Stochastic Programming: Tutorial

Part I

W. Römisch

Humboldt-University Berlin
Institute of Mathematics
10099 Berlin, Germany

http://www.math.hu-berlin.de/~romisch

Optimization: Computation, Theory and Modeling Workshop *Optimization under Uncertainty*
Singapore, December 10-11, 2012

DFG Research Center MATHEON
Mathematics for key technologies

Introduction

What is Stochastic Programming ?

- Mathematics for Decision Making under Uncertainty
- subfield of Mathematical Programming (MSC 90C15)

Stochastic programs are **optimization models**

- having special properties and structures,
- depending on the underlying probability distribution,
- requiring specific approximation and numerical approaches,
- having close relations to practical applications.

Selected recent monographs:

P. Kall, S.W. Wallace 1994, A. Prekopa 1995, J.R. Birge, F. Louveaux 1997
A. Ruszczynski, A. Shapiro (eds.): Stochastic Programming, Handbook, Elsevier, 2003
S.W. Wallace, W.T. Ziemba (eds.): Applications of Stochastic Programming, MPS-SIAM, 2005,
P. Kall, J. Mayer: Stochastic Linear Programming, Kluwer, 2005,
A. Shapiro, D. Dentcheva, A. Ruszczyński: Lectures on Stochastic Programming, MPS-SIAM, 2009.
G. Infanger (ed.): Stochastic Programming - The State-of-the-Art, Springer, 2010.

Contents (Part I)

Motivating example: Newsvendor problem

- ξ uncertain daily demand for a (daily) newspaper
- x decision about the quantity of newspapers to be purchased from a distributor
- c cost to be paid by the newsvendor for one newspaper at the beginning of the day
- s selling price for one newspaper
- r return price for one unsold newspaper at the end of the day

Revenue function: (Assumption: $0 \le r < c < s$)

$$f(x,\xi) = \begin{cases} (s-c)x & , x \le \xi, \\ s\xi + r(x-\xi) - cx & , x > \xi \end{cases}$$

Expected revenue:

$$\mathbb{E}[f(x,\xi)] = \int_0^\infty f(x,\xi)\,dF(\xi) = \sum_{k=1}^{\infty} p_k f(x,k),$$

where $F(w) = \mathbb{P}(\xi \le w) = \sum_{k=1, k\le w} p_k$ is the piecewise constant (cumulative) probability distribution function of the demand ξ.

Maximization of the expected revenue:

$$\max\left\{ \sum_{k=1,k\le x} p_k[(r-c)x + (s-r)k] + \sum_{k>x}^{\infty} p_k(s-c)x : x \ge 0 \right\}$$

or

$$\max\left\{ \sum_{k=1,k\le x} p_k\Big[(s-c)x + (s-r)(k-x)\Big] + \sum_{k>x}^{\infty} p_k(s-c)x : x \ge 0 \right\}$$

or

$$\max\left\{ (s-c)x + (s-r) \sum_{k=1,k\le x} p_k(k-x) : x \ge 0 \right\}$$

or

$$\max\left\{ (s-c)x - (s-r)\,\mathbb{E}\Big[\max\{0, x-\xi\}\Big] : x \ge 0 \right\}$$

or

$$\max\left\{ [(s-c) - (s-r)F(x)]x + (s-r) \sum_{k=1,k\le x} kp_k : x \ge 0 \right\}$$

Hence, x can be maximized as long as $[(s-c)-(s-r)F(x)] \geq 0$, i.e.,

$$F(x) \leq \frac{s-c}{s-r}.$$

Hence, the optimal decision x_* is the minimal $n \in \mathbb{N}$ such that

$$F(n) = \sum_{k=1}^{n} p_k \geq \frac{s-c}{s-r}.$$

The latter model will be called two-stage stochastic program with first-stage decision x and optimal recourse $\max\{0, x-\xi\}$.

Of course, the newsvendor needs knowledge on the distribution function F (at least, approximately).

Basic assumption in stochastic programming: The probability distribution is independent on the decision.

The problem may occur that the random variable $f(x_*, \xi)$ has a high variance $\mathbb{V}[f(x_*, \xi)] = \mathbb{E}[f(x_*, \xi)^2] - [\mathbb{E}[f(x_*, \xi)]]^2$. Then the decision x_* has high risk and one should be interested in a risk averse decision whose expected revenue is still close to $\mathbb{E}[f(x_*, \xi)]$.

An alternative is to consider the risk averse optimization problem

$$\max \left\{ \mathbb{E}[f(x, \xi)] - \gamma \mathbb{V}[f(x, \xi)] : x \geq 0 \right\}$$

with a risk aversion parameter $\gamma \geq 0$.

In general, one might be interested in a risk averse alternative with certain risk functional $\mathbb{F}$ instead of the variance $\mathbb{V}$ in order to maintain good properties of the optimization problem.

The newsvendor may also be interested in making a specific amount of money b with high probability, but minimal work.

Optimization model with probabilistic constraints:

$$\min \{x \in \mathbb{R} : \mathbb{P}(f(x,\xi) \geq b) \geq p\}$$

with $p \in (0,1)$ close to 1. The model is equivalent to

$$\min \left\{x \in \mathbb{R} : (s-c)x \geq b, \mathbb{P}\left(\xi \geq \frac{b+(c-r)x}{s-r}\right) \geq p\right\}$$

or

$$\min \left\{x \in \mathbb{R} : (s-c)x \geq b, \frac{b+(c-r)x}{s-r} \leq F^{-1}(1-p)\right\}$$

A feasible solution of the optimization model exists if

$$b \leq (s-c)F^{-1}(1-p)\,,$$

leading to the optimal solution $\hat{x} = \frac{b}{s-c}$.

Approaches to optimization models under stochastic uncertainty

Let us consider the optimization model

$$\min\{f(x,\xi) \,:\, x \in X,\, g(x,\xi) \leq 0\}\,,$$

where $\xi : \Omega \to \Xi$ is a random vector defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, Ξ and X are closed subsets of $\mathbb{R}^s$ and $\mathbb{R}^m$, respectively, $f : X \times \Xi \to \mathbb{R}$ and $g : X \times \Xi \to \mathbb{R}^d$ are lower semicontinuous.

Aim: Finding optimal decisions before knowing the random outcome of ξ (here-and-now decision).

Main approaches:

- Replace the objective by $\mathbb{E}[f(x,\xi)]$ or by $\mathbb{F}[f(x,\xi)]$, where $\mathbb{E}$ denotes expectation (w.r.t. $\mathbb{P}$) and $\mathbb{F}$ some functional on the space of real random variables (e.g., playing the role of a *risk functional*).

- (i) Replace the random constraints by the constraint

$$\mathbb{P}(\{\omega \in \Omega : g(x, \xi(\omega)) \leq 0\}) = \mathbb{P}(g(x, \xi) \leq 0) \geq p$$

  where $p \in [0, 1]$ denotes a probability level, **or** (ii) go back to the *modeling stage* and introduce a recourse action to compensate constraint violations and add the optimal recourse cost to the objective.

The first variant leads to stochastic programs with probabilistic or chance constraints:

$$\min\{\mathbb{E}[f(x, \xi)] : x \in X,\ \mathbb{P}(g(x, \xi) \leq 0) \geq p\}$$

The second variant leads to two-stage stochastic programs with recourse:

$$\min\{\mathbb{E}[f(x, \xi)] + \mathbb{E}[q(y, \xi)] : x \in X,\ y \in Y,\ g(x, \xi) + h(y, \xi) \leq 0\}.$$

or $\mathbb{E}$ replaced by a risk functional $\mathbb{F}$.

Properties of expectation functions

We consider analytical properties of functions having the form

$$\mathbb{E}[f(x,\xi)] = \int_{\mathbb{R}^s} f(x,\xi)P(d\xi), \quad (x \in \mathbb{R}^m)$$

where $f : \mathbb{R}^m \times \mathbb{R}^s \to \overline{\mathbb{R}}$, $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$ denoting the extended real numbers, is an integrand such that

$$f(x,\cdot) \text{ is measurable and } \mathbb{E}[[f(x,\xi)]_{\pm}] < +\infty$$

and P is a (Borel) probability measure on $\mathbb{R}^s$.

Aim: Properties of the expectation function

$$x \mapsto \mathbb{E}[f(x,\xi)] \quad (\text{on } \mathbb{R}^s)$$

under reasonable assumptions on the integrand f.

Proposition 1: Assume that
(i) $f(\cdot,\xi)$ is lower semicontinuous at $x_0 \in \mathbb{R}^m$ for P-almost all $\xi \in \mathbb{R}^s$,
(ii) there exists a P-integrable function $z : \mathbb{R}^s \to \overline{\mathbb{R}}$, such that $f(x,\xi) \geq z(\xi)$ for P-almost all $\xi \in \mathbb{R}^s$ and all x in a neighborhood of x_0.
Then the function $x \mapsto \mathbb{E}[f(x,\xi)]$ is lower semicontinuous at x_0.

Proof: follows by applying Fatou's Lemma.

Proposition 2: Assume that
(i) $f(\cdot,\xi)$ is continuous at $x_0 \in \mathbb{R}^m$ for P-almost all $\xi \in \mathbb{R}^s$,
(ii) there exists a P-integrable function $z : \mathbb{R}^s \to \overline{\mathbb{R}}$, such that $|f(x,\xi)| \leq z(\xi)$ for P-almost all $\xi \in \mathbb{R}^s$ and all x in a neighborhood of x_0.
Then the function $x \mapsto \mathbb{E}[f(x,\xi)]$ is finite in a neighborhood of x_0 and continuous at x_0.

Proof: follows by applying Lebesgue's dominated convergence theorem.

Example:

For $f(x,\xi) = -\mathbb{1}_{(-\infty,x]}(\xi)$, $(x,\xi) \in \mathbb{R} \times \mathbb{R}$, where $\mathbb{1}_A$ denotes the characteristic function of $A \subset \mathbb{R}$, the function $x \to \mathbb{E}[f(x,\xi)]$ is lower semicontinuous on $\mathbb{R}$, but continuous at $x_0 \in \mathbb{R}$ only if $P(\{x_0\}) = 0$.

