Stochastic Programming:
A Variational Analysis Perspective

W. Römisch

Humboldt-University Berlin
Institute of Mathematics
10099 Berlin, Germany

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

Spring School Variational Analysis, Paseky, April 24 - 29, 2006

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, J. Mayer/P. Kall 2005

A. Ruszczynski/A. Shapiro (eds.), Stochastic Programming, Handbook, Elsevier, 2003

S.W. Wallace/W.T. Ziemba (eds.), Applications of Stochastic Programming, MPS-SIAM Series on Optimization, 2005.

Contents:

Part I

Chance Constraints and Nonsmooth Analysis

(R. Henrion (Berlin))

Optimization models under stochastic uncertainty

Let us consider the optimization model

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

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

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

Main approaches:

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

- Replace the random constraints by the constraint

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

  where $p \in [0, 1]$ denotes a probability level, or go back to the *modeling stage* and introduce a recourse action to compensate violations of the constraint.

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

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

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

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

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

Properties of chance constraints

Special forms of chance constraints:

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

  $$G(x) = \mathbb{P}(\xi \leq h(x)) = F_\mu(h(x)) \geq p\,,$$

  where $F_\mu(y) := \mathbb{P}(\{\omega \in \Omega : \xi(\omega) \leq y\}) = \mu(\{\xi \in \Xi : \xi \leq y\})$ $(y \in \mathbb{R}^s)$ denotes the (multivariate) probability distribution function of ξ and $\mu := \mathbb{P} \cdot \xi^{-1}$ its probability distribution.

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

  $$G(x) := \mu(\{\xi \in \Xi : A(\xi)x \geq b(\xi)\}) \geq p.$$

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

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

is closed.

What about continuity and differentiability properties of G or convexity of $\mathcal{X}_p(\mu)$?

Examples:

(i) Let $H(x) = x + \mathbb{R}^s_-$ $(\forall x \in \mathbb{R}^s)$ and μ be discrete with finite support, i.e., $\mu = \sum_{i=1}^n p_i \delta_{\xi_i}$, where δ_ξ denotes the Dirac measure placing unit mass at ξ and $p_i > 0$, $i = 1, \ldots, n$, $\sum_{i=1}^n p_i = 1$.
Then $\mathcal{X}_p(\mu) = X \cap (\cup_{i \in I}(\xi_i + \mathbb{R}^s_+))$ holds with some index set $I \subset \{1, \ldots, n\}$ and, hence, it is non-convex in general.
Moreover, $G = F_\mu$ 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 μ have a density f_μ with respect to the Lebesgue measure on $\mathbb{R}^s$, i.e.,

$$G(x) = F_\mu(x) = \int_{-\infty}^{x} f_\mu(y)dy = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_s} f_\mu(y_1, \ldots, y_s)dy_s \cdots dy_1.$$

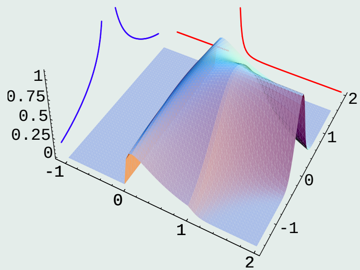
Conjecture: $G = F_\mu$ is Lipschitz continuous if the density f_μ is continuous and bounded.

Answer: The conjecture is true for $s = 1$, but holds no longer for $s > 1$ in general.

Example: (A. Wakolbinger)

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

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



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

Proposition:

A probability distribution function F_μ with density f_μ is locally Lipschitz continuous if its (one-dimensional) marginal density functions f_μ^i, $i = 1, \ldots, s$, are locally bounded.

F_μ is (globally) Lipschitz continuous iff its marginal density functions are bounded.

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

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

Definition:

A probability measure $\mu \in \mathcal{P}(\mathbb{R}^s)$ is called quasi-concave whenever

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

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)

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

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

is closed and convex.

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

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

Theorem: (Borell 75)
If $\mu \in \mathcal{P}(\mathbb{R}^s)$ is quasi-concave and has a density f_μ, the function $f_\mu^{-\frac{1}{s}} : \mathbb{R}^s \to \overline{\mathbb{R}}$ is convex.

Theorem: (Henrion/Römisch 05)
The probability distribution function F_μ of a quasi-concave probability measure $\mu \in \mathcal{P}(\mathbb{R}^s)$ is Lipschitz continuous iff $\operatorname{supp} \mu$ is not contained in a $(s-1)$-dimensional hyperplane.

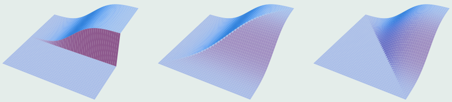
Question: Are distribution functions of quasi-concave measures differentiable, too ?

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.

Example: (singular normal distributions)
The probability distribution functions F_μ 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 $\mathbb{R}^2$.



Home Page
Title Page
Contents
Page 14 of 66
Go Back
Full Screen
Close
Quit

Theorem: (Henrion/Römisch 05)
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 $b \in I\!R^m$.
Then F_η is infinitely many times differentiable at any $\bar{x} \in I\!R^m$ 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 $z \in I\!R^s$ 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 $R^2 \setminus \{(x,x) : x \in I\!R\}$.

Example: (Henrion)

Let $\mu \in \mathcal{P}(I\!R)$ be the standard normal $(N(0,1))$ distribution with probability distribution function

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

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

$$\begin{aligned} G(x) &= \mu(\{\xi \in I\!R : Ax \geq b(\xi)\}) \\ &= \mu(\{\xi \in I\!R : x \geq \xi, -x \geq \xi\}) = \Phi(\min\{-x, x\}). \end{aligned}$$

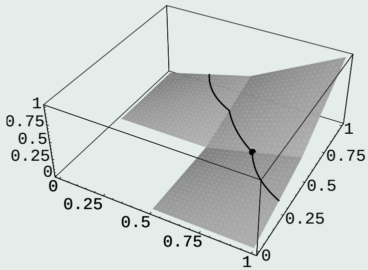
Hence, although Φ is in $C^\infty(I\!R)$, G is non-differentiable.

Example: (Henrion/Römisch 99)

Let $m = s = 2$, $X = [0,2] \times [0,2]$, $A := I$, $p = 1/6$ and μ be the uniform distribution on $\Xi := ([0,1] \times [0,1]) \setminus ([0,1/2] \times [0,1/2])$. Around the feasible point $\bar{x} = (3/4, 1/2)$ (the probability level is binding at $\bar{x}$) the constraint function is of the form

$$G(x) := F_\mu(x) = 4/3 \max\{x_2(x_1 - 1/2), x_1(x_2 - 1/2), x_1 x_2 - 1/4\}$$

and is non-differentiable at $\bar{x}$, although $\bar{x}$ lies in the interior of the support of the underlying constant density. Note that μ is not quasi-concave since the support of μ is non-convex.



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 $\mu \in \mathcal{P}(\mathbb{R}^s)$. We consider the set-valued mapping (from $\mathbb{R}$ to $\mathbb{R}^m$)

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

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

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

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

Motivation: Continuity properties of the feasible set $\mathcal{X}_p(\mu)$ with respect to perturbations of $\mu \in \mathcal{P}(\mathbb{R}^s)$ measured in terms of a suitable distance on $\mathcal{P}(\mathbb{R}^s)$, e.g., the $\mathcal{B}$-discrepancy
$\alpha_{\mathcal{B}}(\mu, \nu) := \sup_{B \in \mathcal{B}} |\mu(B) - \nu(B)|$ with $\mathcal{B} := \{H(x) : x \in X\}$.

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 $\mu \in \mathcal{P}(\mathbb{R}^s)$ be r-concave for some $r \in (-\infty, +\infty]$. Suppose there exists a Slater point $\bar{x} \in X$ such that $\mu(H(\bar{x})) > p$.
Then $\mu(H(\cdot)) - p$ is metrically regular with respect to X at each $x \in \mathcal{X}_p(\mu)$.

The proof is based on the Robinson-Ursescu theorem applied to the set-valued mapping $\Gamma(x) := \{v \in \mathbb{R} : x \in X, p^r - (\mu(H(x)))^r \geq v\}$ 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., $\mu(H(x)) = F_\mu(h(x))$, where h has concave components. However, even for $h(x) = Ax$ the matrix A has to be non-stochastic.
For stochastic A there exist only specific results (Henrion/Strugarek 06).
Metric regularity results for the general case are an open problem.

Definition:

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

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

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.

