STK4400 – Week 5

K. R. Dahl & A. B. Huseby

Department of Mathematics
University of Oslo, Norway

To do: Summarize convexity, optimization and risk measures

Convexity and optimization: Overview

- Convexity theory: Convex sets, hyperplanes, 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.

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

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.

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 theorem

Theorem (Supporting hyperplane)

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

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 $\mathcal{T} \subseteq \Pi^+$.

Properties of convex sets

K. R. Dahl & A. B. Huseby

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) $\mathrm{conv}(A)$ *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, $\mathrm{int}(C)$, *the relative interior* $\mathrm{rint}(C)$ *and the closure of* C, $\mathrm{cl}(C)$, *are convex sets as well.*

The proof is left as an exercise.

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

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

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 $\mathrm{epi}(f) = \{(x, \alpha) : x \in X, \alpha \in \mathbb{R}, \alpha \geq f(x)\}$.

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 $X \times \mathbb{R}$).

The definitions are equivalent

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 $x_1, \ldots, x_n \in X, \lambda_1, \ldots, \lambda_n \in \mathbb{R}, \lambda_k \geq 0$, for $k = 1, \ldots, n$, $\sum_{k=1}^n \lambda_k = 1$, the following inequality holds*

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

Properties of convex functions, ctd.

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.*

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

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

Optimization

Optimization is the mathematical theory of maximization and minimization problems.

Useful in many applications, for example in logistic problems, finding the best spot to set up a wind-farm, and in mathematical finance.

Example (finance): Consider an investor who wants to maximize her utility, given various constraints (for instance her salary).

How can we solve this problem?

Basic optimization problem

Let X be a vector space, $f : X \to \bar{\mathbb{R}}$, $g : X \to \mathbb{R}^n$ and $S \subseteq X$.

Consider an optimization problem of the form

$$\begin{array}{ll} \min & f(x) \\ \text{subject to} & \\ & g(x) \leq 0 \text{ (componentwise)} \\ & x \in S. \end{array} \tag{2}$$

In problem (2), f is called the *objective function*.

Furthermore, $g(x) \leq 0$, $x \in S$ are called the *constraints* of the problem.

Transforming an optimization problem

K. R. Dahl & A. B. Huseby

A useful technique when dealing with optimization problems is transforming the problem.

Example: A constraint of the form $h(x) \geq y$ (for $h : X \to \mathbb{R}^n$, $y \in \mathbb{R}^n$) is equivalent to $y - h(x) \leq 0$, which is of the form $g(x) \leq 0$ with $g(x) = y - h(x)$.

Similarly, any maximization problem can be turned into a minimization problem (and visa versa) by using that $\inf f(x) = -\sup(-f(x))$.

Any equality constraint can be transformed into two inequality constraints: $h(x) = 0$ is equivalent to $h(x) \leq 0$ *and* $h(x) \geq 0$.

Extreme value theorem

One of the most important theorems of optimization is the extreme value theorem (see Munkres [?]).

Theorem (The extreme value theorem)

If $f : X \to \mathbb{R}$ is a continuous function from a compact set into the real numbers, then there exist points $a, b \in X$ such that $f(a) \geq f(x) \geq f(b)$ for all $x \in X$. That is, f attains a maximum and a minimum.

The extreme value theorem gives the existence of a maximum and a minimum in a fairly general situation.

NOTE: These may not be unique.

Lagrange duality

The concept of deriving a "dual" problem to handle constraints is the idea of Lagrange duality as well.

Lagrange duality begins with a problem of the form (2) (or the corresponding maximization problem), and derives a dual problem which gives lower (upper) bounds on the optimal value of the problem.

Linear programming duality is a special case of Lagrange duality.

Since Lagrange duality is more general, one cannot get the strong theorems of linear programming.

The duality concept is generalized even more in convex duality theory.

The convex duality method: Summary

K. R. Dahl & A. B. Huseby

- To begin, there is a minimization problem $\min_{x \in X} f(x)$ which cannot be solved directly.
- Find a function $F : X \times U \to \bar{\mathbb{R}}$, where U is a vector space, such that $f(x) = F(x, 0)$.
- Introduce the linear space Y, paired with U, and define the Lagrange function $K : X \times Y \to \bar{\mathbb{R}}$ by $K(x, y) = \inf_{u \in U}\{F(x, u) + \langle u, y \rangle\}$.
- Try to find a saddle point for K. If this succeeds, Theorem **??** tells us that this gives the solution of (P) and (D).
- Theorem **??** tells us that K has a saddle point if and only if $\varphi(0) = (\mathrm{cl}(\mathrm{co}(\varphi)))(0)$. Hence, if the value function φ is convex, the lower semi-continuity of φ is a sufficient condition for the absence of a duality gap.

Convex risk measures

Definition (Convex risk measure)

A convex risk measure is a function $\rho : \mathbb{X} \to \mathbb{R}$ which satisfies the following for each $X, Y \in \mathbb{X}$:

(*i*) (Convexity) $\rho(\lambda X + (1-\lambda)Y) \leq \lambda \rho(X) + (1-\lambda)\rho(Y)$ for $0 \leq \lambda \leq 1$.