Proposition 3: Assume

(i) $\mathbb{E}[|f(x_0,\xi)|] < +\infty$ for some $x_0 \in \mathbb{R}^m$,

(ii) there exists a P-integrable function $L : \mathbb{R}^s \to \mathbb{R}$ such that

$$|f(x,\xi) - f(\tilde{x},\xi)| \leq L(\xi)\|x - \tilde{x}\|$$

holds for all x and $\tilde{x}$ in a neighborhood U of x_0 in $\mathbb{R}^m$ and P-almost all $\xi \in \mathbb{R}^s$.

Then the function $x \mapsto \mathbb{E}[f(x,\xi)]$ is Lipschitz continuous on U.

(iii) Assume, in addition, $f(\cdot,\xi)$ is differentiable at x_0 for P-almost all $\xi \in \mathbb{R}^s$.

Then the function $F(x) = \mathbb{E}[f(x,\xi)]$ is differentiable at x_0 and

$$\nabla F(x_0) = \mathbb{E}[\nabla_x f(x_0,\xi)].$$

Proposition 4: Assume that
(i) the function $x \mapsto \mathbb{E}[f(x,\xi)]$ is finite on some neighborhood U of x_0,
(ii) $f(\cdot,\xi) : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is convex for P-almost all $\xi \in \mathbb{R}^s$.
Then the function $F(x) = \mathbb{E}[f(x,\xi)]$ from $\mathbb{R}^m$ to $\mathbb{R} \cup \{+\infty\}$ is convex and directionally differentiable at x_0 with

$$F'(x_0; h) = \mathbb{E}[f'(x_0, \xi; h)] \quad (\forall h \in \mathbb{R}^m).$$

(iii) Assume, in addition, that f is a normal integrand and $\operatorname{dom} F$ has nonempty interior.
Then F is subdifferentiable at x_0 and

$$\partial F(x_0) = \int_{\mathbb{R}^s} \partial f(x_0, \xi) P(d\xi) + N_{\operatorname{dom} F}(x_0).$$

(Ruszczyński/Shapiro, Handbook, 2003)

Two-stage stochastic programming models with recourse

Consider a linear program with stochastic parameters of the form

$$\min\{\langle c, x\rangle : x \in X,\, T(\xi)x = h(\xi)\},$$

where $\xi : \Omega \to \Xi$ is a random vector defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $c \in \mathbb{R}^m$, Ξ and X are polyhedral subsets of $\mathbb{R}^s$ and $\mathbb{R}^m$, respectively, and the $d \times m$-matrix $T(\cdot)$ and vector $h(\cdot) \in \mathbb{R}^d$ are affine functions of ξ.

Idea: Introduce a recourse variable $y \in \mathbb{R}^{\overline{m}}$, recourse costs $q(\xi) \in \mathbb{R}^{\overline{m}}$, a fixed recourse $d \times \overline{m}$-matrix W, a polyhedral cone $Y \subseteq \mathbb{R}^{\overline{m}}$, and solve the second-stage or recourse program

$$\min\{\langle q(\xi), y\rangle : y \in Y, Wy = h(\xi) - T(\xi)x\}.$$

Add the expected minimal recourse costs $\mathbb{E}[\Phi(x, \xi)]$ (depending on the first-stage decision x) to the original objective and consider

$$\min\left\{\langle c, x\rangle + \mathbb{E}[\Phi(x, \xi)] : x \in X\right\},$$

where $\Phi(x, \xi) := \inf\{\langle q(\xi), y\rangle : y \in Y, Wy = h(\xi) - T(\xi)x\}$.

Two formulations of two-stage models

Deterministic equivalent of the two-stage model:

$$\min\left\{\langle c, x\rangle + \int_{\Xi} \Phi(x,\xi)P(d\xi) : x \in X\right\},$$

where $P := \mathbb{P}\xi^{-1} \in \mathcal{P}(\Xi)$ is the probability distribution of the random vector ξ and $\Phi(\cdot,\cdot)$ is the infimum function of the second-stage program.

Infinite-dimensional optimization model:

$$\min\left\{\langle c, x\rangle \ + \ \int_{\Xi}\langle q(\xi), y(\xi)\rangle P(d\xi) : x \in X,\ y \in L_r(\Xi, \mathcal{B}(\Xi), P),\right.$$
$$\left. y(\xi) \in Y,\ Wy(\xi) = h(\xi) - T(\xi)x\right\},$$

where $r \in [1, +\infty]$ is selected properly.

If the probability distribution P of ξ is assumed to have p-th order moments, i.e., $\int_{\Xi}\|\xi\|^p P(d\xi) < \infty$, with $p > 1$, r should be chosen such that the constraints of y are consistent with these moment conditions and $\mathbb{E}[\langle q(\xi), y(\xi)\rangle]$ is finite. For example, $r = \frac{p}{p-1}$ is consistent.

Structural properties of two-stage models

We consider the infimum function $v(\cdot,\cdot)$ of the parametrized linear (second-stage) program, namely,

$$\begin{aligned} v(u,t) &= \inf\left\{\langle u,y\rangle : Wy = t, y\in Y\right\} \; ((u,t)\in\mathbb{R}^{\overline{m}}\times\mathbb{R}^d) \\ &= \sup\left\{\langle t,z\rangle : W^\top z - u \in Y^*\right\} \\ \mathcal{D} &= \left\{u : \{z\in\mathbb{R}^r : W^\top z - u\in Y^*\} \neq \emptyset\right\} \end{aligned}$$

where $W^\top$ is the transposed of W and Y^* the polar cone of Y. Hence, we have

$$\Phi(x,\xi) = v(q(\xi), h(\xi) - T(\xi)x).$$

Theorem: (Walkup/Wets 69)
The function $v(\cdot,\cdot)$ is finite and continuous on the polyhedral cone $\mathcal{D}\times W(Y)$. Furthermore, the function $v(u,\cdot)$ is piecewise linear convex on the polyhedral set $W(Y)$ for fixed $u\in\mathcal{D}$, and $v(\cdot,t)$ is piecewise linear concave on $\mathcal{D}$ for fixed $t\in W(Y)$.

Assumptions:

(A1) *relatively complete recourse*: for any $(\xi, x) \in \Xi \times X$, $h(\xi) - T(\xi)x \in W(Y)$;

(A2) *dual feasibility*: $q(\xi) \in \mathcal{D}$ holds for all $\xi \in \Xi$.

(A3) *finite second order moment*: $\int_\Xi \|\xi\|^2 P(d\xi) < \infty$.

Note that (A1) is satisfied if $W(Y) = \mathbb{R}^d$ (complete recourse). In general, (A1) and (A2) impose a condition on the support of P.

Proposition:

Assume (A1) and (A2). Then the deterministic equivalent of the two-stage model represents a convex program (with polyhedral constraints) if the integrals $\int_\Xi v(q(\xi), h(\xi) - T(\xi)x)P(d\xi)$ are finite for all $x \in X$. For the latter it suffices to assume (A3).

An element $x \in X$ minimizes the convex program if and only if

$$0 \in \int_\Xi \partial\Phi(x, \xi)P(d\xi) + N_X(x)\,,$$

$$\partial\Phi(x, \xi) = c - T(\xi)^\top \arg\max_{z \in D(\xi)} z^\top(h(\xi) - T(\xi)x).$$

Discrete approximations of two-stage stochastic programs

Replace the (original) probability measure P by measures P_n having (finite) discrete support $\{\xi_1, \dots, \xi_n\}$ $(n \in \mathbb{N})$, i.e.,

$$P_n = \sum_{i=1}^{n} p_i \delta_{\xi_i},$$

and insert it into the infinite-dimensional stochastic program:

$$\min\{\langle c, x\rangle + \sum_{i=1}^{n} p_i \langle q(\xi_i), y_i\rangle : x \in X, y_i \in Y, \, i = 1, \dots, n,$$

$$\begin{array}{ccccccc} Wy_1 & & & & +T(\xi_1)x & = & h(\xi_1) \\ & Wy_2 & & & +T(\xi_2)x & = & h(\xi_2) \\ & & \ddots & & \vdots & = & \vdots \\ & & & Wy_n & +T(\xi_n)x & = & h(\xi_n)\} \end{array}$$

Hence, we arrive at a (finite-dimensional) large scale block-structured linear program which allows for specific decomposition methods.

(Ruszczyński/Shapiro, Handbook, 2003)

Mixed-integer two-stage stochastic programs

Applied optimization models often contain continuous and integer decisions (e.g. on/off decisions, quantities). If such decisions enter the second-stage program, its optimal value function is no longer continuous and/or convex in general.
We consider

$$\min \left\{ \langle c, x \rangle + \int_{\Xi} \Phi(q(\xi), h(\xi) - T(\xi)x) P(d\xi) : x \in X \right\},$$

where Φ is given by

$$\Phi(u,t) := \inf \left\{ \langle u_1, y_1 \rangle + \langle u_2, y_2 \rangle \;\middle|\; \begin{array}{l} W_1 y_1 + W_2 y_2 \le t \\ y_1 \in \mathbb{R}^{m_1}_+, y_2 \in \mathbb{Z}^{m_2}_+ \end{array} \right\}$$

for all pairs $(u,t) \in \mathbb{R}^{m_1+m_2} \times \mathbb{R}^d$, and $c \in \mathbb{R}^m$, X is a closed subset of $\mathbb{R}^m$, Ξ a polyhedron in $\mathbb{R}^s$, $T \in \mathbb{R}^{d \times m}$, $W_1 \in \mathbb{R}^{d \times m_1}$, $W_2 \in \mathbb{R}^{d \times m_2}$, and $q(\xi) \in \mathbb{R}^{m_1+m_2}$ and $h(\xi) \in \mathbb{R}^d$ are affine functions of ξ, and P is a Borel probability measure.

Assumptions:

(C1) The matrices W_1 and W_2 have rational elements.

(C2) For each pair $(x, \xi) \in X \times \Xi$ it holds that $h(\xi) - T(\xi)x \in \mathcal{T}$ (relatively complete recourse), where

$$\mathcal{T} := \{t \in \mathbb{R}^d | \exists y = (y_1, y_2) \in \mathbb{R}^{m_1} \times \mathbb{Z}^{m_2} \text{ with } W_1 y_1 + W_2 y_2 \leq t\}.$$

(C3) For each $\xi \in \Xi$ the recourse cost $q(\xi)$ belongs to the dual feasible set (dual feasibility)

$$\mathcal{U} := \{u = (u_1, u_2) \in \mathbb{R}^{m_1+m_2} | \exists z \in \mathbb{R}^d_- \text{ with } W_j^\top z = u_j, j = 1, 2\}.$$

(C4) $P \in \mathcal{P}_r(\Xi)$, i.e., $\int_\Xi \|\xi\|^r P(d\xi) < +\infty$, $r \in \{1, 2\}$.