A non-convex and non-differentiable situation

Proposition: (Henrion/Römisch 99)
Let $\mu \in \mathcal{P}(\mathbb{R}^s)$ have a density f_μ on $\mathbb{R}^s$, $p \in (0,1)$ and $h : \mathbb{R}^m \to \mathbb{R}^s$ be locally Lipschitz continuous.
Then $F_\mu(h(\cdot)) - p$ is metrically regular with respect to X at $\bar{x} \in \mathcal{X}_p(\mu)$ if the following conditions are satisfied:

(i) $(h(\bar{x}) + \mathrm{bd}\, \mathbb{R}^s_-) \cap \mathcal{D}^+ \neq \emptyset$ if $F_\mu(h(\bar{x})) = p$, where
$\mathcal{D}^+ := \{\xi \in \mathbb{R}^s : \exists \varepsilon > 0$ such that $f_\mu(z) \geq \varepsilon, \forall z \in \mathbb{B}(\xi, \varepsilon)\}$.

(ii) $\partial_a \langle y^*, h \rangle(\bar{x}) \cap (-N_a(X; \bar{x})) = \emptyset$, $\forall y^* \in \mathbb{R}^s_- \setminus \{0\}$, where
N_a and ∂_a denote the approximate normal cone and subdifferential (of Mordukhovich), respectively.

(decomposition into growth condition and constraint qualification)

Corollary:
If X is closed and convex, and all components of h are concave, condition (ii) is satisfied if there exists $\hat{x} \in X$ with $h(\hat{x}) > h(\bar{x})$.

Illustration of quantitative stability results

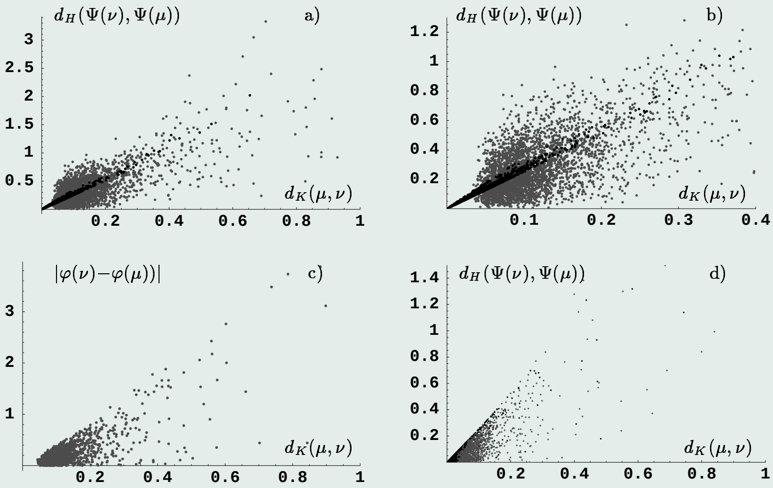
We consider the 2-dimensional example

$$\min\{x_1 + x_2 | I\!P(\xi_1 \le x_1, \xi_2 \le x_2) \ge 1/2\},$$

where ξ is assumed to have a distribution μ which is normal with independent $N(0,1)$ components. The solution set consists of a singleton $\Psi(\mu) = \{(q,q)\}$, where $q \approx 0.55$ is the $1/\sqrt{2}$-quantile of the $N(0,1)$ distribution.

We consider two specific approximations of μ:
(i) The empirical measure $\nu = N^{-1} \sum_{i=1}^{N} \delta_{\xi_i}$, where ξ_i, $i = 1, \ldots, N$ are i.i.d. observations of ξ.
(ii) A parametric estimate for the mean m and the covariance matrix C in $\nu \sim N(m, C)$.

The figure shows $d_H(\Psi(\mu), \Psi(\nu))$ relative to the Kolmogorov distance $d_K(\mu, \nu) = \sup_{\xi \in I\!R^2} |F_\mu(\xi) - F_\nu(\xi)|$. The grey dots correspond to the empirical estimates and the black dots to the parametric estimates.


$d_H(\Psi(\nu),\Psi(\mu))$
a)
3
2.5
2
1.5
1
0.5
$d_K(\mu,\nu)$
0.2 0.4 0.6 0.8 1
$d_H(\Psi(\nu),\Psi(\mu))$
b)
1.2
1
0.8
0.6
0.4
0.2
$d_K(\mu,\nu)$
0.1 0.2 0.3 0.4
$|\varphi(\nu)-\varphi(\mu))|$
c)
3
2
1
$d_K(\mu,\nu)$
0.2 0.4 0.6 0.8 1
$d_H(\Psi(\nu),\Psi(\mu))$
d)
1.4
1.2
1
0.8
0.6
0.4
0.2
$d_K(\mu,\nu)$
0.2 0.4 0.6 0.8 1

Part II

Solution Estimates for Two-Stage Models

(S. T. Rachev (Karlsruhe), R. J-B Wets (Davis))

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

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

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

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

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

Two formulations of two-stage models

Deterministic equivalent of the two-stage model:

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

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

Infinite-dimensional optimization model:

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

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 $I\!\!E[\langle q(\xi), y(\xi)\rangle]$ is finite. For example, if the recourse matrix is fixed (i.e., $W(\xi) \equiv W$), $r = \frac{p}{p-1}$ is consistent.

Structural properties of two-stage models

We consider the infimum function of the parametrized linear (second-stage) program and the dual feasible set of the second-stage program, namely,

$$\Phi(\xi,u,t):=\inf\{\langle u,y\rangle : W(\xi)y=t,\, y\in Y\}\;((\xi,u,t)\in\Xi\times\mathbb{R}^{\overline{m}}\times\mathbb{R}^{d})$$
$$D(\xi):=\{z\in\mathbb{R}^{r} : W(\xi)^{\top}z-q(\xi)\in Y^{*}\}\;(\xi\in\Xi),$$

where $W(\xi)^{\top}$ is the transposed of $W(\xi)$ and Y^{*} the polar cone of Y. Then we have

$$\hat{\Phi}(\xi,x)=\Phi(\xi,q(\xi),h(\xi)-T(\xi)x)=\sup\{\langle h(\xi)-T(\xi)x,z\rangle : z\in D(\xi)\}.$$

Theorem: (Walkup/Wets 69)
For any $\xi\in\Xi$, the function $\Phi(\xi,\cdot,\cdot)$ is finite and continuous on the polyhedral set $D(\xi)\times W(\xi)Y$. Furthermore, the function $\Phi(\xi,u,\cdot)$ is piecewise linear convex on the polyhedral set $W(\xi)Y$ for fixed $u\in D(\xi)$, and $\Phi(\xi,\cdot,t)$ is piecewise linear concave on $D(\xi)$ for fixed $t\in W(\xi)Y$.

Assumptions:

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

(A2) *dual feasibility*: $D(\xi) \neq \emptyset$ holds for all $\xi \in \Xi$.

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

Proposition:

Then the deterministic equivalent of the two-stage model represents a finite convex program (with polyhedral constraints) if the integrals $\int_\Xi \Phi(\xi, q(\xi), h(\xi) - T(\xi)x)P(d\xi)$ are finite for all $x \in X$.
For fixed recourse ($W(\xi) \equiv W$), it suffices to assume

$$\int_\Xi \|\xi\|^2 P(d\xi) < \infty.$$

Convex subdifferentials, optimality conditions, conditions for differentiability, duality results are well known.

(Ruszczyński/Shapiro, Handbook, 2003)

Towards stability

We define the integrand $f_0 : \Xi \times I\!R^m \to \overline{I\!R}$ by

$$f_0(\xi, x) = \begin{cases} \langle c, x\rangle + \Phi(\xi, q(\xi), h(\xi) - T(\xi)x) & \text{if } h(\xi) - T(\xi)x \in W(\xi)Y,\ D(\xi) \neq \emptyset, \\ +\infty & \text{otherwise,} \end{cases}$$

and note that f_0 is a convex random lsc function with $\Xi \times X \subseteq \text{dom } f_0$ if (A1) and (A2) are satisfied.

The two-stage stochastic program can thus be expressed as

$$\min\Big\{ \int_{\Xi} \mathbf{f_0}(\xi, \mathbf{x}) \mathbf{P}(\mathbf{d}\xi) : \ \mathbf{x} \in \mathbf{X} \Big\}.$$

We are interested in studying the behavior of its solutions when perturbing (approximating, estimating) the probability measure P. By $v(P)$, $S(P)$ and $S_\varepsilon(P)$ ($\varepsilon \geq 0$) we denote its optimal value, solution set and set of ε-approximate solutions, i.e.,

$$v(P) := \inf\{\int_\Xi f_0(\xi,x)P(d\xi) : x\in X\}$$
$$S(P) := \text{argmin}_{x\in X}\int_\Xi f_0(\xi,x)P(d\xi) := S_0(P),$$
$$S_\varepsilon(P) := \{x\in X : \int_\Xi f_0(\xi,x)P(d\xi)\le v(P)+\varepsilon\}.$$

We consider classes of relevant functions and probability measures, namely, $\mathcal{F} = \{f_0(\cdot,x) : x\in X\}$ and

$$\mathcal{P}_\mathcal{F} = \{Q\in\mathcal{P}(\Xi) : \int_\Xi \inf_{x\in X\cap\rho\mathbb{B}} f_0(\xi,x)Q(d\xi) > -\infty \text{ , and } \sup_{x\in X\cap\rho\mathbb{B}}\int_\Xi f_0(\xi,x)Q(d\xi)<\infty \text{ , for all } \rho>0\},$$

where $\mathbb{B}$ is the closed unit ball in $\mathbb{R}^m$. We note that the infimum function $\xi\mapsto\inf_{x\in X\cap\rho\mathbb{B}} f_0(\xi,x)$ is measurable for each $\rho>0$ as f_0 is a random lsc function.