(*ii*) (Monotonicity) If $X \leq Y$, then $\rho(X) \geq \rho(Y)$.

(*iii*) (Translation invariance) If $m \in \mathbb{R}$, then $\rho(X + m\mathbf{1}) = \rho(X) - m$.

If $\rho(X) \leq 0$, X is acceptable since it does not have a positive risk. On the other hand, if $\rho(X) > 0$, X is unacceptable.

Coherent risk measures

If a convex risk measure also satisfies positive homogeneity, that is if

$$\lambda \geq 0 \Rightarrow \rho(\lambda X) = \lambda \rho(X)$$

then ρ is called a coherent risk measure.

Interpretations of convex risk measures

Interpret ρ as a capital requirement, that is: $\rho(X)$ is the extra amount of money which should be added to the portfolio in a risk free way to make the position acceptable for an agent.

Interpretation of the conditions in Definition 13:

Convexity: Diversification reduces risk. The total risk of loss in two portfolios should be reduced when the two are weighed into a mixed portfolio.

Interpretations of convex risk measures, ctd.

K. R. Dahl & A. B. Huseby

Monotonicity: The risk of loss is reduced by choosing a portfolio that has a higher value in every possible state of the world.

Translation invariance: ρ is the amount of money one needs to add to the portfolio in order to make it acceptable for an agent. Hence, if one adds a risk free amount m to the portfolio, the capital requirement should be reduced by the same amount.

Convex combinations and max of convex risk measures are convex risk measures

K. R. Dahl & A. B. Huseby

Starting with n convex risk measures, one can derive more convex risk measures:

Theorem

Let $\rho_1, \rho_2, \ldots, \rho_n$ be convex risk measures.

1. *If $\lambda_1, \lambda_2, \ldots, \lambda_n \geq 0$ and $\sum_{i=1}^n \lambda_i = 1$, then $\rho = \sum_{i=1}^n \lambda_i \rho_i$ is a convex risk measure as well.*
2. *$\rho = \max\{\rho_1, \rho_2, \ldots, \rho_n\}$ is a convex risk measure.*

The acceptance set

K. R. Dahl & A. B. Huseby

Associated with every convex risk measure ρ, there is a natural set of all acceptable portfolios: The acceptance set, $\mathcal{A}_\rho$, of ρ.

Definition (The acceptance set of a convex risk measure, $\mathcal{A}_\rho$)

A convex risk measure ρ induces a set

$$\mathcal{A}_\rho = \{X \in \mathbb{X} : \rho(X) \leq 0\}$$

The set $\mathcal{A}_\rho$ is called the acceptance set of ρ.

Associated measure of risk

K. R. Dahl & A. B. Huseby

Conversely, given a class $\mathcal{A} \subseteq \mathbb{X}$, one can associate a quantitative risk measure $\rho_{\mathcal{A}}$ to it.

Definition (Associated measure of risk)

Let $\mathcal{A} \subseteq \mathbb{X}$ be a set of "acceptable" random variables. This set has an associated measure of risk $\rho_{\mathcal{A}}$ defined as follows: For $X \in \mathbb{X}$, let

$$\rho_{\mathcal{A}}(X) = \inf\{m \in \mathbb{R} : X + m \in \mathcal{A}\}. \tag{3}$$

$\rho_{\mathcal{A}}(X)$ measures how much one must add to the portfolio X, in a risk free way, to get the portfolio into the set $\mathcal{A}$ of acceptable portfolios.

Connection between risk measure and acceptance sets

Theorem

Let ρ be a convex risk measure with acceptance set $\mathcal{A}_\rho$. Then:

(i) $\rho_{\mathcal{A}_\rho} = \rho$

(ii) *$\mathcal{A}_\rho$ is a nonempty, convex set.*

(iii) *If $X \in \mathcal{A}_\rho$ and $Y \in \mathbb{X}$ are such that $X \leq Y$, then $Y \in \mathcal{A}_\rho$*

(iv) *If ρ is a coherent risk measure, then $\mathcal{A}_\rho$ is a convex cone.*

Conversely, let $\mathcal{A}$ be a nonempty, convex subset of $\mathbb{X}$. Let $\mathcal{A}$ be such that if $X \in \mathcal{A}$ and $Y \in \mathbb{X}$ satisfy $X \leq Y$, then $Y \in \mathcal{A}$. Then, the following holds:

(v) *$\rho_{\mathcal{A}}$ is a convex risk measure.*

(vi) *If $\mathcal{A}$ is a convex cone, then $\rho_{\mathcal{A}}$ is a coherent risk measure.*

(vii) $\mathcal{A} \subseteq \mathcal{A}_{\rho_{\mathcal{A}}}$.