Condition (C2) means that a feasible second stage decision always exists. Both (C2) and (C3) imply $\Phi(u, t)$ to be finite for all $(u, t) \in \mathcal{U} \times \mathcal{T}$. Clearly, it holds $(0, 0) \in \mathcal{U} \times \mathcal{T}$ and $\Phi(0, t) = 0$ for every $t \in \mathcal{T}$.

$r = 1$ holds if either $q(\xi)$ is the only quantity depending on ξ or $q(\xi)$ does not depend on ξ. Otherwise, we set $r = 2$.

With the convex polyhedral cone

$$\mathcal{K} := \{t \in \mathbb{R}^d \mid \exists y_1 \in \mathbb{R}^{m_1} \text{ such that } t \geq W_1 y_1\} = W_1(\mathbb{R}^{m_1}) + \mathbb{R}^d_+$$

one obtains the representation

$$\mathcal{T} = \bigcup_{z \in \mathbb{Z}^{m_2}} (W_2 z + \mathcal{K}).$$

The set $\mathcal{T}$ is always (path) connected (i.e., there exists a polygon connecting two arbitrary points of $\mathcal{T}$) and condition (C1) implies that $\mathcal{T}$ is closed. If, for each $t \in \mathcal{T}$, $Z(t)$ denotes the set

$$Z(t) := \{z \in \mathbb{Z}^{m_2} \mid \exists y_1 \in \mathbb{R}^{m_1} \text{ such that } W_1 y_1 + W_2 z \leq t\},$$

the representation of $\mathcal{T}$ implies that it is decomposable into subsets of the form

$$\begin{aligned} \mathcal{T}(t_0) &:= \{t \in \mathcal{T} \mid Z(t) = Z(t_0)\} \\ &= \bigcap_{z \in Z(t_0)} (W_2 z + \mathcal{K}) \setminus \bigcup_{z \in \mathbb{Z}^{m_2} \setminus Z(t_0)} (W_2 z + \mathcal{K}) \end{aligned}$$

for every $t_0 \in \mathcal{T}$.

In general, the set $Z(t_0)$ is finite or countable, but condition (C1) implies that there exist countably many elements $t_i \in \mathcal{T}$ and $z_{ij} \in \mathbb{Z}^{m_2}$ for j belonging to a finite subset N_i of $\mathbb{N}$, $i \in \mathbb{N}$, such that

$$\mathcal{T} = \bigcup_{i \in \mathbb{N}} \mathcal{T}(t_i) \quad \text{with} \quad \mathcal{T}(t_i) = (t_i + \mathcal{K}) \setminus \bigcup_{j \in N_i} (W_2 z_{ij} + \mathcal{K}).$$

The sets $\mathcal{T}(t_i)$, $i \in \mathbb{N}$, are nonempty and star-shaped, but nonconvex in general.

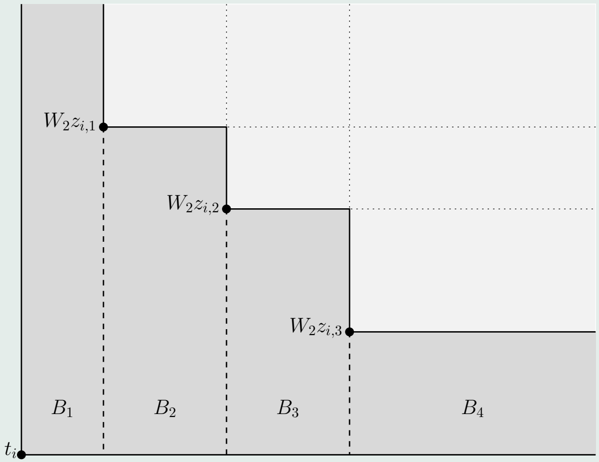


Illustration of $\mathcal{T}(t_i)$ for $W_1 = 0$ and $d = 2$, i.e., $\mathcal{K} = \mathbb{R}^2_+$, with $N_i = \{1, 2, 3\}$ and its decomposition into the sets B_j, $j = 1, 2, 3, 4$, whose closures are rectangular.

If for some $i \in \mathbb{N}$ the set $\mathcal{T}(t_i)$ is nonconvex, it can be decomposed into a finite number of subsets.

This leads to a countable number of subsets B_j, $j \in \mathbb{N}$, of $\mathcal{T}$ whose closures are convex polyhedra with facets parallel to $W_1(\mathbb{R}^{m_1})$ or to suitable facets of $\mathbb{R}^r_+$ and form a partition of $\mathcal{T}$.

Since the sets $Z(t)$ of feasible integer decisions do not change if t varies in some B_j, the function $(u,t) \mapsto \Phi(u,t)$ from $\mathcal{U} \times \mathcal{T}$ to $\mathbb{R}$ has the (local) Lipschitz continuity regions $\mathcal{U} \times B_j$, $j \in \mathbb{N}$ and the estimate

$$|\Phi(u,t) - \Phi(\tilde{u},\tilde{t})| \leq L(\max\{1, \|t\|, \|\tilde{t}\|\}\|u-\tilde{u}\| + \max\{1, \|u\|, \|\tilde{u}\|\}\|t-\tilde{t}\|)$$

holds for all pairs $(u,t), (\tilde{u},\tilde{t}) \in \mathcal{U} \times B_j$ and some (uniform) constant $L > 0$.

(Blair-Jeroslow 77, Bank-Guddat-Kummer-Klatte-Tammer 1982)

For the integrand

$$f_0(x,\xi) = \langle c, x\rangle + \Phi(q(\xi), h(\xi) - T(\xi)x) \quad ((x,\xi) \in X \times \Xi)$$

it holds

$$|f_0(x,\xi) - f_0(x,\tilde{\xi})| \leq \hat{L} \max\{1, \|\xi\|^{r-1}, \|\tilde{\xi}\|^{r-1}\} \|\xi - \tilde{\xi}\| \; (\xi, \tilde{\xi} \in \Xi_{x,j})$$

$$|f_0(x,\xi)| \leq C \max\{1, \|x\|\} \max\{1, \|\xi\|^r\} \; (\xi \in \Xi)$$

for all $x \in X$ with some constants $\hat{L}$ and C and

$$\Xi_{x,j} = \{\xi \in \Xi \,|\, h(\xi) - T(\xi)x \in B_j\} \quad (j \in \mathbb{N})$$

Proposition: (Schultz 93, 95)

Assume (C1)–(C4). Then the objective function

$$F_P(x) = \langle c, x\rangle + \int_\Xi \Phi(q(\xi), h(\xi) - T(\xi)x) P(d\xi)$$

is lower semicontinuous on X and solutions exist if X is compact. If the probability distribution P has a density, the objective function is continuous, but nonconvex in general.

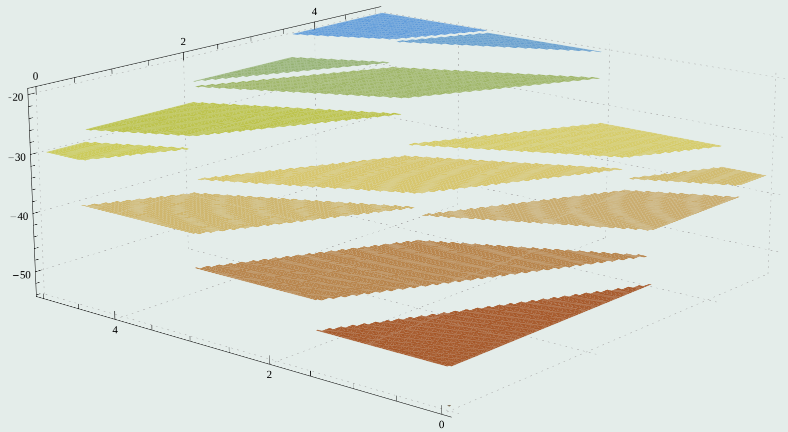
If the support of P is finite, the objective function is piecewise continuous with a finite number of continuity regions, whose closures are polyhedral.

Example: (Schultz-Stougie-van der Vlerk 98)

$m = d = s = 2$, $m_1 = 0$, $m_2 = 4$, $c = (0, 0)$, $X = [0, 5]^2$, $h(\xi) = \xi$, $q(\xi) \equiv q = (-16, -19, -23, -28)$, $y_i \in \{0, 1\}$, $i = 1, 2, 3, 4$, $P \sim \mathcal{U}\{5, 10, 15\}^2$ (discrete)

Second stage problem: MILP with 36 binary variables and 18 constraints.

$$T = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix} \qquad W = \begin{pmatrix} 2 & 3 & 4 & 5 \\ 6 & 1 & 3 & 2 \end{pmatrix}$$



Optimal value function

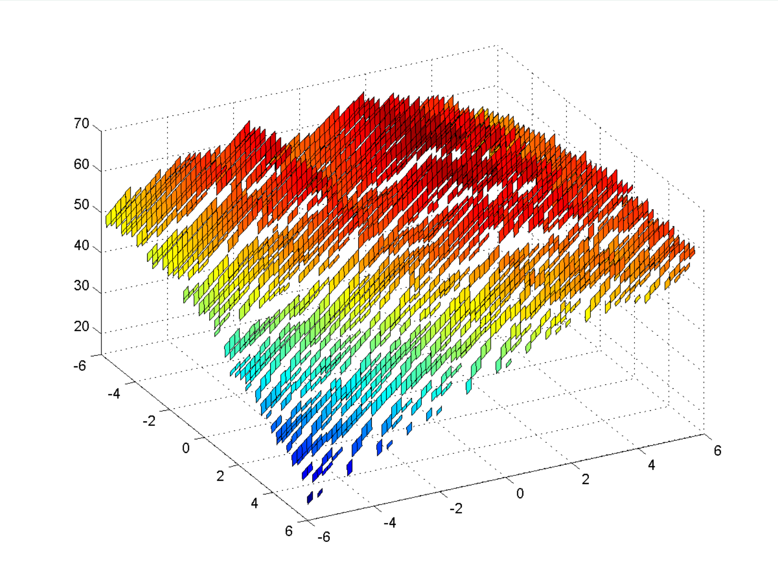
Example: (Schultz-Stougie-van der Vlerk 98)

Stochastic multi-knapsack problem:

$m = d = s = 2,\ m_1 = 0,\ m_2 = 4,\ c = (1.5, 4),\ X = [-5, 5]^2,$
$h(\xi) = \xi,\ q(\xi) \equiv q = (16, 19, 23, 28),\ y_i \in \{0, 1\},\ i = 1, 2, 3, 4,$
$P \sim \mathcal{U}\{5, 5.5, \ldots, 14.5, 15\}^2$ (discrete)

Second stage problem: MILP with 1764 Boolean variables and 882 constraints.

$$T = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix} \qquad W = \begin{pmatrix} 2 & 3 & 4 & 5 \\ 6 & 1 & 3 & 2 \end{pmatrix}$$



Stochastic programs with probabilistic constraints

We consider the stochastic program

$$\min\left\{f(x) : x \in X, P(g(x,\xi) \leq 0) \geq p\right\},$$

where X is a closed subset of $\mathbb{R}^m$, $f : \mathbb{R}^m \to \mathbb{R}$, $g : \mathbb{R}^m \times \mathbb{R}^s \to \mathbb{R}^r$, ξ a random vector with probability distribution P and $p \in (0.1)$.

