicator function for a set M, δ_M)

Let $M \subseteq X$ be a set. The indicator function for the set M, $\delta_M : X \to \bar{\mathbb{R}}$ is defined as

$$\delta_M(x) = \begin{cases} 0 & \text{if } x \in M \\ +\infty & \text{if } x \notin M. \end{cases}$$

Example of indicator function in optimization

K. R. Dahl & A. B. Huseby

Consider the constrained minimization problem

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & x \in M \end{array}$$

for some function $f : X \to \bar{\mathbb{R}}$ and some set $M \subseteq X$.

Transform into unconstrained minimization problem: Altering the objective function,

$$\min f(x) + \delta_M(x).$$

Same problem as before: The minimum cannot be achieved for $x \notin M$, because then $\delta_M = +\infty$, so the objective function is infinitely large as well.

Convex functions

V.I.D. (very important definition):

Definition (Convex function)

Let $C \subseteq X$ be a convex set. A function $f : C \to \mathbb{R}$ is called convex if the inequality

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) \tag{9}$$

holds for all $x, y \in C$ and every $0 \leq \lambda \leq 1$.

Alternative way of defining convex functions via epigraphs:

Definition (Epigraph, $\mathrm{epi}(\cdot)$)

Let $f : X \to \bar{\mathbb{R}}$ be a function. Then the epigraph of f is defined as $\mathrm{epi}(f) = \{(x, \alpha) : x \in X,\ \alpha \in \mathbb{R},\ \alpha \geq f(x)\}$.

Illustration of the epigraph

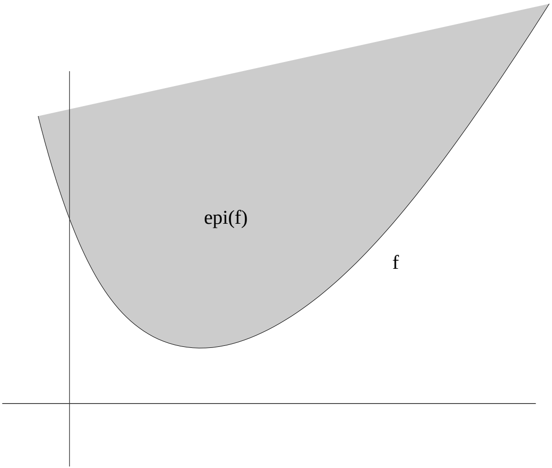


Figure: The epigraph of a function f.

Definition of convex functions via the epigraph

Definition (Convex function)

Let $A \subseteq X$. A function $f : A \to \bar{\mathbb{R}}$ is called convex if the epigraph of f is convex (as a subset of the vector space $X \times \mathbb{R}$).

Illustration of a convex function

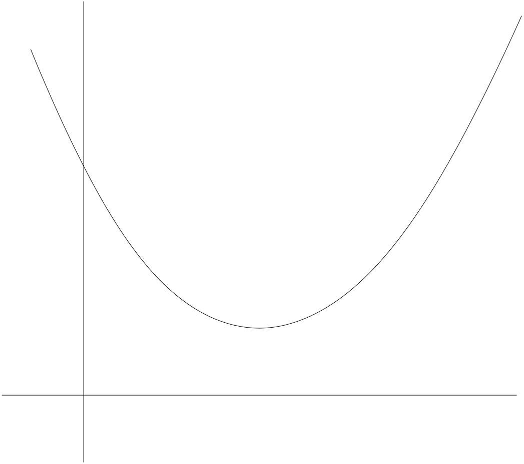


Figure: A convex function.

The two definitions of convex functions are equivalent

K. R. Dahl & A. B. Huseby

Theorem

Definitions 17 and 19 are equivalent if the set A in Definition 19 is convex (A must be convex in order for Definition 17 to make sense).

Proof that the definitions are equivalent

Proof:

17 ⇒ 19: Assume that f is a convex function according to Definition 17. Let $(x, a), (y, b) \in \mathrm{epi}(f)$ and let $\lambda \in [0, 1]$. Then

$$\lambda(x, a) + (1 - \lambda)(y, b) = (\lambda x + (1 - \lambda)y, \lambda a + (1 - \lambda)b).$$

But $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$ from Definition 17, so

$$\begin{aligned} f(\lambda x + (1 - \lambda)y) &\leq \lambda f(x) + (1 - \lambda)f(y) \\ &\leq \lambda a + (1 - \lambda)b. \end{aligned}$$

So $(\lambda x + (1 - \lambda)y, \lambda a + (1 - \lambda)b) \in \mathrm{epi}(f)$.

19 ⇒ 17: Same type of arguments, so omitted. This concludes the proof.

Concave function

Definition (Concave function)

A function g is concave if the function $f := -g$ is convex.