For any $\rho > 0$ and probability measures P, $Q \in \mathcal{P}_{\mathcal{F}}$ we consider the following distance

$$d_{\mathcal{F},\rho}(P,Q) = \sup_{x \in X \cap \rho \mathbb{B}} \left| \int_{\Xi} f_0(\xi,x) P(d\xi) - \int_{\Xi} f_0(\xi,x) Q(d\xi) \right| .$$

It is nonnegative, finite, symmetric and satisfies the triangle inequality, i.e., it is a semi-metric on $\mathcal{P}_{\mathcal{F}}$. In general, however, the class $\mathcal{F}_\rho$ will not be rich enough to guarantee $d_{\mathcal{F},\rho}(P,Q) = 0$ implies $P = Q$.

Lemma:

For any $Q \in \mathcal{P}_{\mathcal{F}}$, the function $x \mapsto \int_{\Xi} f_0(\xi,x) Q(d\xi)$ is convex and lsc on $\mathbb{R}^m$.

Proof by using Fatou's lemma.

Set- and epi-distances (Rockafellar/Wets 98)

Let $d_C(x) = d(x, C) = \inf_{y\in C} \|x - y\|$ denote the distance of $x \in I\!R^m$ to a non-empty closed subset of $I\!R^m$. The ρ−distance between two non-empty closed sets is by definition

$$d\!l_\rho(C, D) = \sup_{||x||\le\rho} |d_C(x) - d_D(x)|.$$

In fact, it is just a semi-distance from which one can build a metric on the hyperspace of closed sets (metrizing the topology of Painlevé-Kuratowski convergence), for example, by setting

$$d\!l(C, D) = \int_0^\infty d\!l_\rho(C, D)e^{-\rho}\, d\rho.$$

Estimates for the ρ-distance can be obtained by relying on a 'truncated' Pompeiu-Hausdorff type distance:

$$\hat{d\!l}_\rho(C, D) = \inf\{\eta \ge 0 : C \cap \rho I\!B \subset D + \eta I\!B; D \cap \rho I\!B \subset C + \eta I\!B\}.$$

Indeed one always has,

$$\hat{d\!l}_\rho(C_1, C_2) \le d\!l_\rho(C_1, C_2) \le \hat{d\!l}_{\rho'}(C_1, C_2)$$

for $\rho' \ge 2\rho + \max\{d_{C_1}(0), d_{C_2}(0)\}$.

If we let $\rho \to \infty$, we end up with $dl_\rho(C,D)$ and $\hat{dl}_\rho(C,D)$ tending to $dl_\infty(C,D)$, the Pompeiu-Hausdorff distance between the closed non-empty sets C and D.

The distance between (lsc) functions is measured in terms of the distance between their epigraphs, so for $\rho > 0$,

$$dl_\rho(f,g) = dl_\rho(\operatorname{epi} f, \operatorname{epi} g), \qquad \hat{dl}_\rho(f,g) = \hat{dl}_\rho(\operatorname{epi} f, \operatorname{epi} g).$$

and $dl(f,g) = dl(\operatorname{epi} f, \operatorname{epi} g)$. However, since our sets are epigraphs (in $\mathbb{R}^{m+1}$), it is convenient to rely on the 'unit ball' to be $\mathbb{B} \times [-1,1]$, this brings us to an 'auxiliary' distance $\hat{dl}^+_\rho(f_1,f_2)$ defined as the infimum of all $\eta \geq 0$ such that for all $x \in \rho\mathbb{B}$,

$$\begin{aligned} \min\nolimits_{B(x,\eta)} f_2 &\leq \max\{ f_1(x), -\rho \} + \eta \\ \min\nolimits_{B(x,\eta)} f_1 &\leq \max\{ f_2(x), -\rho \} + \eta. \end{aligned}$$

For lsc $f_1,\ f_2 : \mathbb{R}^n \to \overline{\mathbb{R}}$, not identically ∞, one has,

$$\hat{dl}^+_{\rho/\sqrt{2}}(f_1,f_2) \leq \hat{dl}_\rho(f_1,f_2) \leq \sqrt{2}\,\hat{dl}^+_\rho(f_1,f_2).$$

Quantitative stability of two-stage models

Theorem: (Römisch/Wets 06)
Let $P \in \mathcal{P}_\mathcal{F}$ and suppose $S(P)$ is non-empty and bounded. Then there exist constants $\rho > 0$ and $\delta > 0$ such that

$$\begin{aligned} |v(P) - v(Q)| &\leq d_{\mathcal{F},\rho}(P,Q) \\ \emptyset \neq S(Q) &\subset S(P) + \Psi_P(d_{\mathcal{F},\rho}(P,Q))\mathbb{B} \end{aligned}$$

holds for all $Q \in \mathcal{P}_\mathcal{F}$ with $d_{\mathcal{F},\rho}(P,Q) < \delta$, where Ψ_P is a *conditioning function* associated with our given program, more precisely,

$$\Psi_P(\eta) := \eta + \psi_P^{-1}(2\eta),\ \eta \geq 0, \quad \text{with}$$

$$\psi_P(\tau) := \min\{\int_\Xi f_0(\xi,x)P(d\xi) - v(P) : d(x,S(P)) \geq \tau\},\ \tau \geq 0.$$

Simple examples of two-stage stochastic programs show that, in general, the set-valued mapping $S(.)$ is not inner semicontinuous at P (Römisch 03). Furthermore, explicit descriptions of conditioning functions ψ_P of stochastic programs (like linear or quadratic growth at solution sets) are only known in some specific cases, for example, for linear two-stage stochastic programs with finite discrete distribution or with strictly positive densities of random right-hand sides (Schultz 94).

We are in much better shape, when we consider the stability properties of the sets $S_\varepsilon(\cdot)$ of ε-approximate solutions.

Theorem:
Let $P \in \mathcal{P}_\mathcal{F}$ and $S(P)$ be non-empty, bounded. Then there exist constants $\hat{\rho} > 0$ and $\hat{\varepsilon} > 0$ such that

$$\mathbb{d}l_\infty(S_\varepsilon(P), S_\varepsilon(Q)) \leq \frac{4\hat{\rho}}{\varepsilon} d_{\mathcal{F},\hat{\rho}+\varepsilon}(P, Q)$$

holds for any $\varepsilon \in (0, \hat{\varepsilon})$ and $Q \in \mathcal{P}_\mathcal{F}$ such that $d_{\mathcal{F},\hat{\rho}+\varepsilon}(P, Q) < \varepsilon$.

The preceding stability results remain valid if the set $\mathcal{F}_\rho$ is enlarged to a set $\hat{\mathcal{F}}$ and the set $\mathcal{P}_\mathcal{F}$ reduced to a subset on which the new distance

$$d_{\hat{\mathcal{F}}}(P, Q) = \sup_{f \in \hat{\mathcal{F}}} \left| \int_\Xi f(\xi) P(d\xi) - \int_\Xi f(\xi) P(d\xi) \right|$$

is finite and well-defined. Which classes $\hat{\mathcal{F}}$ of functions contain $\mathcal{F}_\rho = \{f_0(\cdot, x) : x \in X \cap \rho \mathbb{B}\}$ for any $\rho > 0$?

In the context of two-stage models, function classes of the form

$$\mathcal{F}_H := \{ f : \Xi \to I\!R \;:\; f(\xi) - f(\tilde{\xi}) \le \max\{1, H(\|\xi\|), H(\|\tilde{\xi}\|)\} \cdot \|\xi - \tilde{\xi}\|, \forall \xi, \tilde{\xi} \in \Xi \}$$

are of particular interest, where $H : I\!R_+ \to I\!R_+$ is nondecreasing, $H(0) = 0$. The corresponding distances are

$$d_{\mathcal{F}_H}(P,Q) = \sup_{f \in \mathcal{F}_H} \left| \int_\Xi f(\xi) P(d\xi) - \int_\Xi f(\xi) Q(d\xi) \right| =: \zeta_H(P,Q)$$

are so-called Fortet-Mourier metrics defined on

$$\mathcal{P}_H(\Xi) := \{ Q \in \mathcal{P}(\Xi) : \int_\Xi \max\{1, H(\|\xi\|)\} \|\xi\| Q(d\xi) < \infty \}$$

Important special case: $H(t) := t^{r-1}$ for $r \ge 1$.