Problem: If the original optimization problem is smooth, convex or even linear, the probabilistic constraint function

$$G(x) := P(g(x,\xi) \leq 0)$$

may be non-differentiable, non-Lipschitzian and non-convex.

Special forms of probabilistic constraints:

- $g(x,\xi) := \xi - h(x)$, where $h : \mathbb{R}^m \to \mathbb{R}^s$, i.e.,

  $$G(x) = P(\xi \leq h(x)) = F_P(h(x)) \geq p\,,$$

  where $F_P(y) := P(\{\xi \leq y\})$ $(y \in \mathbb{R}^s)$ denotes the (multivariate) probability distribution function of ξ.

- $g(x,\xi) := b(\xi) - A(\xi)x$, where the matrix $A(\cdot)$ and the vector $b(\cdot)$ are affine functions of ξ. Then

$$G(x) := P(\{\xi : A(\xi)x \geq b(\xi)\})$$

corresponds to the probability of a polyhedron depending on x.

Proposition: (Prekopa)
If $H : \mathbb{R}^m \to \mathbb{R}^s$ is a set-valued mapping with closed graph, the function $G : \mathbb{R}^m \to \mathbb{R}$ defined by $G(x) := P(H(x))$ $(x \in \mathbb{R}^m)$ is upper semicontinuous for every probability distribution P on $\mathbb{R}^s$. Hence, the feasible set

$$\mathcal{X}_p(P) = \{x \in X : G(x) = P(H(x)) \geq p\}$$

is closed.
(In particular, H is of the form $H(x) = \{\xi \in \mathbb{R}^s : g(x,\xi) \leq 0\}$, $\mathrm{gph}\, H = \{(x,\xi) \in \mathbb{R}^m \times \mathbb{R}^s : g(x,\xi) \leq 0\}$.)

Proposition: (Henrion 02)
For any $i = 1, \dots, r$ let $g_i(\cdot, \xi)$ be quasiconvex for all $\xi \in \mathbb{R}^s$ and min stable w.r.t. X, i.e., for any $x, \tilde{x} \in X$ there exists $\bar{x} \in X$ such that

$$g_i(\bar{x}, \xi) \leq \min\{g_i(x, \xi), g_i(\tilde{x}, \xi)\} \quad \forall \xi \in \mathbb{R}^s.$$

Then the set $\mathcal{X}_p(P) = \{x \in X : P(g(x, \xi) \leq 0) \geq p\}$ is (path) connected for any $p \in [0, 1]$ and probability distribution P on $\mathbb{R}^s$.

Corollary:
Let A be a (s, m)-matrix and ξ a s-dimensional random vector with distribution P. If the rows of A are positively linear independent, the set $\mathcal{X}_p(P) = \{x \in \mathbb{R}^m : P(Ax \geq \xi) \geq p\}$ is path connected for any $p \in [0, 1]$ and probability distribution P on $\mathbb{R}^s$.

Problem:
Which conditions imply continuity and differentiability properties of $G(x) = P(H(x))$ or convexity of $\mathcal{X}_p(P) = \{x \in X : P(H(x)) \geq p\}$?

Examples:

(i) Let $H(x) = x + \mathbb{R}^s_-$ $(\forall x \in \mathbb{R}^s)$ and P have finite support, i.e.,

$$P = \sum_{i=1}^{n} p_i \delta_{\xi_i},$$

where δ_ξ denotes the Dirac measure placing unit mass at ξ and $p_i > 0$, $i = 1, \dots, n$, $\sum_{i=1}^{n} p_i = 1$. Then

$$\mathcal{X}_p(P) = X \cap (\cup_{i \in I}(\xi_i + \mathbb{R}^s_+))$$

holds for some index set $I \subset \{1, \dots, n\}$ and, hence, is non-convex in general. Moreover, $G = F_P$ is discontinuous with jumps at $\mathrm{bd}(\xi_i + \mathbb{R}^s_-)$.

(ii) Let $H(x) = x + \mathbb{R}^s_-$ $(\forall x \in \mathbb{R}^s)$ and P have a density f_P with respect to the Lebesgue measure on $\mathbb{R}^s$, i.e.,

$$G(x) = F_P(x) = \int_{-\infty}^{x} f_P(y)dy = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_s} f_P(y_1, \dots, y_s)dy_s \cdots dy_1.$$

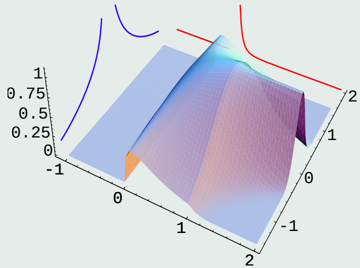
Conjecture: $G = F_P$ is Lipschitz continuous if the density f_P is continuous and bounded.

Answer: The conjecture is true for $s = 1$, but wrong for $s > 1$ in general.

<u>Example:</u> (Wakolbinger)

$$f_P(x_1, x_2) = \begin{cases} 0 & x_1 < 0 \\ c x_1^{1/4} e^{-x_1 x_2^2} & x_1 \in [0, 1] \\ c e^{-x_1^4 x_2^2} & x_1 > 1, \end{cases}$$

where c is chosen such that $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_P(x_1, x_2) dx_1 dx_2 = 1$.



The density f_P is continuous and bounded. However, F_P is not locally Lipschitz continuous (as the marginal density functions are not bounded).

Proposition:
A probability distribution function F_P with density f_P is locally Lipschitz continuous if its (one-dimensional) marginal density functions f_P^i, $i = 1, \ldots, s$, are locally bounded.
F_P is (globally) Lipschitz continuous iff its marginal density functions are bounded.

$$f_P^i(x_i) := \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_P(x_1, \ldots, x_s) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_s$$

Question: Is there a reasonable class of probability distributions to which the proposition applies?

Definition:

A probability measure P on $\mathbb{R}^s$ is called quasi-concave whenever

$$P(\lambda B + (1-\lambda)\tilde{B}) \geq \min\{P(B), P(\tilde{B})\}$$

holds true for all Borel measurable convex subsets B, $\tilde{B} \subseteq \mathbb{R}^s$ and all $\lambda \in [0,1]$ such that $\lambda B + (1-\lambda)\tilde{B}$ is Borel measurable.

Proposition: (Prekopa)

Let $H : \mathbb{R}^m \to \mathbb{R}^s$ be a set-valued mapping with closed convex graph and P be quasi-concave on $\mathbb{R}^s$. Then the function $G(x) := P(H(x))$ $(x \in \mathbb{R}^m)$ is quasi-concave on $\mathbb{R}^m$. Hence, if X is closed and convex, the feasible set

$$\mathcal{X}_p(P) = \{x \in X : G(x) = P(H(x)) \geq p\}$$

is closed and convex.

Proof: Let $x, \tilde{x} \in \mathbb{R}^m$, $\lambda \in [0,1]$.

$$\begin{aligned} G(\lambda x + (1-\lambda)\tilde{x}) &= P(H(\lambda x + (1-\lambda)\tilde{x})) \geq P(\lambda H(x) + (1-\lambda)H(\tilde{x})) \\ &\geq \min\{P(H(x)), P(H(\tilde{x}))\} = \min\{G(x), G(\tilde{x})\}. \qquad \square \end{aligned}$$

Theorem: (Borell 75)

Assume that the probability distribution on $\mathbb{R}^s$ has a density f_P. Then P is quasi-concave iff $f_P^{-\frac{1}{s}} : \mathbb{R}^s \to \overline{\mathbb{R}}$ is convex.

Examples: (of quasi-concave probability measures)

Multivariate normal distributions $N(m, C)$ (with mean $m \in \mathbb{R}^s$ and $s \times s$ symmetric, positive semidefinite covariance matrix C; nondegenerate or singular), uniform distributions on convex compact subsets of $\mathbb{R}^s$, Dirichlet-, Pareto-, Gamma-distributions etc.

Theorem: (Henrion/Römisch 10)

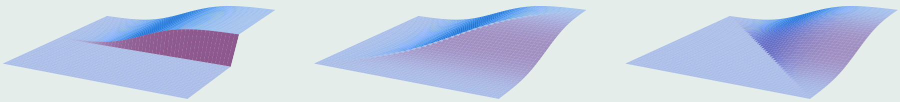
The probability distribution function F_P of a quasi-concave probability measure P on $\mathbb{R}^s$ is Lipschitz continuous iff the support $\operatorname{supp} P$ is not contained in a $(s-1)$-dimensional hyperplane.

Question: Are distribution functions of quasi-concave measures differentiable, too?

Example: (singular normal distributions)
The probability distribution functions F_P of 2-dimensional normal distributions $N(0, C)$ with

$$C = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

are not differentiable on $\mathbb{R}^2$.



Theorem: (Henrion/Römisch 10)

Let ξ be an s-dimensional normal random vector whose covariance matrix is nonsingular. Let F_η denote the probability distribution function of the random vector $\eta = A\xi + b$ where A is an $m \times s$-matrix and $b \in \mathbb{R}^m$.