When minimizing a function, the points where it is infinitely large are uninteresting:

Definition (Effective domain, $\mathrm{dom}(\cdot)$)

Let $A \subseteq X$ and let $f : A \to \bar{\mathbb{R}}$ be a function. The effective domain of f is defined as $\mathrm{dom}(f) = \{x \in A : f(x) < +\infty\}$.

Proper functions

Definition (Proper function)

Let $A \subseteq X$ and let $f : A \to \bar{\mathbb{R}}$ be a function. f is called proper if $\mathrm{dom}(f) \neq \emptyset$ and $f(x) > -\infty$ for all $x \in A$.

For definitions of general topological terms, such as convergence, continuity and neighborhood, see any basic topology book, for instance Munkres.

Lower semi-continuity

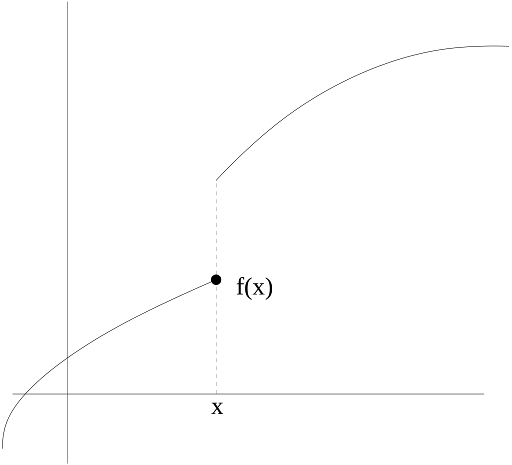
Definition (Lower semi-continuity, lsc)

Let $A \subseteq X$ be a set, and let $f : A \to \bar{\mathbb{R}}$ be a function.

f is called lower semi-continuous, lsc, at a point $x_0 \in A$ if for each $k \in \mathbb{R}$ such that $k < f(x_0)$ there exists a neighborhood U of x_0 such that $f(u) > k$ for all $u \in U$.

Equivalently: f is lower semi-continuous at x_0 if and only if $\liminf_{x \to x_0} f(x) \geq f(x_0)$.

A lower semi-continuous function f



Sublevel sets

Lower semi-continuity can be described via sublevel sets:

Definition (α-sublevel set of a function, $S_\alpha(f)$)

Let $f : X \to \bar{\mathbb{R}}$ be a function and let $\alpha \in \mathbb{R}$. The α-sublevel set of f, $S_\alpha(f)$, is defined as

$$S_\alpha(f) = \{x \in X : f(x) \leq \alpha\}.$$

Theorem

Let $f : X \to \bar{\mathbb{R}}$ be a function. Then, f is lower semi-continuous if and only if the sublevel sets $S_\alpha(f)$ are closed for all $\alpha \in \bar{\mathbb{R}}$.

Proof: The sublevel sets $S_\alpha(f) := \{x \in X : f(x) \leq \alpha\}$ are closed for all $\alpha \in \mathbb{R}$ iff. the complement sets $Y = X - S_\alpha(f) = \{x \in X : f(x) > \alpha\}$ are open for all α.

This happens iff. all $y \in Y$ are interior points, which is equivalent with that for each $y \in Y$ there is a neighborhood U such that $U \subseteq Y$, i.e. $f(U) > \alpha$.

This is the definition of f being lower semi-continuous at the point y.

Since this argument holds for all $y \in X$ (by choosing different α), f is lower semi-continuous.

Convex hull

Definition (Convex hull of a function, $\mathrm{co}(f)$)

Let $A \subseteq X$ be a set, and let $f : A \rightarrow \bar{\mathbb{R}}$ be a function. Then the convex hull of f is the (pointwise) largest convex function h such that $h(x) \leq f(x)$ for all $x \in A$.

Clearly, if f is a convex function $\mathrm{co}(f) = f$. One can define the lower semi-continuous hull, $\mathrm{lsc}(f)$ of a function f in a similar way.

Closure of a function

Definition (Closure of a function, $\mathrm{cl} f$)

Let $A \subseteq X$ be a set, and let $f : A \to \bar{\mathbb{R}}$ be a function. We define: $\mathrm{cl}(f)(x) = \mathrm{lsc}(f(x))$ for all $x \in A$ if $\mathrm{lsc}(f(x)) > -\infty$ $\forall$ $x \in X$ and $\mathrm{cl}(f)(x) = -\infty$ for all $x \in A$ if $\mathrm{lsc}(f(x')) = -\infty$ for some $x' \in A$.

We say that a function f is closed if $\mathrm{cl}(f) = f$. Hence, f is closed if it is lower semi-continuous and $f(x) > -\infty$ for all x or if $f(x) = -\infty$ for all x.