The finite dimensional case

K. R. Dahl & A. B. Huseby

Choose a specific set of paired spaces, $\mathbb{X}$ and V: Theorem by Föllmer and Schied:

Let $\mathbb{X} = \mathbb{R}^n$, $V = \mathbb{R}^n$ be paired spaces with the standard Euclidean inner product, denoted $\cdot$, as pairing.

Let $(\Omega, \mathcal{F})$ be a measurable space and let $\mathcal{P}$ denote the set of all probability measures over Ω.

Dual representation of convex risk measures: Finite Ω

Theorem

Assume that Ω is finite. Then, any convex risk measure $\rho : \mathbb{X} \to \mathbb{R}$ can be represented in the form

$$\rho(X) = \sup_{Q \in \mathcal{P}} \{E_Q[-X] - \alpha(Q)\} \tag{4}$$

where $E_Q[\cdot]$ denotes the expectation with respect to Q and $\alpha : \mathcal{P} \to (-\infty, \infty]$ is a "penalty function" which is convex and closed. Actually, $\alpha(Q) = \rho^*(-Q)$ for all $Q \in \mathcal{P}$.

Interpretation of the dual representation

Theorem 18 says that any convex risk measure $\rho : \mathbb{R}^n \to \mathbb{R}$ is the expected value of the negative of a contingent claim, $-X$, minus a penalty function, $\alpha(\cdot)$, under the worst case probability.

Note that we consider the expectation of $-X$, not X, since losses are negative in our context.

The penalty function $\alpha(Q)$

K. R. Dahl & A. B. Huseby

Know: The penalty function α in Theorem 18 is of the form $\alpha(Q) = \rho^*(-Q)$.

Luthi and Doege proved that it is possible to derive a more intuitive representation of α.

Theorem

Let $\rho : \mathbb{R}^n \to \mathbb{R}$ be a convex risk measure, and let $\mathcal{A}_\rho$ be its acceptance set (in the sense of Definition 15). Then, Theorem 18 implies that $\rho(X) = \sup_{Q \in \mathcal{P}}\{E_Q[-X] - \alpha(Q)\}$, where $\alpha : \mathcal{P} \to \mathbb{R}$ is a penalty function. Then, α is of the form

$$\alpha(Q) = \sup_{X \in \mathcal{A}_\rho} \{E_Q[-X]\}.$$

Value-at-Risk

Value at risk, VaR, is the most commonly used risk measure in practice (insurance, banks, investment funds etc.).

Interpretation: For a given portfolio, time horizon and probability λ, VaR is the maximum potential loss over the time period after excluding the λ percent worst cases.

Let X be a random variable representing a financial position.

X may represent one stock, a portfolio of stocks or the financial holdings of an entire firm.

Negative values of $X(\omega)$ correspond to losses, and positive values to profit.

Mathematical definition of VaR

VaR is defined as follows: Fix some level $\lambda \in (0,1)$ (typically close to 0), and define $Y := -X$.

Note that for the random variable Y, losses are positive numbers.

Then, VaR_λ is defined as the $(1-\lambda)$-quantile of Y

$$\mathrm{VaR}_\lambda(X) := F_Y^{-1}(1-\lambda) \tag{5}$$

where F_Y is the cumulative distribution function of Y.

Despite its use in practice, value at risk has major drawbacks:

- In general, VaR_λ is not convex, see Föllmer and Knispel: Diversification may increase the risk w.r.t. VaR. This is economically unreasonable.
- From equation (5): VaR_λ ignores extreme losses which occur with small probability. This tail insensitivity makes it an unsuitable measure of risk in situations where the consequences of large losses are very bad (e.g., insurance companies not being able to pay their customers).

These drawbacks of VaR lead to the development of the theory of convex and coherent risk measures.

Still widely used in practice, despite its deficiencies.

Average value at risk

K. R. Dahl & A. B. Huseby

Average value at risk (AVaR), also called expected shortfall (ES) or conditional value at risk (CVaR), was introduced to mend the deficiencies of value at risk.

For $\lambda \in (0, 1]$, the average value at risk is defined as

$$\mathrm{AVaR}_\lambda(X) := \frac{1}{\lambda} \int_0^\lambda \mathrm{VaR}_\alpha(X) d\alpha. \tag{6}$$

Average value at risk can be interpreted as the expected loss in a presupposed percentage of worst cases.

AVaR is a coherent risk measure

Föllmer and Schied prove that AVaR_λ is a coherent risk measure, with a dual representation

$$\mathrm{AVaR}_\lambda(X) = \max_{Q \in \mathcal{Q}_\lambda} E_Q[-X]$$

where $\mathcal{Q}_\lambda := \{Q << P | \frac{dQ}{dP} \leq \lambda\}$. That is, $\mathcal{Q}_\lambda$, is the set of all measures Q that are absolutely continuous w.r.t. the measure P given that the Radon-Nikodym derivative of Q w.r.t. P is less than or equal λ (see Shilling for more on these measure theoretical concepts).

Note also that for $\lambda = 1$, average value at risk reduces to $E_P[-X]$, i.e., the expected loss.