Then F_η is infinitely many times differentiable at any $\bar{x} \in \mathbb{R}^m$ for which the system $(A, \bar{x} - b)$ satisfies the *Linear Independence Constraint Qualification* (LICQ), i.e., the rows a_i, $i = 1, \ldots, m$, of A satisfy the condition $\operatorname{rank}\{a_i : i \in I\} = \#I$ for every index set $I \in \{1, \ldots, m\}$ such that there exists $z \in \mathbb{R}^s$ with

$$a_i^T z = \bar{x}_i - b_i \quad (i \in I), \quad a_i^T z < \bar{x}_i - b_i \quad (i \in \{1, \ldots, m\} \setminus I).$$

Example:

Our second example of singular normal distributions corresponds to the probability distribution function F_η of

$$\eta = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \xi, \quad \xi \sim N(0, 1).$$

The result implies the C^∞-property of F_η on $R^2 \setminus \{(x, x) : x \in \mathbb{R}\}$.

Let us consider the chance constraint set

$$\mathcal{X}_p(P) = \{x \in \mathbb{R}^m : P(\Xi x \leq a) \geq p\}$$

where Ξ is a stochastic matrix whose rows ξ_i have multivariate normal distributions with mean μ_i and covariance matrix Σ_i, $i = 1, \ldots, r$, and P is the distribution of $(\xi_1, \ldots, \xi_r)$.
For $r = 1$ convexity of $\mathcal{X}_p(P)$ for $p \in [\frac{1}{2}, 1)$ is a classical result.
(van de Panne/Popp 63)

Proposition: (Henrion/Strugarek 08)
Assume that the rows ξ_i of Ξ are pairwise independent.
Then $\mathcal{X}_p$ is convex for $p > \Phi(u^*)$, where Φ is the one-dimensional standard normal distribution function and $u^* \geq \sqrt{3}$ is computable and depends on the means μ_i and the eigenvalues of Σ_i.

Furthermore, the function $G(x) = P(\Xi x \leq a)$ is differentiable and the gradients of G can be explicitly computed if Ξ is Gaussian.
(van Ackooij/Henrion/Möller/Zorgati 11)

Example: (Henrion)

Let P be the standard normal ($N(0,1)$) distribution with probability distribution function

$$F(x) = \frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-\infty}^{x} \exp(-\frac{\xi^2}{2}) d\xi,$$

$A = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and $b(\xi) = \begin{pmatrix} \xi \\ \xi \end{pmatrix}$ for each $\xi \in \mathbb{R}$. Then we have

$$\begin{aligned} G(x) &= P(\{\xi \in \mathbb{R} : Ax \geq b(\xi)\}) \\ &= P(\{\xi \in \mathbb{R} : x \geq \xi, -x \geq \xi\}) = F(\min\{-x, x\}). \end{aligned}$$

Hence, although F is in $C^\infty(\mathbb{R})$, G is non-differentiable.

Hence, tools from nonsmooth analysis should be used for studying the behavior of constraints sets, in general.

Metric regularity of chance constraints

Let $H : \mathbb{R}^m \to \mathbb{R}^s$ be a set-valued mapping with closed graph, $X \subseteq \mathbb{R}^m$ be closed and P be a probability distribution on $\mathbb{R}^s$. We consider the set-valued mapping (from $\mathbb{R}$ to $\mathbb{R}^m$)

$$y \mapsto \mathcal{X}_y(P) = \{x \in X : P(H(x)) \geq y\}.$$

Definition:

The chance constraint function $P(H(\cdot)) - p$ is metrically regular with respect to X at $\bar{x} \in \mathcal{X}_p(P)$ if there exist positive constants a and ε such that

$$d(x, \mathcal{X}_y(P)) \leq a \max\{0, y - P(H(x))\}$$

holds for all $x \in X \cap \mathbb{B}(\bar{x}, \varepsilon)$ and $|p - y| \leq \varepsilon$.

Motivation: Continuity properties of the feasible set $\mathcal{X}_p(P)$ with respect to perturbations of P measured in terms of a suitable distance on the space of all probability distributions on $\mathbb{R}^s$.

The convex case

Proposition: (Römisch/Schultz 91)
Let the set-valued mapping H have closed and convex graph, X be closed and convex, $p \in (0,1)$ and the probability distribution P on $\mathbb{R}^s$ be r-concave for some $r \in (-\infty, +\infty]$. Suppose there exists a Slater point $\bar{x} \in X$ such that $P(H(\bar{x})) > p$.
Then $P(H(\cdot)) - p$ is metrically regular with respect to X at each $x \in \mathcal{X}_p(P)$.

The proof is based on the Robinson-Ursescu theorem applied to the set-valued mapping $\Gamma(x) := \{v \in \mathbb{R} : x \in X, p^r - (P(H(x)))^r \geq v\}$ for some $r < 0$ (w.l.o.g.).

The proposition applies to $H(x) = \{\xi \in \mathbb{R}^s : h(x) \geq \xi\}$, i.e., $P(H(x)) = F_P(h(x))$, where h has concave components. However, even for linear h, i.e., $h(x) = Ax$ the matrix A has to be non-stochastic.

Definition:

A probability measure P on $\mathbb{R}^s$ is called r- concave for some $r \in [-\infty, +\infty]$ if the inequality

$$P(\lambda B + (1-\lambda)\tilde{B}) \geq m_r(P(B), P(\tilde{B}); \lambda)$$

holds for all $\lambda \in [0,1]$ and all convex Borel subsets $B, \tilde{B}$ of $\mathbb{R}^s$ such that $\lambda B + (1-\lambda)\tilde{B}$ is Borel.

Here, the generalized mean function m_r on $\mathbb{R}_+ \times \mathbb{R}_+ \times [0,1]$ for $r \in [-\infty, \infty]$ is given by

$$m_r(a,b;\lambda) := \begin{cases} (\lambda a^r + (1-\lambda)b^r)^{1/r} & , \ r > 0 \text{ or } r < 0, ab > 0, \\ 0 & , \ ab = 0, r < 0, \\ a^\lambda b^{1-\lambda} & , \ r = 0, \\ \max\{a,b\} & , \ r = \infty, \\ \min\{a,b\} & , \ r = -\infty. \end{cases}$$

Notice that $r = -\infty$ corresponds to quasi-concavity.

Optimization problems with stochastic dominance constraints

Optimization model with kth order stochastic dominance constraint

$$\min\{f(x) : x \in D,\, G(x,\xi) \succeq_{(k)} Y\},$$

where $k \in \mathbb{N}$, D is a nonempty convex closed subset of $\mathbb{R}^m$, Ξ a closed subset of $\mathbb{R}^s$, $f : \mathbb{R}^m \to \mathbb{R}$ is convex, ξ is a random vector with support Ξ and Y a real random variable on some probability space both having finite moments of order $k-1$, and $G : \mathbb{R}^m \times \mathbb{R}^s \to \mathbb{R}$ is continuous, concave with respect to the first argument and satisfies the linear growth condition

$$|G(x,\xi)| \leq C(B) \max\{1, \|\xi\|\} \quad (x \in B, \xi \in \Xi)$$

for every bounded subset $B \subset \mathbb{R}^m$ and some constant $C(B)$ (depending on B). The random variable Y plays the role of a benchmark outcome.

D. Dentcheva, A. Ruszczyński: Optimization with stochastic dominance constraints, *SIAM J. Optim.* 14 (2003), 548–566.

Stochastic dominance relation $\succeq_{(k)}$

$$X \succeq_{(1)} Y \quad \Leftrightarrow \quad F_X(\eta) \leq F_Y(\eta) \quad (\forall \eta \in \mathbb{R})$$

where X and Y are real random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. P_X denotes the probability distribution of X and F_X its distribution function, i.e.,

$$F_X(\eta) = \mathbb{P}(\{X \leq \eta\}) = \int_{-\infty}^{\eta} P_X(d\xi) \quad (\forall \eta \in \mathbb{R})$$

Equivalent characterization:

$$X \succeq_{(1)} Y \quad \Leftrightarrow \quad \mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)]$$

for each nondecreasing $u : \mathbb{R} \to \mathbb{R}$ such that the expectations are finite.

Expected utility hypotheses: (von Neumann-Morgenstern)

Outcome X is preferred over outcome Y if and only if

$$\mathbb{E}[u(X)] > \mathbb{E}[u(Y)]$$

for some utility $u(\cdot)$.

$$X \succeq_{(k)} Y \quad \Leftrightarrow \quad F_X^{(k)}(\eta) \leq F_Y^{(k)}(\eta) \quad (\forall \eta \in \mathbb{R})$$

where X and Y are real random variables having moments of order $k-1$ and we define $F_X^{(1)} = F_X$ and recursively

$$F_X^{(k+1)}(\eta) = \int_{-\infty}^{\eta} F_X^{(k)}(\xi) d(\xi) = \int_{-\infty}^{\eta} \frac{(\eta-\xi)^k}{k!} P_X(d\xi)$$
$$= \frac{1}{k!} \| \max\{0, \eta - X\} \|_k^k \quad (\forall \eta \in \mathbb{R}),$$

where

$$\|X\|_k = \left(\mathbb{E}(|X|^k)\right)^{\frac{1}{k}} \quad (\forall k \geq 1).$$

Equivalent characterization of $\succeq_{(2)}$:

$$X \succeq_{(2)} Y \quad \Leftrightarrow \quad \mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)]$$

for each nondecreasing concave $u : \mathbb{R} \to \mathbb{R}$ such that the expectations are finite.

A. Müller and D. Stoyan: *Comparison Methods for Stochastic Models and Risks*, Wiley, Chichester, 2002.

Relaxation, theory and discretization

We consider the relaxed kth order stochastic dominance (SD) constrained optimization model

$$\min \left\{ f(x) : x \in D,\ F^{(k)}_{G(x,\xi)}(\eta) \le F^{(k)}_{Y}(\eta),\ \forall \eta \in I \right\},$$

where $I \subset \mathbb{R}$ is a compact interval.
Split-variable formulation:

$$\min \left\{ f(x) : x \in D,\ G(x,\xi) \ge X,\ F^{(k)}_{X}(\eta) \le F^{(k)}_{Y}(\eta),\ \forall \eta \in I \right\}$$

Since the function $F^{(k)}_X : \mathbb{R} \to \mathbb{R}$ is nondecreasing for $k \ge 1$ and convex for $k \ge 2$, the SD constrained optimization model is a convex semi-infinite program.