The corresponding classes of functions and measures, and the distances are denoted by $\mathcal{F}_r$, $\mathcal{P}_r(\Xi)$ and ζ_r, respectively, where the measures are in the class

$$\mathcal{P}_r(\Xi) := \{ Q \in \mathcal{P}(\Xi) : \int_\Xi \|\xi\|^r Q(d\xi) < \infty \}.$$

Convergence with respect to ζ_r means weak conergence of the probability measures and $| \int_\Xi \|\xi\|^r P(d\xi) - \int_\Xi \|\xi\|^r Q(d\xi) | \le r \zeta_r(P,Q)$, i.e., convergence of the r-th order moments (Rachev 91).

Home Page
Title Page
Contents
Page 36 of 66
Go Back
Full Screen
Close
Quit

Under which conditions appear relevant classes $\mathcal{F}_H$ containing $\mathcal{F}_\rho$?

Proposition:

Suppose the stochastic program satisfies (A1) and (A2). Assume that the mapping $\xi \mapsto D(\xi)$ is bounded-valued and there exists a constant $L > 0$, and a nondecreasing function $h : \mathbb{R}_+ \to \mathbb{R}_+$ with $h(0) = 0$ such that

$$d\!l_\infty(D(\xi), D(\tilde{\xi})) \le L \max\{1, h(\|\xi\|), h(\|\tilde{\xi}\|)\}\|\xi - \tilde{\xi}\|$$

holds for all $\xi, \tilde{\xi} \in \Xi$.

Then, for any $\rho > 0$, there exist $\hat{L} > 0$ and $\hat{L}(\rho) > 0$ such that

$$\begin{aligned} f_0(\xi, x) - f_0(\tilde{\xi}, x) &\le \hat{L}(\rho) \max\{1, H(\|\xi\|), H(\|\tilde{\xi}\|)\}\|\xi - \tilde{\xi}\| \\ f_0(\xi, x) - f_0(\xi, \tilde{x}) &\le \hat{L} \max\{1, H(\|\xi\|)\|\xi\|\}\|x - \tilde{x}\| \end{aligned}$$

for all $\xi, \tilde{\xi} \in \Xi$, $x, \tilde{x} \in X \cap \rho\mathbb{B}$, where H is defined by

$$H(t) := h(t)t, \; \forall t \in \mathbb{R}_+.$$

Note that $h(t) = \begin{cases} 1 & \text{, fixed recourse} \\ t^k & \text{, lower diagonal randomness with } k \text{ blocks.} \end{cases}$

Example:

Let $\overline{m} = 4$, $d = 2$, $Y = I\!R^4_+$, $\Xi = I\!R$ and consider the random (second-stage) costs and recourse matrix

$$W(\xi) = \begin{pmatrix} 1 & -1 & 0 & 0 \\ -\xi & 0 & 1 & -1 \end{pmatrix} \qquad q(\xi) = \begin{pmatrix} 0 \\ 0 \\ \xi \\ -\xi \end{pmatrix}$$

Then $W(\xi)Y = I\!R^2$ (complete recourse) and $D(\xi)=[0, \xi^2] \times \{\xi\}$. Hence, the conditions (A1), (A2) are satisfied and the local Lipschitz continuity property of $D(\cdot)$ holds with $h(t) = t$, $t \in I\!R_+$.

Remark: (convergence of empirical estimates)

For the empirical measure $P_n = n^{-1} \sum_{i=1}^n \delta_{\xi_i}$, where ξ_i, $i \in I\!N$ are i.i.d. samples from P, exponential estimates for the convergence in probability of $d_{\mathcal{F},\rho}(P_n, P) = \sup_{f \in \mathcal{F}_\rho} |\int_\Xi f(\xi)(P_n - P)(d\xi)|$ can be obtained by showing that the covering number of $\mathcal{F}_\rho$, i.e., the minimal number of balls with radius ε in $L_2(\Xi, P)$, grows at most with ε^{-r} for some $r \geq 1$.

Discrete approximations of two-stage stochastic programs

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

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

and insert it into the infinite-dimensional stochastic program:

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

$$\begin{array}{lllllll} W(\xi_1)y_1 & & & & +T(\xi_1)x & = & h(\xi_1) \\ & W(\xi_2)y_2 & & & +T(\xi_2)x & = & h(\xi_2) \\ & & \ddots & & \vdots & = & \vdots \\ & & & W(\xi_n)y_n & +T(\xi_n)x & = & h(\xi_n)\} \end{array}$$

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

(Ruszczyński/Shapiro, Handbook, 2003)

How to choose the discrete approximation ?

The quantitative stability results suggest to determine P_n such that it forms the best approximation of P with respect to the semi-distance $d_{\mathcal{F},\rho}$ or the probability metric ζ_r, i.e., given $n \in I\!N$ solve

$$\min\{\zeta_r(P, \frac{1}{n}\sum_{i=1}^{n}\delta_{\xi_i}) : \xi_i \in \Xi,\ i = 1, \ldots, n\}$$

Such best approximations $P_n^* = \frac{1}{n}\sum_{i=1}^{n}\delta_{\xi_i^*}$ are known as optimal quantizations of the probability distribution P (Graf/Luschgy, LNM 2000).

Convergence properties of optimal quantizations and numerical methods for solving the best approximation problems in case of the ℓ_r-minimal metrics (or Wasserstein metrics)

$$\ell_r(P,Q) := \left(\inf\left\{\int_{\Xi\times\Xi}\|\xi-\tilde{\xi}\|^r\eta(d\xi,d\tilde{\xi})\mid \pi_1\eta = P,\ \pi_2\eta = Q\right\}\right)^{\frac{1}{r}},$$

are already known. Here, π_i is the projection onto the i-th component. The convergence rates are in some cases better than $O(n^{-\frac{1}{2}})$ for sampling methods. Note that $\zeta_r(P,Q) \le (1 + \int_{\Xi}\|\xi\|^r(P+Q)(d\xi))^{\frac{r-1}{r}}\ell_r(P,Q)$.

Scenario reduction

We consider discrete distributions P with scenarios ξ_i and probabilities p_i, $i = 1, \dots, N$, and Q being supported by a given subset of scenarios ξ_j, $j \notin J \subset \{1, \dots, N\}$, of P.

Optimal reduction of a given scenario set J:
The best approximation of P with respect to $\zeta_r = \mu_{\hat{c}_r}$ by such a distribution Q exists and is denoted by Q^*. It has the distance

$$D_J := \zeta_r(P, Q^*) = \min_Q \mu_{\hat{c}_r}(P, Q) = \sum_{i \in J} p_i \min_{j \notin J} \hat{c}_r(\xi_i, \xi_j)$$
$$= \sum_{i \in J} p_i \min\{\sum_{k=1}^{n-1} c_r(\xi_{l_k}, \xi_{l_{k+1}}) : n \in \mathbb{N}, l_k \in \{1, \dots, N\}, l_1 = i, l_n = j \notin J\}$$

and the probabilities $q_j^* = p_j + \sum_{i \in J_j} p_i$, $\forall j \notin J$, where

$J_j := \{i \in J : j = j(i)\}$ and $j(i) \in \arg\min_{j \notin J} \hat{c}_r(\xi_i, \xi_j)$, $\forall i \in J$.

(Dupačová/Gröwe-Kuska/Römisch 03)

We needed the following notation:

$$c_r(\xi, \tilde{\xi}) := \max\{1, \|\xi\|^{r-1}, \|\tilde{\xi}\|^{r-1}\}\|\xi - \tilde{\xi}\| \quad (\xi, \tilde{\xi} \in \Xi).$$

Proposition: (Rachev/Rüschendorf 98)

$$\zeta_r(P,Q) = \mu_{\hat{c}_r}(P,Q) = \inf\left\{\int_{\Xi\times\Xi} \hat{c}_r(\xi,\tilde{\xi})\eta(d\xi,d\tilde{\xi}) : \pi_1\eta = P, \pi_2\eta = Q\right\}$$

where $\hat{c}_r \leq c_r$ and $\hat{c}_r$ is the metric (reduced cost)

$$\hat{c}_r(\xi,\tilde{\xi}) := \inf\left\{\sum_{i=1}^{n-1} c_r(\xi_{l_i},\xi_{l_{i+1}}) : n \in \mathbb{N}, \xi_{l_i} \in \Xi, \xi_{l_1} = \xi, \xi_{l_n} = \tilde{\xi}\right\}.$$

Determining the optimal scenario index set with prescribed cardinality n is, however, a combinatorial optimization problem of set covering type:

$$\min\{D_J = \sum_{i\in J} p_i \min_{j\notin J} \hat{c}_r(\xi_i,\xi_j) : J \subset \{1,\ldots,N\}, \#J = N - n\}$$

Hence, the problem of finding the optimal set J to delete is $\mathcal{NP}$-hard and polynomial time solution algorithms do not exist.