Properties of the indicator function

Theorem

Let $M \subseteq X$, and consider the indicator function for the set M, δ_M, as defined in Definition 16. Then, the following properties hold:

- *If $N \subseteq X$, then $M \subseteq N \iff \delta_N \leq \delta_M$.*
- *M is a convex set $\iff$ δ_M is a convex function.*
- *δ_M is lower semi-continuous $\iff$ M is a closed set.*

Proof of properties of the indicator function

K. R. Dahl & A. B. Huseby

Proof:

- From Definition 16: $\delta_N \leq \delta_M$ iff. (If $\delta_M(x) < +\infty$ then $\delta_N(x) < +\infty$) iff. $(x \in M \Rightarrow x \in N)$ iff. $M \subseteq N$.
- δ_M is convex if and only if $\delta_M(\lambda x + (1-\lambda)y) \leq \lambda\delta_M(x) + (1-\lambda)\delta_M(y)$ holds for all $0 \leq \lambda \leq 1$ and all $x, y \in X$ such that $\delta_M(x), \delta_M(y) < +\infty$, that is, for all $x, y \in M$. But this means that $\lambda x + (1-\lambda)y \in M$, equivalently, M is convex.
- Assume δ_M is lower semi-continuous. Then it follows from Theorem 26 that $S_\alpha(\delta_M)$ is closed for all $\alpha \in \mathbb{R}$. But, for any $\alpha \in \mathbb{R}$, $S_\alpha(\delta_M) = \{x \in X : \delta_M(x) \leq \alpha\} = M$ (from the definition of δ_M), so M is closed. Conversely, assume that M is closed. Then, for any $\alpha \in \mathbb{R}$, $S_\alpha(\delta_M) = M$, hence δ_M is lower semi-continuous from Theorem 26.

This concludes the proof.

Global minimum

A global minimum for a function $f : A \to \bar{\mathbb{R}}$, where $A \subset X$, is an $x' \in A$ such that $f(x') \leq f(x)$ for all $x \in A$.

A local minimum for f is an $x' \in A$ such that there exists a neighborhood U of x' such that $x \in U \Rightarrow f(x') \leq f(x)$.

Properties of convex functions

K. R. Dahl &
A. B. Huseby

Theorem (Properties of convex functions)

Let $C \subseteq X$ be a convex set, $f : C \to \mathbb{R}$ be a convex function. Then the following properties hold:

- *If f has a local minimum x', then x' is also a global minimum for f.*
- *If $C = \mathbb{R}$, so that $f : \mathbb{R} \to \mathbb{R}$ and f is differentiable, then f' is monotonically increasing.*
- *If a function $g : \mathbb{R} \to \mathbb{R}$ is twice differentiable and $g''(x) > 0$, then g is convex.*
- *Jensen's inequality: For $x_1, \dots, x_n \in X, \lambda_1, \dots, \lambda_n \in \mathbb{R}, \lambda_k \geq 0$, for $k = 1, \dots, n$, $\sum_{k=1}^{n} \lambda_k = 1$, the following inequality holds*

$$f(\sum^{n} \lambda_k x_k) \leq \sum^{n} \lambda_k f(x_k).$$

Properties of convex functions, ctd.

K. R. Dahl & A. B. Huseby

Theorem

- *The sum of convex functions is convex.*
- αf *is convex if* $\alpha \in \mathbb{R}, \alpha \geq 0$.
- *If* $(f_n)_{n \in \mathbb{N}}$ *is a sequence of convex functions,* $f_n : X \to \mathbb{R}$, *and* $f_n \to f$ *pointwise as* $n \to \infty$, *then* f *is convex.*
- $\mathrm{dom}(f)$ *is a convex set*
- *If* $\alpha \in \bar{\mathbb{R}}$, *then the sublevel set for* f, $S_\alpha(f)$ *is a convex set. Similarly,* $\{x \in C : f(x) < \alpha\}$ *is a convex set.*
- *Maximization: Let* $\{f_\lambda\}$ *be an arbitrary family of convex functions, then* $g(x) = \sup_\lambda f_\lambda(x)$ *is convex. Also,* $g(x) = \sup_y f(x, y)$ *is convex if* $f(x, y)$ *is convex in* x *for all* y.
- *Minimization: Let* $f : X \times X \to \bar{\mathbb{R}}$ *be a convex function. Then* $g(x) = \inf_y f(x, y)$ *is convex.*

Proof of properties of convex functions

Proof:

- Suppose x' is a local minimum for f, that is: There exists a neighborhood $U \subseteq C$ of x' such that $f(x') \leq f(x)$ for all $x \in U$. We want to show that $f(x') \leq f(x)$ for all $x \in C$. Let $x \in C$. Consider the convex combination $\lambda x + (1-\lambda)x'$. This convex combination converges towards x' as $\lambda \to 0$. Therefore, for a sufficiently small λ^*, $\lambda^* x + (1-\lambda^*)x' \in U$, so since f is convex

$$\begin{aligned} f(x') &\leq f(\lambda^* x + (1-\lambda^*)x') \\ &\leq \lambda^* f(x) + (1-\lambda^*)f(x') \end{aligned}$$

  which, by rearranging the terms, shows that $f(x') \leq f(x)$. Therefore, x' is a global minimum as well.