Constraint qualification:

kth order uniform dominance condition: There exists $\bar{x} \in D$ such that

$$\min_{\eta \in I} \left(F^{(k)}_{Y}(\eta) - F^{(k)}_{G(\bar{x},\xi)}(\eta) \right) > 0\,.$$

Optimality conditions and duality results can be derived when imposing the kth order uniform dominance condition.

Let X_j and Y_j the scenarios of X and Y with probabilities p_j, $j = 1, \ldots, n$. Then the second order dominance constraints can be expressed as

$$\sum_{j=1}^{n} p_j[\eta - X_j]_+ \le \sum_{j=1}^{n} p_j[\eta - Y_j]_+ \quad \forall \eta \in I.$$

The latter condition can be shown to be equivalent to

$$\sum_{j=1}^{n} p_j[Y_k - X_j)]_+ \le \sum_{j=1}^{n} p_j[Y_k - Y_j]_+ \quad \forall k = 1, \ldots, n.$$

if $Y_k \in I$, $k = 1, \ldots, n$. Here, $[\,\cdot\,]_+ = \max\{0, \,\cdot\,\}$.
Hence, the second order dominance constraints may be reformulated as linear constraints.

D. Dentcheva, A. Ruszczyński: Optimality and duality theory for stochastic optimization problems with nonlinear dominance constraints, *Mathematical Programming* 99 (2004), 329-350.

Stochastic programs with equilibrium constraints

Such optimization models are extensions of two-stage stochastic programs. We consider the SMPEC

$$\min\left\{\inf\{\mathbb{E}[f(x,y,\xi)] : y \in S(x,\xi)\} : x \in X\right\},$$

where $S(x,\xi)$ is the solution set of the variational inequality

$$g(x,y,\xi) \in N_{C(x,\xi)}(y),$$

$f, g : \mathbb{R}^m \times \mathbb{R}^{\bar{m}} \times \mathbb{R}^s \longrightarrow \mathbb{R}$, C is a set-valued mapping from $\mathbb{R}^m \times \mathbb{R}^s$ to $R^{\bar{m}}$ and $N_C(y)$ denotes the normal cone to the set C at y. If we assume that $C(x,\xi)$ is of the form

$$C(x,\xi) = \{y \in \mathbb{R}^{\bar{m}} : h(x,y,\xi) \in V\}$$

with a closed convex cone V in $\mathbb{R}^r$ and a mapping h which is differentiable with respect to y, the variational inequality may be rewritten as

$$-g(x,y,\xi) + \nabla_y h(x,y,\xi)^\top \lambda = 0, \quad \lambda \in N_V(h(x,y,\xi)).$$

The condition $\lambda \in N_V(h(x,y,\xi))$ is equivalent to

$$\lambda \in V^*,\ h(x,y,\xi) \in V,\ \lambda^\top h(x,y,\xi) = 0.$$

or equivalently

$$h(x,y,\xi) \in N_{V^*}(\lambda)$$

Hence, the introduction of the *new variable* λ allows to rewrite the original variational inequality into (Robinson 80)

$$H(x,(y,\lambda),\xi) \in N_K(\lambda),$$

where H maps from $\mathbb{R}^m \times \mathbb{R}^{\bar{m}+r} \times \mathbb{R}^s$ to $\mathbb{R}^{\bar{m}+r}$ and a (fixed) closed convex cone K in $\mathbb{R}^{\bar{m}+r}$ given by

$$H(x,(y,\lambda),\xi) = \begin{pmatrix} -g(x,y,\xi) + \nabla_y h(x,y,\xi)^\top \lambda \\ h(x,y,\xi) \end{pmatrix},\ K = \mathbb{R}^{\bar{m}} \times V^*.$$

Let $\bar{S}(x,\xi) \subset \mathbb{R}^{\bar{m}+r}$ denote the solution set of the previous variational inequality. Then $S(x,\xi)$ equals the projection of $\bar{S}(x,\xi)$ to the first component.

The original SMPEC is equivalent to

$$\min \left\{ \mathbb{E}[f(x,y,\xi)] : (y,\lambda) \in \bar{S}(x,\xi), x \in X \right\}$$

Proposition: (Shapiro, JOTA 06)
Let the functions $f, g, h, \nabla_y h$ be continuous and there exist a P-integrable function w such that

$$\theta(x,\xi) = \inf\{f(x,y,\xi) : (y,\lambda) \in \bar{S}(x,\xi)\} \geq w(\xi)$$

holds for all ξ and all x in a neighborhood of some $\bar{x} \in X$. Assume that the solution set $\bar{S}(x,\xi)$ is nonempty and uniformly bounded (in a neighborhood of $\bar{x}$).
Then the objective $x \mapsto \mathbb{E}[\theta(x,\xi)]$ is (at least) lower semicontinuous at $\bar{x}$.

Under stronger assumptions (Lipschitz) continuity and directional differentiability of the objective may be derived, too.

Multistage stochastic programs

New constraints: Measurability or information constraints

Let $\{\xi_t\}_{t=1}^T$ be a discrete-time stochastic data process defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and with ξ_1 deterministic. The stochastic decision x_t at period t is assumed to be measurable with respect to $\mathcal{F}_t := \sigma(\xi_1, \ldots, \xi_t)$ (nonanticipativity).

Multistage stochastic optimization model:

$$\min \left\{ \mathbb{E}\left[\sum_{t=1}^{T} \langle b_t(\xi_t), x_t \rangle\right] \middle| \begin{array}{l} x_t \in X_t, t = 1, \ldots, T, A_{1,0}x_1 = h_1(\xi_1), \\ x_t \text{ is } \mathcal{F}_t\text{-measurable}, t = 1, \ldots, T, \\ A_{t,0}x_t + A_{t,1}(\xi_t)x_{t-1} = h_t(\xi_t), t = 2, ., T \end{array} \right\}$$

where the sets X_t, $t = 1, \ldots, T$, are polyhedral cones, the vectors $b_t(\cdot)$, $h_t(\cdot)$ and $A_{t,1}(\cdot)$ are affine functions of ξ_t, where ξ varies in a polyhedral set Ξ.

If the process $\{\xi_t\}_{t=1}^T$ has a finite number of scenarios, they exhibit a scenario tree structure.

To have the model well defined, we assume
$x_t \in L_{r'}(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{m_t})$ and $\xi_t \in L_r(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^d)$,
where $r \geq 1$ and

$$r' := \begin{cases} \frac{r}{r-1} & , \text{ if only costs are random} \\ r & , \text{ if only right-hand sides are random} \\ \infty & , \text{ if all technology matrices are random and } r = T. \end{cases}$$

Then nonanticipativity may be expressed as

$$x \in \mathcal{N}_{na}$$

$$\mathcal{N}_{na} = \{x \in \times_{t=1}^{T} L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{m_t}) : x_t = \mathbb{E}[x_t|\mathcal{F}_t]\,,\ \forall t\},$$

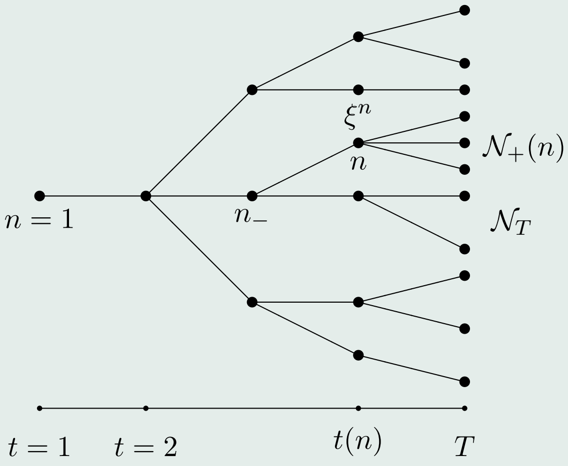
i.e., as a subspace constraint, by using the conditional expectation $\mathbb{E}[\cdot\,|\mathcal{F}_t]$ with respect to the σ-algebra $\mathcal{F}_t$.

For $T = 2$ we have $\mathcal{N}_{na} = \mathbb{R}^{m_1} \times L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{m_2})$.

$\longrightarrow$ infinite-dimensional (linear) optimization problem

Data process approximation by scenario trees

The process $\{\xi_t\}_{t=1}^T$ is approximated by a process forming a scenario tree based on a finite set of scenarios and nodes $\mathcal{N} \subset \mathbb{N}$.



Scenario tree with $T = 5$, $N = 22$ and 11 leaves

$n = 1$ root node, n_- unique predecessor of node n, $\text{path}(n) = \{1, \ldots, n_-, n\}$, $t(n) := |\text{path}(n)|$, $\mathcal{N}_+(n)$ set of successors to n, $\mathcal{N}_T := \{n \in \mathcal{N} : \mathcal{N}_+(n) = \emptyset\}$ set of leaves, $\text{path}(n)$, $n \in \mathcal{N}_T$, scenario with (given) probability π^n, $\pi^n := \sum_{\nu \in \mathcal{N}_+(n)} \pi^\nu$ probability of node n, ξ^n realization of $\xi_{t(n)}$.

Tree representation of the optimization model

$$\min\left\{\sum_{n\in\mathcal{N}}\pi^n\langle b_{t(n)}(\xi^n),x^n\rangle \,\middle|\, \begin{array}{l} x^n\in X_{t(n)}, n\in\mathcal{N}, A_{1,0}x^1=h_1(\xi^1) \\ A_{t(n),0}x^n+A_{t(n),1}x^{n-}=h_{t(n)}(\xi^n), n\in\mathcal{N}\end{array}\right\}$$

How to solve the optimization model ?

- Standard software (e.g., CPLEX)

- Decomposition methods for (very) large scale models
  (Ruszczynski/Shapiro (Eds.): Stochastic Programming, Handbook, 2003)

Open question:

How to generate (multivariate) scenario trees ?