Fast reduction heuristics

Starting point ($n = N - 1$): $\min_{l \in \{1,\ldots,N\}} p_l \min_{j \neq l} \hat{c}_r(\xi_l, \xi_j)$

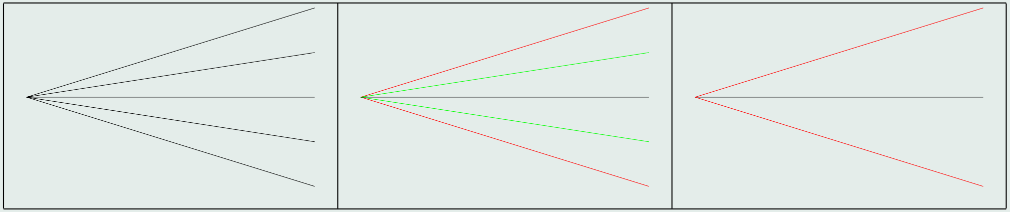
Algorithm 1: (Backward reduction)

Step [0]: $J^{[0]} := \emptyset$.

Step [i]: $l_i \in \arg\min_{l \notin J^{[i-1]}} \sum_{k \in J^{[i-1]} \cup \{l\}} p_k \min_{j \notin J^{[i-1]} \cup \{l\}} \hat{c}_r(\xi_k, \xi_j)$.

$J^{[i]} := J^{[i-1]} \cup \{l_i\}$.

Step [N-n+1]: Optimal redistribution.



Starting point ($n = 1$): $\min\limits_{u \in \{1,\ldots,N\}} \sum\limits_{k=1}^{N} p_k \hat{c}_r(\xi_k, \xi_u)$

Algorithm 2: (Forward selection)

Step [0]: $J^{[0]} := \{1, \ldots, N\}$.

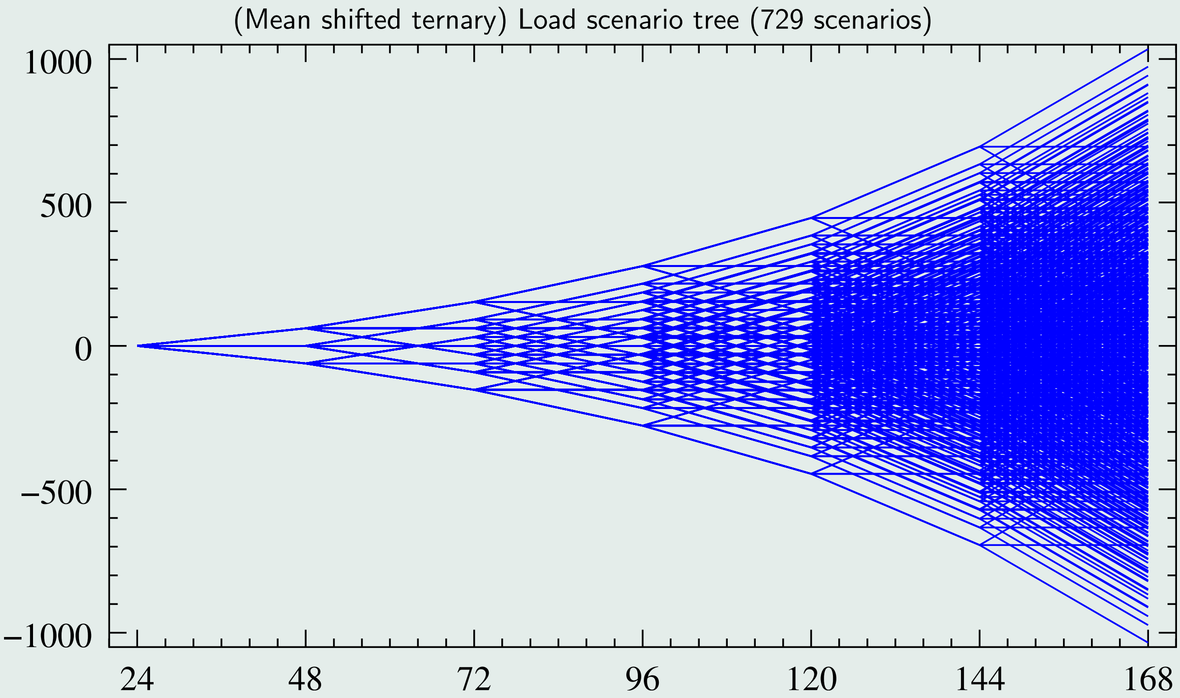
Step [i]: $u_i \in \arg \min\limits_{u \in J^{[i-1]}} \sum\limits_{k \in J^{[i-1]} \setminus \{u\}} p_k \min\limits_{j \notin J^{[i-1]} \setminus \{u\}} \hat{c}_r(\xi_k, \xi_j)$,

$$J^{[i]} := J^{[i-1]} \setminus \{u_i\}.$$

Step [n+1]: Optimal redistribution.

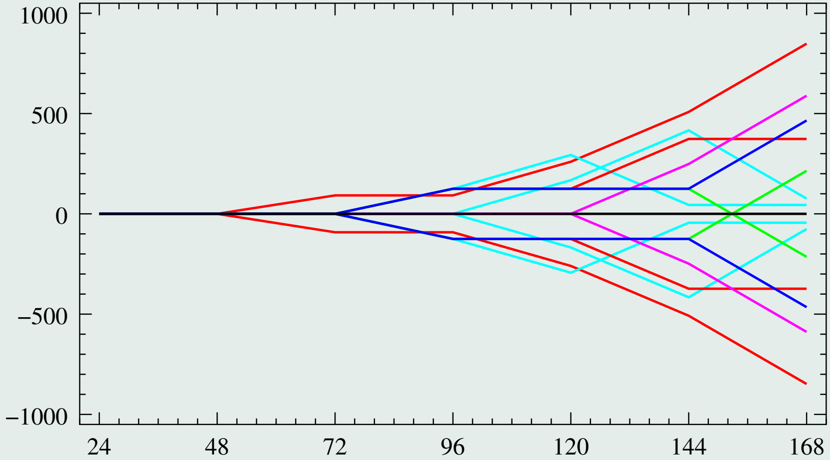
Home Page
Title Page
Contents
Page 44 of 66
Go Back
Full Screen
Close
Quit

Example: (Electrical load scenario tree)

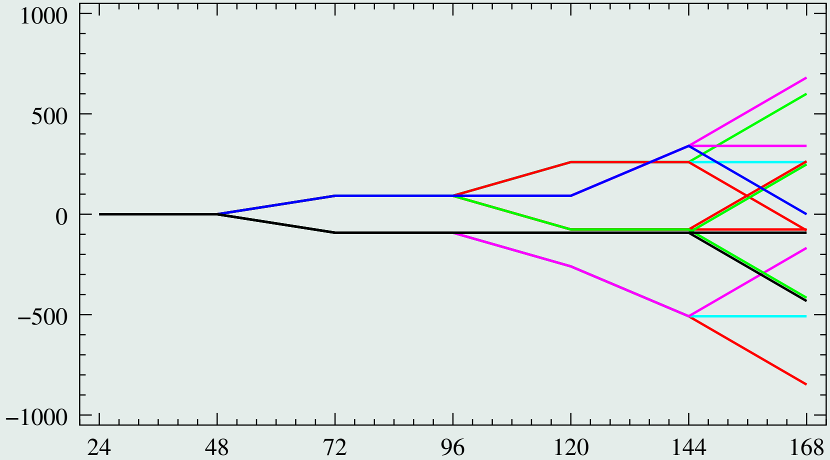


<Start Animation>

Reduced load scenario tree obtained by the forward selection method (15 scenarios)



Reduced load scenario tree obtained by the backward reduction method (12 scenarios)



Home Page

Title Page

Contents

Go Back

Full Screen

Close

Quit

Part III

Optimization in L_r-spaces – Multistage stochastic programs

(H. Heitsch (Berlin), C. Strugarek (EdF, Clamart))

Multistage stochastic programs

Let $\{\xi_t\}_{t=1}^T$ be a discrete-time stochastic data process defined on some probability space $(\Omega, \mathcal{F}, I\!P)$ 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:

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

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.