- Follows from Definition 17 and the definition of the derivative.

Proof of properties of convex functions, ctd.

- Left as an exercise.
- Left as an exercise.
- Left as an exercise.
- Follows from Definition 17.
- Use Definition 17 and the homogeneity and additivity of limits.
- Follows from the definitions.
- Follows from the definitions, but is included here as an example of a typical basic proof. Let $x, y \in S_\alpha(f)$. Then $f(x), f(y) \leq \alpha$. Then $\lambda x + (1-\lambda)y \in S_\alpha(f)$ because

  $$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) \leq \lambda\alpha + (1-\lambda)\alpha =$$

  where the first inequality follows from the convexity of f, and the second inequality follows from that $x, y \in S_\alpha(f)$.
- sup is a limit, so the result is a consequence of property 7.
- Same as property 10.

Quasiconvex functions

K. R. Dahl &
A. B. Huseby

Goal: Introduce quasiconvex functions.

Quasi-convexity is weaker than convexity.

Still: Strong enough to be useful!

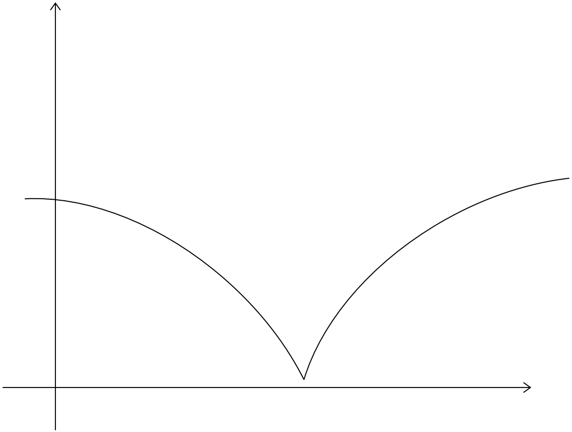
Applications in optimization, game theory and economics.

Definition (Quasiconvex function)

Let $S \subseteq X$ be convex. A function $f : S \to \mathbb{R}$ is quasiconvex if for all $x, y \in S$ and $\lambda \in [0, 1]$, we have

$$f(\lambda x + (1 - \lambda)y) \leq \max\{f(x), f(y)\}.$$

A quasiconvex, but not convex, function



Defining quasiconvexity via sublevel sets

An equivalent way to define quasiconvexity is via convexity of the sublevel sets $S_\alpha := \{x \in S : f(x) \leq \alpha\}$ for all α.

Theorem

Let $S \subseteq X$ be convex and let $f : S \to \mathbb{R}$. Then, f is quasiconvex if and only if the α-sublevel sets

$$S_\alpha = \{x \in S : f(x) \leq \alpha\}$$

are convex for all $\alpha \in \mathbb{R}$.

Proof: Left as an exercise.

Convex functions are quasiconvex

K. R. Dahl & A. B. Huseby

All convex functions are quasiconvex.

The opposite is not true: There exists quasiconvex functions that are not convex, see Figure 61.

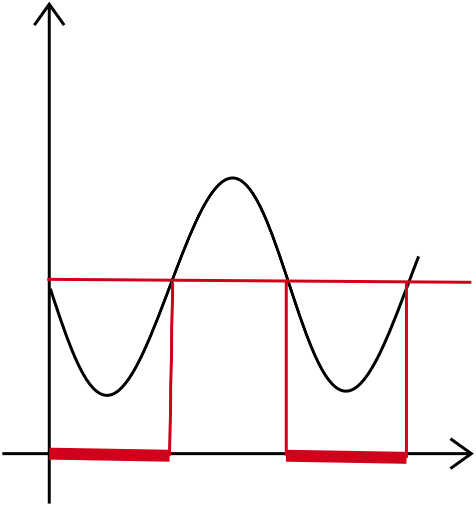
Concave functions can be quasiconvex:An example of this is $f(x) = log(x)$, defined on the positive real numbers.

Not all functions are quasiconvex

Not all functions are quasiconvex.

An example of a function which is not quasiconvex is illustrated in Figure 61.

A function which is not quasiconvex



Why is this function not quasiconvex?

K. R. Dahl & A. B. Huseby

This fuction is not quasiconvex because the set of points in the domain where the function values are below the horizontal red line is the union of the two bold, red intervals, which is not a convex set.

Hence, the sublevel set S_α for this particular α is not convex, and therefore the function does not satisfy the condition in Proposition 33