Dynamic programming

Theorem: (Evstigneev 76, Rockafellar/Wets 76)
Under weak assumptions the multistage stochastic program is equivalent to the (first-stage) convex minimization problem

$$\min \left\{ \int_{\Xi} f(x_1,\xi)P(d\xi) : x_1 \in \mathcal{X}_1(\xi_1) \right\},$$

where f is an integrand on $\mathbb{R}^{m_1} \times \Xi$ given by

$$f(x_1,\xi) := \langle b_1(\xi_1), x_1\rangle + \Phi_2(x_1,\xi^2),$$

$$\Phi_t(x_1,\ldots,x_{t-1},\xi^t) := \inf \Big\{ \langle b_t(\xi_t), x_t\rangle + \mathbb{E}\big[\Phi_{t+1}(x_1,\ldots,x_t,\xi^{t+1})|\mathcal{F}_t\big] : x_t \in X_t, A_{t,0}x_t + A_{t,1}(\xi_t)x_{t-1} = h_t(\xi_t) \Big\}$$

for $t = 2,\ldots,T$, where $\Phi_{T+1}(x_1,\ldots,x_T,\xi^{T+1}) := 0$, $\mathcal{X}_1(\xi_1) := \{x_1 \in X_1 : A_{1,0}x_1 = h_1(\xi_1)\}$ and $P \in \mathcal{P}(\Xi)$ is the probability distribution of ξ.

$\longrightarrow$The integrand f depends on the probability measure $\mathbb{P}$ in a non-linear way !

Risk Functionals

A risk functional or risk measure ρ assigns a real number to any (real) random variable Y (possibly satisfying certain moment conditions). Recently, it was suggested that ρ should satisfy the following axioms for all random variables Y, $\tilde{Y}$, $r \in \mathbb{R}$, $\lambda \in [0,1]$:

(A1) $\rho(Y + r) = \rho(Y) - r$ (translation-invariance),

(A2) $\rho(\lambda Y + (1-\lambda)\tilde{Y}) \leq \lambda\rho(Y) + (1-\lambda)\rho(\tilde{Y})$ (convexity),

(A3) $Y \leq \tilde{Y}$ implies $\rho(Y) \geq \rho(\tilde{Y})$ (monotonicity).

A risk functional ρ is called coherent if it is, in addition, positively homogeneous, i.e., $\rho(\lambda Y) = \lambda\rho(Y)$ for all $\lambda \geq 0$ and random variables Y.

Given a risk functional ρ, the mapping $\mathcal{D} = \mathbb{E} + \rho$ is also called deviation risk functional.

References: Artzner-Delbaen-Eber-Heath 99, Föllmer-Schied 02, Fritelli-Rosazza Gianin 02

Examples:

(a) Conditional Value-at-Risk or Average Value-at-Risk $\mathbb{AV@R}_\alpha$:

$$\mathbb{AV@R}_\alpha(Y) := \frac{1}{\alpha}\int_0^\alpha \mathbb{V@R}_u(Y)(u)du = \frac{1}{\alpha}\int_0^\alpha G^{-1}(u)du$$
$$= \inf\left\{x + \frac{1}{\alpha}\mathbb{E}([Y+x]_-) : x \in \mathbb{R}\right\}$$
$$= \sup\left\{-\mathbb{E}(YZ) : \mathbb{E}(Z) = 1, 0 \le Z \le \frac{1}{\alpha}\right\}$$

where $\alpha \in (0,1]$, $\mathbb{V@R}_\alpha := \inf\{y \in \mathbb{R} : \mathbb{P}(Y \le y) \ge \alpha\}$ is the Value-at-Risk, $[a]_- := -\min\{0, a\}$ and G the distribution function of Y.

Reference: Rockafellar-Uryasev 02

(b) Lower semi standard deviation corrected expectation:

$$\rho(Y) := -\mathbb{E}(Y) + (\mathbb{E}([Y - \mathbb{E}(Y)]_-)^2)^{\frac{1}{2}}$$

Reference: Markowitz 52

Multiperiod risk measurement

Let $\mathfrak{F} = \{\mathcal{F}_t : t = 1, \dots, T\}$ be a filtration generated by some stochastic process on $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathcal{F}_1 = \{\emptyset, \Omega\}$.
A functional $\rho_{\mathfrak{F}}$ on $\mathcal{Z} = \times_{t=1}^{T} L_p(\Omega, \mathcal{F}, \mathbb{P})$ is called a multiperiod risk measure if the following conditions (i)–(iii) hold:

(i) Monotonicity: if $z_t \leq \tilde{z}_t$ a.s, $t = 1, \dots, T$, then
$\rho_{\mathfrak{F}}(z_1, \dots, z_T) \geq \rho_{\mathfrak{F}}(\tilde{z}_1, \dots, \tilde{z}_T)$;

(ii) Translation invariance: for each $r \in \mathbb{R}$ we have
$\rho_{\mathfrak{F}}(z_1 + r, \dots, z_T + r) = \rho_{\mathfrak{F}}(z_1, \dots, z_T) - r$;

(iii) Convexity: for each $\lambda \in [0, 1]$ and $z, \tilde{z} \in \mathcal{Z}$ we have
$\rho_{\mathfrak{F}}(\lambda z + (1 - \lambda)\tilde{z}) \leq \lambda \rho_{\mathfrak{F}}(z) + (1 - \lambda)\rho_{\mathfrak{F}}(\tilde{z})$.

It is called coherent if in addition condition (iv) holds:

(iv) Positive homogeneity: for each $\lambda \geq 0$ we have
$\rho_{\mathfrak{F}}(\lambda z_1, \dots, \lambda z_T) = \lambda \rho_{\mathfrak{F}}(z_1, \dots, z_T)$.

(Artzner-Delbaen-Eber-Heath-Ku 07)

A multiperiod risk measure $\rho_{\mathfrak{F}}$ is called information monotone if $\mathfrak{F} \subseteq \mathfrak{F}'$ (i.e. $\mathcal{F}_t \subseteq \mathcal{F}'_t$, $t = 1, \ldots, T$) implies

$$\rho_{\mathfrak{F}'}(z) \leq \rho_{\mathfrak{F}}(z) \quad \forall z \in \mathcal{Z}.$$

A multiperiod risk measure $\rho_{\mathfrak{F}}$ is time consistent if it is constructed by conditional risk mappings $\rho_t(\cdot|\mathfrak{F}^{(t)})$ from $\times_{\tau=t}^{T} L_p(\Omega, \mathcal{F}_t, \mathbb{P})$ to $L_p(\Omega, \mathcal{F}_t, \mathbb{P})$ with $\mathfrak{F}^{(t)} = \{\mathcal{F}_t, \ldots, \mathcal{F}_T\}$, $t = 1, \ldots, T$, such that $\rho_{\mathfrak{F}}(z) = \rho_1(z|\mathfrak{F}^{(1)})$ and if the conditions

$$\rho_t(z^{(t)}|\mathfrak{F}^{(t)}) \geq \rho_t(\tilde{z}^{(t)}|\mathfrak{F}^{(t)}) \text{ and } z_{t-1} \leq \tilde{z}_{t-1}$$

imply $\rho_{t-1}(z^{(t-1)}|\mathfrak{F}^{(t-1)}) \geq \rho_{t-1}(\tilde{z}^{(t-1)}|\mathfrak{F}^{(t-1)})$ for all $t = 2, \ldots, T$.

Remark:

There appear different requirements in the literature instead of the translation invariance (ii).

(e.g. Fritelli-Scandalo 06, Pflug-Römisch 07)

Theorem: (dual representation)
Let $\rho_{\mathfrak{F}} : \times_{t=1}^{T} L_p(\Omega, \mathcal{F}, \mathbb{P}) \to \bar{\mathbb{R}}$ be proper (i.e. $\rho_{\mathfrak{F}}(z) > -\infty$ and $\operatorname{dom} \rho_{\mathfrak{F}} = \{z : \rho(z) < \infty\} \neq \emptyset$) and lower semicontinuous. Then $\rho_{\mathfrak{F}}$ is a multiperiod convex risk measure if and only if it admits the representation

$$\rho_{\mathfrak{F}}(z) = \sup \left\{ -\mathbb{E}\Big[\sum_{t=1}^{T} \lambda_t z_t\Big] - \rho_{\mathfrak{F}}^*(\lambda) \; : \; \lambda \in \mathcal{P}_\rho(\mathfrak{F}) \right\},$$

where

$$\mathcal{P}_\rho(\mathfrak{F}) \subseteq \mathcal{D}_T = \left\{ \lambda \in \times_{t=1}^{T} L_q(\Omega, \mathcal{F}_t, \mathbb{P}) : \lambda_t \geq 0, \sum_{t=1}^{T} \mathbb{E}[\lambda_t] = 1 \right\}$$

with $\frac{1}{p} + \frac{1}{q} = 1$ is closed and convex, and $\rho_{\mathfrak{F}}^*$ is the conjugate of $\rho_{\mathfrak{F}}$. The functional $\rho_{\mathfrak{F}}$ is a multiperiod coherent risk measure if and only if the conjugate $\rho_{\mathfrak{F}}^*$ is the indicator function of $\mathcal{P}_\rho(\mathfrak{F})$.

(Ruszczyński-Shapiro 06)

Multiperiod extended polyhedral risk measures

A multiperiod risk measure $\rho_{\mathfrak{F}}$ on $\mathcal{Z}$ is called extended polyhedral if there exist matrices $A_t, B_{t,\tau}$, vectors a_t, c_t, and functions $h_t(z) = (h_{t,1}(z), \ldots, h_{t,n_{t,2}}(z))^\top$ with $h_{t,i} : \mathcal{Z} \to \mathcal{Z}$ such that

$$\rho_{\mathfrak{F}}(z) = \inf \left\{ \mathbb{E}\Big[\sum_{t=1}^{T} c_t^\top y_t\Big] \;\middle|\; \begin{array}{l} y_t \in L_p(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{k_t}), A_t y_t \leq a_t \\ \sum_{\tau=0}^{t-1} B_{t,\tau} y_{t-\tau} = h_t(z_t) \\ (t = 1, \ldots, T) \end{array} \right\}$$

(Guigues-Römisch, SIOPT 12)

Motivation: Characterizing the largest class of multiperiod risk measures that maintains important theoretical and algorithmic properties when incorporated into (linear) multistage stochastic programs instead of the expectation functional.
Most important case: h_t affine.

First version: $a_t = 0$, $B_{t,\tau}$ row vectors, h_t identity
(Eichhorn-Römisch 05)

Examples of multiperiod extended polyhedral risk measures

Let increasing risk measuring time steps t_j, $j = 1, \ldots, J$, with $t_J = T$, and weights $\gamma_j \geq 0$, $j = 1, \ldots, J$, with $\sum_{j=1}^{J} \gamma_j = 1$ be given.