The model is (multiperiod) two-stage if $\mathcal{F}_t = \mathcal{F}$, $t = 2, \ldots, T$.

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
$x_t \in L_{r'}(\Omega, \mathcal{F}_t, I\!P; I\!R^{m_t})$ and $\xi_t \in L_r(\Omega, \mathcal{F}_t, I\!P; I\!R^d)$,
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

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

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

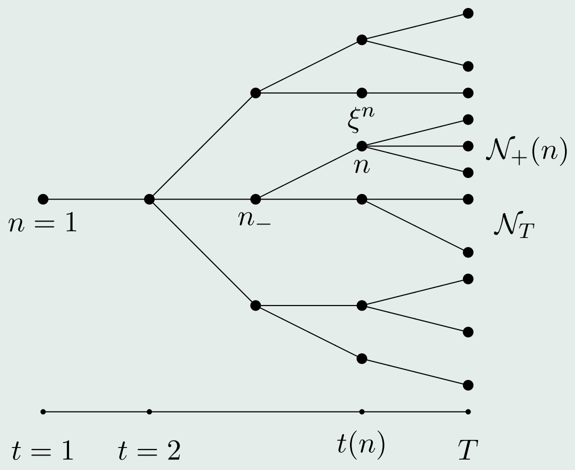
i.e., as a subspace constraint, by using the conditional expectation $I\!E[\cdot|\mathcal{F}_t]$ with respect to the σ-algebra $\mathcal{F}_t$.

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

$\longrightarrow$ infinite-dimensional optimization problem

1. Data process approximation by scenario trees

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



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

$n = 1$ root node, n_- unique predecessor of node n, $\text{path}(n) = \{1, \dots, n_-, n\}$, $t(n) := |\text{path}(n)|$, $\mathcal{N}_+(n)$ 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 ?

Idea:

Utilizing quantitative stability results !?

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

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

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

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 nonlinear way !

Quantitative Stability

Let us introduce some notations. Let F denote the objective function defined on $L_r(\Omega, \mathcal{F}, I\!P; I\!R^s) \times L_{r'}(\Omega, \mathcal{F}, I\!P; I\!R^m) \to I\!R$ by $F(\xi, x) := I\!E[\sum_{t=1}^{T} \langle b_t(\xi_t), x_t \rangle]$, let

$$\mathcal{X}_t(x_{t-1}; \xi_t) := \{x_t \in X_t : A_{t,0} x_t + A_{t,1}(\xi_t) x_{t-1} = h_t(\xi_t)\}$$

denote the t-th feasibility set for every $t = 2, \ldots, T$ and

$$\mathcal{X}(\xi) := \{x \in L_{r'}(\Omega, \mathcal{F}, I\!P; I\!R^m) : x_1 \in \mathcal{X}_1(\xi_1), x_t \in \mathcal{X}_t(x_{t-1}; \xi_t)\}$$

the set of feasible elements with input ξ.

Then the multistage stochastic program may be rewritten as

$$\min\{F(\xi, x) \,:\, x \in \mathcal{X}(\xi) \cap \mathcal{N}_{r'}(\xi)\}.$$

Let $v(\xi)$ denote its optimal value and, for any $\alpha \geq 0$,

$$\begin{aligned} l_\alpha(F(\xi, \cdot)) &:= \{x \in \mathcal{X}(\xi) \cap \mathcal{N}_{r'}(\xi) : F(\xi, x) \leq v(\xi) + \alpha\} \\ S(\xi) &:= l_0(F(\xi, \cdot)) \end{aligned}$$

denote the α-level set and the solution set of the stochastic program with input ξ.

Assumptions:

(A1) $\xi \in L_r(\Omega, \mathcal{F}, I\!\!P; I\!\!R^s)$ for some $r \geq 1$.

(A2) There exists a $\delta > 0$ such that for any $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, I\!\!P; I\!\!R^s)$ with $\|\tilde{\xi} - \xi\|_r \leq \delta$, any $t = 2, \ldots, T$ and any $x_1 \in X_1$, $x_\tau \in \mathcal{X}_\tau(x_{\tau-1}; \tilde{\xi}_\tau)$, $\tau = 2, \ldots, t-1$, the set $\mathcal{X}_t(x_{t-1}; \tilde{\xi}_t)$ is nonempty (relatively complete recourse locally around ξ).

(A3) For each $\xi \in \Xi$ there exists $z \in \times_{t=1}^{T} L_{r'}(\Omega, \mathcal{F}, I\!\!P; I\!\!R^{n_t})$ with

$$\begin{aligned} A_{t,0}^\top z_t + A_{t+1,1}^\top(\xi_{t+1}) z_{t+1} - h_t(\xi_t) &\in X_t^*, \; t = 1, \ldots, T-1, \\ A_{T,0}^\top z_T - h_T(\xi_T) &\in X_T^*, \end{aligned}$$

where X_t^* denotes the polar to the polyhedral cone X_t, $t = 1, \ldots, T$ (dual feasibility).

(A4) The objective function F is level-bounded locally uniformly at ξ, i.e., for some $\alpha > 0$ there exists a $\delta > 0$ and a bounded subset B of $L_{r'}(\Omega, \mathcal{F}, I\!\!P; I\!\!R^m)$ such that $l_\alpha(F(\tilde{\xi}, \cdot))$ is nonempty and contained in B for all $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, I\!\!P; I\!\!R^s)$ with $\|\tilde{\xi} - \xi\|_r \leq \delta$.

Theorem: (Heitsch/Römisch/Strugarek 06)
Let (A1) – (A4) be satisfied and $\mathcal{X}_1(\xi_1)$ be (uniformly) bounded. Then there exist positive constants L, α and δ such that

$$|v(\xi) - v(\tilde{\xi})| \leq L(\|\xi - \tilde{\xi}\|_r + D_{\rm f}(\xi, \tilde{\xi}))$$

holds for all $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ with $\|\tilde{\xi} - \xi\|_r \leq \delta$.

Assume that only costs and right-hand sides are random and that the solution x^* of the original problem is unique.
If $(\xi^{(n)})$ is a sequence in $L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ such that

$$\|\xi^{(n)} - \xi\|_r \quad \text{and} \quad D_{\rm f}(\xi^n, \xi)$$

converge to 0, then any sequence $(x^{(n)})$ of solutions of the approximate problems converges to x^* with respect to the weak (weak*) topology $\sigma(L_{r'}, L_r)$.

Here, $D_{\rm f}(\xi, \tilde{\xi})$ denotes the filtration distance of ξ and $\tilde{\xi}$ defined by

$$D_{\rm f}(\xi, \tilde{\xi}) = \inf_{\substack{x \in S(\xi) \\ \tilde{x} \in S(\tilde{\xi})}} \sum_{t=2}^{T-1} \max\{\|x_t - \mathbb{E}[x_t|\tilde{\mathcal{F}}_t]\|_{r'}, \|\tilde{x}_t - \mathbb{E}[\tilde{x}_t|\mathcal{F}_t]\|_{r'}\}.$$

Remark:

The convergence of approximate solutions can be supplemented by a quantitative stability property of the set $S_1(\xi)$ of first stage solutions. Namely, there exists a constant $\hat{L} > 0$ such that

$$\sup_{x \in S_1(\tilde{\xi})} d(x, S_1(\xi)) \leq \Psi_\xi^{-1}(\hat{L}(\|\xi - \tilde{\xi}\|_r + D_{\mathrm{f}}(\xi, \tilde{\xi}))),$$

where $\Psi_\xi(\tau) := \inf\{\mathbb{E}[f(x_1, \xi)] - v(\xi) : d(x_1, S_1(\xi)) \geq \tau, x_1 \in X_1\}$ with $\Psi_\xi^{-1}(\alpha) := \sup\{\tau \in \mathbb{R}_+ : \Psi_\xi(\tau) \leq \alpha\}$ $(\alpha \in \mathbb{R}_+)$ is the growth function of the original problem near its solution set $S_1(\xi)$.

Remark:

The filtration distance $D_{\mathrm{f}}(\xi, \tilde{\xi})$ may be further estimated by the distance $d_{\mathrm{f}}(\xi, \tilde{\xi})$ with

$$d_{\mathrm{f}}(\xi, \tilde{\xi}) := \sup_{\|x\|_{r'} \leq 1,\, x \in L_{r'}} \sum_{t=2}^{T-1} \|\mathbb{E}[x_t|\mathcal{F}_t] - \mathbb{E}[x_t|\tilde{\mathcal{F}}_t]\|_{r'}.$$

In case of finite Ω, this distance corresponds to the $l_{r'}$-distance of two matrices representing the information on the filtrations of ξ and $\tilde{\xi}$, respectively.