(a) Weighted sum of Average Value-at-Risk at risk measuring time steps:

$$\rho_s(z) := \sum_{j=1}^{J} \gamma_j \mathbb{A}\mathsf{V@R}_\alpha(z(t_j)),$$

where $\mathbb{A}\mathsf{V@R}_\alpha(z) = \inf_{r\in\mathbb{R}}[r + \frac{1}{\alpha}\mathbb{E}[z + r]^-]$.

(c) Average Value-at-Risk of the weighted average at risk measuring time steps:

$$\rho_a(z) := \mathbb{A}\mathsf{V@R}_\alpha\Big(\sum_{j=1}^{J} \gamma_j z(t_j)\Big)$$

(d) Average Value-at-Risk of the minimum at risk measuring time steps:

$$\rho_m(z) := \mathbb{A}\mathsf{V@R}_\alpha\Big(\min_{j=1,\ldots,J} z(t_j)\Big)$$

Risk-averse multistage stochastic optimization model:

$$\min_{x} \left\{ \rho(z) \;\middle|\; \begin{array}{l} z_t = \sum_{\tau=1}^{t} b_\tau(\xi_\tau)^\top x_\tau \\ x_t \in X_t, x_t \in L_p(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{m_t}) \\ \sum_{\tau=0}^{t-1} A_{t,\tau}(\xi_t) x_{t-\tau} = g_t(\xi_t) \\ (t = 1, \ldots, T) \end{array} \right\}$$

Multiperiod extended polyhedral risk functional:

$$\rho(z) = \inf \left\{ \mathbb{E}\Big[\sum_{t=1}^{T} c_t^\top y_t\Big] \;\middle|\; \begin{array}{l} y_t \in L_p(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{k_t}) \\ A_t y_t \le a_t \\ \sum_{\tau=0}^{t-1} B_{t,\tau} y_{t-\tau} = h_t(z_t) \\ (t = 1, \ldots, T) \end{array} \right\}$$

Equivalent risk-neutral multistage stochastic optimization model:

$$\min_{(y,x)} \left\{ \mathbb{E}\Big[\sum_{t=1}^{T} c_t^\top y_t\Big] \;\middle|\; \begin{array}{l} y_t \in L_p(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{k_t}), x_t \in L_p(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{m_t}) \\ A_t y_t \le a_t,\; x_t \in X_t \\ \sum_{\tau=0}^{t-1} B_{t,\tau} y_{t-\tau} = h_t(\sum_{\tau=1}^{t} b_\tau(\xi_\tau)^\top x_\tau) \\ \sum_{\tau=0}^{t-1} A_{t,\tau}(\xi_t) x_{t-\tau} = g_t(\xi_t) \\ (t = 1, \ldots, T) \end{array} \right\}$$

Conditional risk mappings

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{F}_1$ be a σ-field contained in $\mathcal{F}$. Let $\mathcal{Y} = L_p(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{Y}_1 = L_p(\Omega, \mathcal{F}_1, \mathbb{P})$ for some $p \in [1, +\infty)$, hence $\mathcal{Y}_1 \subseteq \mathcal{Y}$. All (in)equalities between random variables in $\mathcal{Y}$ are intended to hold $\mathbb{P}$-almost surely.

A mapping $\rho : \mathcal{Y} \to \mathcal{Y}_1$ is called conditional risk mapping (with observable information $\mathcal{F}_1$) if the following conditions are satisfied for all Y, $\tilde{Y} \in \mathcal{Y}$, $Y^{(1)} \in \mathcal{Y}_1$, $\lambda \in [0, 1]$:

(i) $\rho(Y + Y^1) = \rho(Y) - Y^{(1)}$ (predictable translation-invariance),

(ii) $\rho(\lambda Y + (1 - \lambda)\tilde{Y}) \leq \lambda\rho(Y) + (1 - \lambda)\rho(\tilde{Y})$ (convexity),

(iii) $Y \leq \tilde{Y}$ implies $\rho(Y) \geq \rho(\tilde{Y})$ (monotonicity).

The conditional risk mapping ρ is called
positively homogeneous if $\rho(\lambda Y) = \lambda\rho(Y)$, $\forall \lambda > 0$.
lower semicontinuous if $\mathbb{E}(\rho(\cdot)\mathbb{1}_B) : \mathcal{Y} \to \overline{\mathbb{R}}$ is lower semicontinuous for every $B \in \mathcal{F}_1$.

Examples:

(a) Conditional expectation: The defining equation for the conditional expectation $\mathbb{E}(\cdot\,|\mathcal{F}_1)$ is

$$\mathbb{E}(\mathbb{E}(Y|\mathcal{F}_1)\,\mathbb{1}_B) = \mathbb{E}(Y\mathbb{1}_B) \quad (\forall B \in \mathcal{F}_1).$$

It is a mapping from $L_p(\mathcal{F})$ onto $L_p(\mathcal{F}_1)$ for $p \in [1, \infty)$.

(b) Conditional average value-at-risk: $\rho(Y|\mathcal{F}_1) = \mathbb{A}\mathsf{V@R}_\alpha(Y\,|\mathcal{F}_1)$ is defined on $L_1(\mathcal{F})$ by the relation

$$\mathbb{E}(\rho(Y|\mathcal{F}_1)\mathbb{1}_B) = \sup\{-\mathbb{E}(YZ) : 0 \leq Z \leq \frac{1}{\alpha}\mathbb{1}_B, \mathbb{E}(Z|\mathcal{F}_1) = \mathbb{1}_B\}$$

for every $B \in \mathcal{F}_1$. The mapping $Y \mapsto \mathbb{A}\mathsf{V@R}_\alpha(Y|\mathcal{F}_1)$ is positively homogeneous, continuous and satisfies (i)–(iii).

Composition of conditional risk mappings

Let a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a filtration $\mathfrak{F} = (\mathcal{F}_0, \dots, \mathcal{F}_T)$ of σ-fields $\mathcal{F}_t$, $t = 0, ..., T$, with $\mathcal{F}_T = \mathcal{F}$ be given. We consider the Banach spaces $\mathcal{Y}_t := L_p(\mathcal{F}_t)$ of $\mathcal{F}_t$-measurable (real) random variables for $t = 1, \dots, T$ and some $p \in [1, +\infty)$.

Let conditional risk mappings $\rho_{t-1} := \rho(\cdot\,|\mathcal{F}_{t-1})$ from $\mathcal{Y}_T$ to $\mathcal{Y}_{t-1}$ be given for each $t = 1, \dots, T$.
We introduce a multi-period risk functional ρ on $\mathcal{Y} := \times_{t=1}^{T} \mathcal{Y}_t$ by nested compositions and a family $(\rho^{(t)})_{t=1}^{T}$ of single-period risk functionals $\rho^{(t)}$ by compositions of the conditional risk mappings ρ_{t-1}, $t = 1, \dots, T$, namely,

$$\begin{aligned} \rho(Y; \mathfrak{F}) &:= \rho_0[Y_1 + \cdots + \rho_{T-2}[Y_{T-1} + \rho_{T-1}(Y_T)] \cdots] \\ \rho^{(t)}(Y_T) &:= \rho_0 \circ \rho_1 \circ \cdots \circ \rho_{t-1}(Y_T) \end{aligned}$$

for every $Y \in \mathcal{Y}$ and $Y_T \in \mathcal{Y}_T$.

(Ruszczyński-Shapiro, Math. OR 06)

Proposition: (Ruszczyński-Shapiro)
Then $\rho(\cdot;\mathfrak{F}):\mathcal{Y}\to\overline{\mathbb{R}}$ is a multi-period risk functional and every $\rho^{(t)}:\mathcal{Y}_T\to\mathbb{R}$ is a (single-period) risk functional. Moreover, it holds

$$\rho(Y;\mathfrak{F})=\rho^{(T)}(Y_1+\cdots+Y_T).$$

The functionals ρ and $\rho^{(t)}$, $t=1,\ldots,T$, are positively homogeneous if all ρ_t are positively homogeneous.

Example:
We consider the conditional average value-at-risk (of level $\alpha\in(0,1]$) as conditional risk mapping

$$\rho_{t-1}(Y_t):=\mathbb{A}\mathsf{V@R}_\alpha(\cdot\,|\mathcal{F}_{t-1})$$

for every $t=1,\ldots,T$. Then

$$n\mathbb{A}\mathsf{V@R}_\alpha(Y;\mathfrak{F})=\mathbb{A}\mathsf{V@R}_\alpha(\cdot\,|\mathcal{F}_0)\circ\cdots\circ\mathbb{A}\mathsf{V@R}_\alpha(\cdot\,|\mathcal{F}_{T-1})\Big(\sum_{t=1}^{T}Y_t\Big)$$

is a multi-period risk functional and is called nested average value-at-risk.

Proposition: (Pflug-Römisch 07)
The nested $n\mathbb{A}\text{V@R}$ has the following dual representation:

$$n\mathbb{A}\text{V@R}_\alpha(Y;\mathfrak{F}) = \sup\{-\mathbb{E}[(Y_1+\cdots+Y_T)Z_T] : 0 \le Z_t \le \frac{1}{\alpha}Z_{t-1}, \mathbb{E}(Z_t|\mathcal{F}_{t-1}) = Z_{t-1}, Z_0 = 1, t = 1,\ldots,T\}.$$

The (dual) process (Z_t) is a martingale and $n\mathbb{A}\text{V@R}$ is not polyhedral and not information monotone, but given by a linear stochastic program (with functional constraints).

Risk-averse multistage stochastic programs:
Replace the conditional expectation in the dynamic programming representation by conditional risk mappings $\rho(\cdot\,|\mathcal{F}_t)$

$$\Phi_t(x_1,\ldots,x_{t-1},\xi^t) := \inf\Big\{\langle b_t(\xi_t),x_t\rangle + \rho\big(\Phi_{t+1}(x_1,\ldots,x_t,\xi^{t+1})|\mathcal{F}_t\big) : x_t \in X_t, A_{t,0}x_t + A_{t,1}(\xi_t)x_{t-1} = h_t(\xi_t)\Big\}$$

for $t = 2,\ldots,T$, where $\Phi_{T+1}(x_1,\ldots,x_T,\xi^{T+1}) := 0$.

Contents (Part II)