Next we compute the distance $d_{\mathrm{f}}(\xi,\tilde{\xi})$ of filtrations for the special case that Ω is finite, say, $\Omega = \{\omega_1,\dots,\omega_S\}$. Let $I\!P(\{\omega_i\}) = p_i$, $i = 1,\dots,S$ and let $\mathcal{E}_t$ and $\tilde{\mathcal{E}}_t$ be partitions of Ω that generate the σ-fields $\mathcal{F}_t$ and $\tilde{\mathcal{F}}_t$, respectively. Then $I\!E[x_t|\mathcal{F}_t] = H_t x_t$, where the matrix H_t is of the form

$$H_t = (e_{\sigma s})_{\sigma,s=1,\dots,S}\,, \quad \text{where} \quad e_{\sigma s} := \begin{cases} \dfrac{p_s}{\sum\limits_{i\in E_{t\sigma}} p_i} & , \ s \in E_{t\sigma} \\ 0 & , \ s \notin E_{t\sigma} \end{cases}$$

and $\omega_\sigma \in E_{t\sigma} \in \mathcal{E}_t$. Analogously, $\tilde{H}_t = (\tilde{e}_{\sigma s})_{\sigma,s=1,\dots,S}$ is defined using the corresponding sets $\tilde{E}_{t\sigma}$ in a generator of the σ-field $\tilde{\mathcal{F}}_t$. Hence, we obtain for $r' = \infty$, i.e., the row sum norm $\|\cdot\|_\infty$ of matrices, that

$$d_{\mathrm{f}}(\xi,\tilde{\xi}) = \sum_{t=1}^{T-1} \|H_t - \tilde{H}_t\|_\infty$$

$$\|H_t-\tilde{H}_t\|_\infty = \max_{\sigma=1,\dots,S}\left\{ \sum_{s\in E_{t\sigma}\setminus\tilde{E}_{t\sigma}} \frac{p_s}{\sum\limits_{i\in E_{t\sigma}} p_i} + \sum_{s\in \tilde{E}_{t\sigma}\setminus E_{t\sigma}} \frac{p_s}{\sum\limits_{i\in \tilde{E}_{t\sigma}} p_i} + \sum_{s\in E_{t\sigma}\cap\tilde{E}_{t\sigma}} \left| \frac{p_s}{\sum\limits_{i\in E_{t\sigma}} p_i} - \frac{p_s}{\sum\limits_{i\in \tilde{E}_{t\sigma}} p_i} \right| \right\}$$

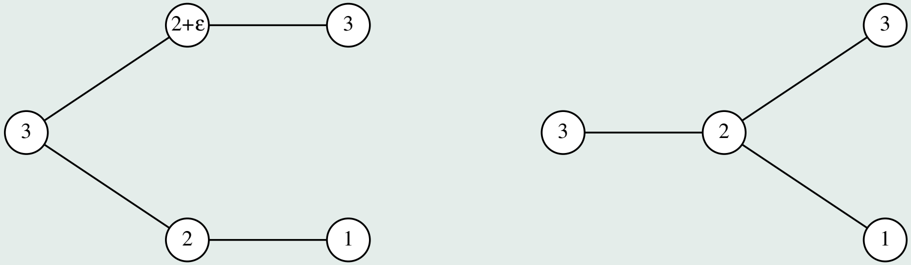
for $t = 2,\dots,T-1$.

The following example shows that the filtration distance D_{f} is indispensable for the stability result to hold.

Example: (Optimal purchase under uncertainty)
The decisions x_t correspond to the amounts to be purchased at each time period with uncertain prices are ξ_t, $t = 1, \ldots, T$, and such that a prescribed amount a is achieved at the end of a given time horizon. The problem is of the form

$$\min \left\{ \mathbb{E}\left[\sum_{t=1}^{T} \xi_t x_t\right] \middle| \begin{array}{l} (x_t, s_t) \in X_t = \mathbb{R}_+^2, \\ (x_t, s_t) \text{ is } (\xi_1, \ldots, \xi_t)\text{-measurable}, \\ s_t - s_{t-1} = x_t,\ t = 2, \ldots, T, \\ s_1 = 0,\ s_T = a. \end{array} \right\},$$

where the state variable s_t corresponds to the amount at time t.
Let $T := 3$ and ξ_ε denote the stochastic price process having the two scenarios $\xi_\varepsilon^1 = (3, 2+\varepsilon, 3)$ $(\varepsilon \in (0,1))$ and $\xi_\varepsilon^2 = (3, 2, 1)$ each endowed with probability $\frac{1}{2}$. Let $\tilde{\xi}$ denote the approximation of ξ_ε given by the two scenarios $\tilde{\xi}^1 = (3, 2, 3)$ and $\tilde{\xi}^2 = (3, 2, 1)$ with the same probabilities $\frac{1}{2}$.



Scenario trees for ξ_ε (left) and $\tilde{\xi}$

We obtain

$$\begin{aligned} v(\xi_\varepsilon) &= \frac{1}{2}((2+\varepsilon)a + a) = \frac{3+\varepsilon}{2}a \\ v(\tilde{\xi}) &= 2a\,, \qquad \text{but} \\ \|\xi_\varepsilon - \tilde{\xi}\|_1 &\le \frac{1}{2}(0+\varepsilon+0) + \frac{1}{2}(0+0+0) = \frac{\varepsilon}{2}. \end{aligned}$$

Hence, the multistage stochastic purchasing model is not stable with respect to $\|\cdot\|_1$.

However, the estimate for $|v(\xi) - v(\tilde{\xi})|$ in the stability theorem is valid with $L = 1$ since $D_{\mathrm{f}}(\xi, \tilde{\xi}) = \frac{a}{2}$.

Generation of scenario trees

(i) Development of a statistical model for the stochastic process ξ (parametric [e.g. time series model], nonparametric [e.g. resampling]) and generation of simulation scenarios;

(ii) Construction of a scenario tree out of the statistical model or of the simulation scenarios.

Approaches for (ii):

(1) Bound-based approximation methods (Frauendorfer 96, Kuhn 05, Edirisinghe 99, Casey/Sen 05).

(2) Monte Carlo-based schemes (inside or outside decomposition methods) (e.g. Shapiro 03, 06, Higle/Rayco/Sen 01, Chiralaksanakul/Morton 04).

(3) the use of Quasi Monte Carlo integration quadratures (Pennanen 05, 06).

(4) EVPI-based sampling schemes (inside decomposition schemes) (Corvera Poire 95, Dempster 04).

(5) Moment-matching principle (Høyland/Wallace 01, Høyland/Kaut/Wallace 03).

(6) (Nearly) best approximations based on probability metrics (Pflug 01, Hochreiter/Pflug 02, Gröwe-Kuska/Heitsch/Römisch 01, 03, Heitsch/Römisch 05).

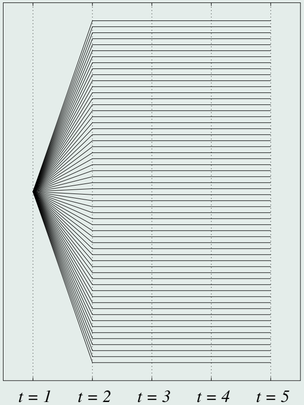
Survey: Dupačová/Consigli/Wallace 2000

Constructing scenario trees

Let ξ be the original stochastic process on some probability space $(\Omega, \mathcal{F}, I\!\!P)$ with parameter set $\{1, \ldots, T\}$ and state space $I\!\!R^d$. We aim at generating a scenario tree ξ^{tr} such that

$$\|\xi - \xi^{\mathrm{tr}}\|_r \qquad \text{and} \qquad D_{\mathrm{f}}(\xi, \xi^{\mathrm{tr}})$$

are small and, hence, the optimal values $v(\xi)$ and $v(\xi^{\mathrm{tr}})$ are close to each other. Since this problem is hardly solvable in general, we replace ξ by a finitely discrete approximation ξ^{f} such that $\|\xi - \xi^{\mathrm{f}}\|_r$ is small and its scenarios $\xi^i = (\xi^i_1, \ldots, \xi^i_T)$ with probabilities π^i, $i = 1, \ldots, N$ form a *fan* of individual scenarios.



An **algorithm** was developed that generates a tree ξ^{tr} by deleting and bundling scenarios at each $t = 2, \ldots, T$ (that are close to each other) and such that

$$\|\xi^{\mathrm{f}} - \xi^{\mathrm{tr}}\|_r$$

may be computed and bounded and that

$$D_{\mathrm{f}}(\xi^{\mathrm{f}}, \xi^{\mathrm{tr}})$$

may be bounded from above. The latter relies on the

Proposition:

Assume that only costs and right-hand sides are random and let (A2) – (A4) be satisfied. Then there exists a constant $\hat{L} > 0$ such that the filtration distance allows the estimate

$$D_{\mathrm{f}}(\xi^{\mathrm{f}}, \xi^{\mathrm{tr}}) \leq \hat{L} \begin{cases} \left(\sum_{i \in I_2} \sum_{j \in I_{2,i}} p_j \|\xi^j - \xi^i\|^{r'} \right)^{\frac{1}{r'}} & , 1 \leq r' < \infty \\ \max_{i \in I_2} \max_{j \in I_{2,i}} \|\xi^j - \xi^i\| & , r' = \infty \, . \end{cases}$$

Tolerances ε_r and ε_f are prescribed for $\|\xi^{\mathrm{f}} - \xi^{\mathrm{tr}}\|_r$ and $D_{\mathrm{f}}(\xi^{\mathrm{f}}, \xi^{\mathrm{tr}})$, respectively, which control the scenario tree generation process.

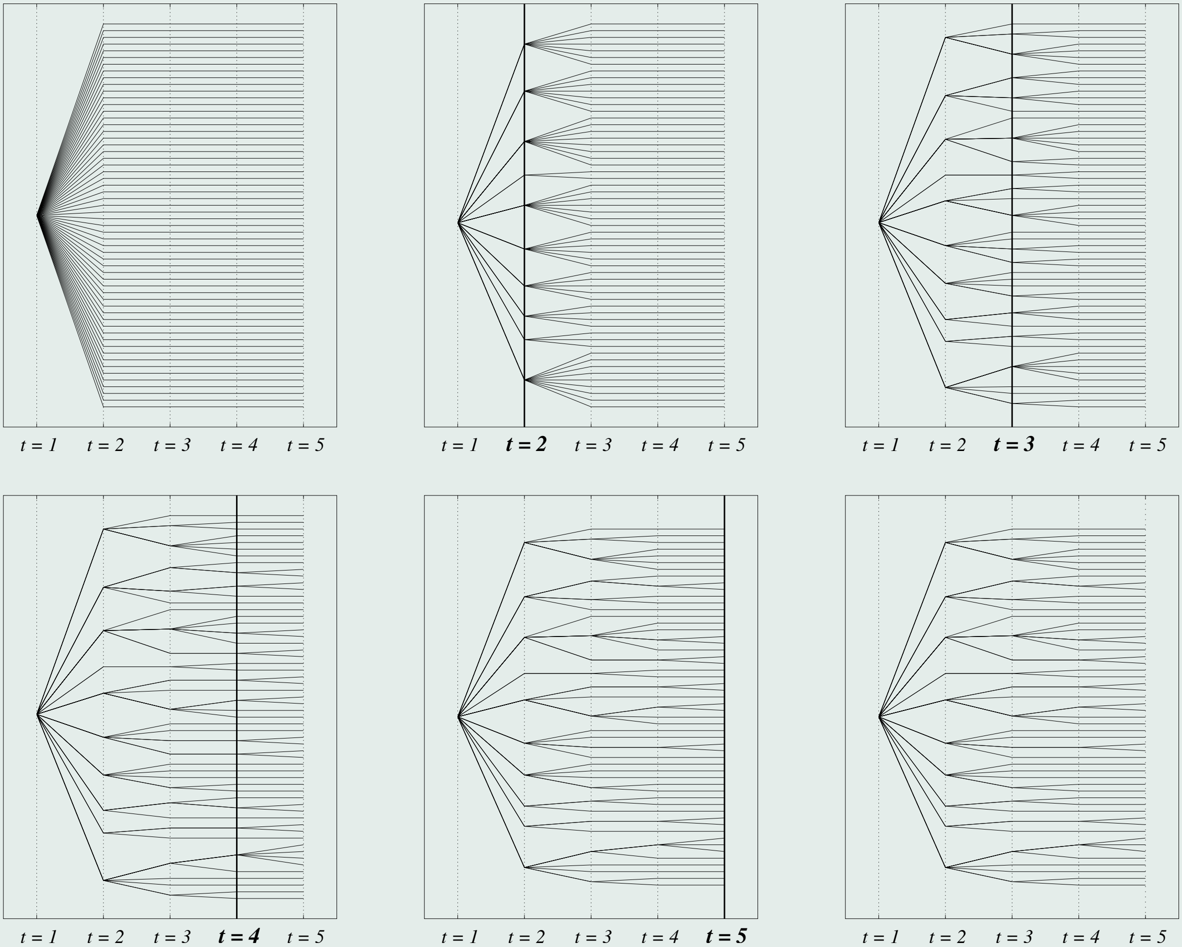


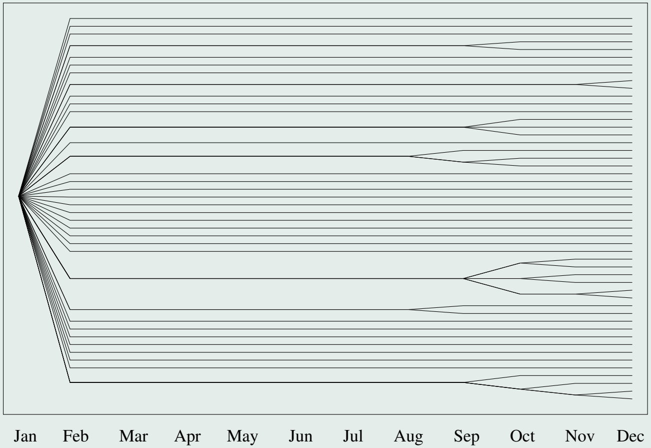
Illustration of the forward tree construction for an example including T=5 time periods starting with a scenario fan containing N=58 scenarios

<Start Animation>

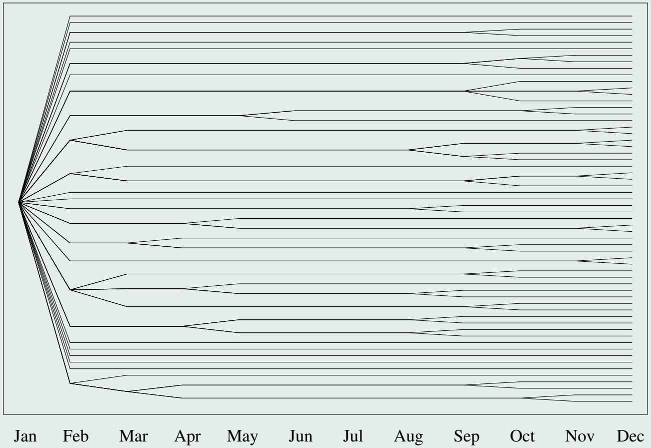
a) Forward tree construction with filtration level 0.35

b) Forward tree construction with filtration level 0.45

Yearly demand-price scenario trees with relative tolerance 0.25



a) Modified forward tree construction with filtration level 0.6



b) Modified forward tree construction with filtration level 0.7

Yearly demand-price scenario trees with relative reduction level 0.5 (Heitsch/Römisch 06)

References:

Dupačová, J., Gröwe-Kuska, N., Römisch, W.: Scenario reduction in stochastic programming: An approach using probability metrics, *Mathematical Programming* 95 (2003), 493–511.

Heitsch, H., Römisch, W., Strugarek, C.: Stability of multistage stochastic programs, Preprint 255, DFG Research Center MATHEON, Berlin 2005 and to appear in *SIAM Journal Optimization* (2006).

Heitsch, H., Römisch, W.: Scenario tree modelling for multistage stochastic programs, Preprint 296, DFG Research Center MATHEON, Berlin 2005, and submitted.

Heitsch, H., Römisch, W.: Stability and scenario trees for multistage stochastic programs, Preprint 324, DFG Research Center MATHEON, Berlin 2006, and submitted.

Henrion, R., Römisch, W.: Metric regularity and quantitative stability in stochastic programs with probabilistic constraints, *Mathematical Programming* 84 (1999), 55–88.

Henrion, R., Römisch, W.: Hölder and Lipschitz stability of solution sets in programs with probabilistic constraints, *Mathematical Programming* 100 (2004), 589–611.

Henrion, R., Römisch, W.: Lipschitz and differentiability properties of quasi-concave and singular normal distribution functions, Preprint, DFG Research Center MATHEON, Berlin 2005, and to appear in *Annals of Operations Research*.

Rachev, S. T., Römisch, W.: Quantitative stability in stochastic programming: The method of probability metrics, *Mathematics of Operations Research* 27 (2002), 792–818.

Römisch, W., Wets, R. J-B: Stability of ε-approximate solutions to convex stochastic programs, Preprint 325, DFG Research Center MATHEON, Berlin 2006, and submitted.

Ruszczyński, A., Shapiro, A. (eds.): Stochastic Programming, Handbook of Operations Research and Management Science Vol. 10, Elsevier, Amsterdam, 2003.