Stochastic Programming: Tutorial

Part II

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

Contents (Part II)

Stability of stochastic programs

Consider the stochastic programming model

$$\min\left\{\int_\Xi f_0(x,\xi)P(d\xi) : x \in M(P)\right\}$$

$$M(P) := \left\{x \in X : \int_\Xi f_j(x,\xi)P(d\xi) \le 0, j = 1, ..., r\right\}$$

where f_j from $\mathbb{R}^m \times \Xi$ to the extended reals $\overline{\mathbb{R}}$ are normal integrands, X is a nonempty closed subset of $\mathbb{R}^m$, Ξ is a closed subset of $\mathbb{R}^d$ and P is a Borel probability measure on Ξ.

(f is a normal integrand if it is Borel measurable and $f(\xi, .)$ is lower semicontinuous $\forall \xi \in \Xi$.)

Let $\mathcal{P}(\Xi)$ the set of all Borel probability measures on Ξ and by

$$v(P) = \inf_{x\in M(P)} \int_\Xi f_0(x,\xi)P(d\xi) \quad \text{(optimal value)}$$

$$S_\varepsilon(P) = \left\{x \in M(P) : \int_\Xi f_0(x,\xi)P(d\xi) \le v(P) + \varepsilon\right\}$$

$$S(P) = S_0(P) = \arg\min_{x\in M(P)} \int_\Xi f_0(x,\xi)P(d\xi) \quad \text{(solution set).}$$

The underlying probability distribution P is often incompletely known in applied models and/or has to be approximated (estimated, discretized).
Hence, the stability behavior of stochastic programs becomes important when changing (perturbing, estimating, approximating) the probability distribution P on Ξ.

Stability refers to (quantitative) continuity properties of the optimal value function $v(.)$ and of the set-valued mapping $S_\varepsilon(.)$ at P, where both are regarded as mappings given on certain subset of $\mathcal{P}(\Xi)$ equipped with some probability metric.

(The corresponding subset of probability measures is determined by imposing certain moment conditions that are related to growth properties of the integrands f_j with respect to ξ.)

Examples: Two-stage and chance constrained stochastic programs.

Survey:
W. Römisch: Stability of stochastic programming problems, in: Stochastic Programming (A. Ruszczynski, A. Shapiro eds.), Handbook, Elsevier, 2003.

Weak convergence in $\mathcal{P}(\Xi)$

$$P_n \to_w P \ \text{ iff } \int_\Xi f(\xi)P_n(d\xi) \to \int_\Xi f(\xi)P(d\xi) \ \ (\forall f \in C_b(\Xi)),$$
$$\text{iff } \ P_n(\{\xi \le z\}) \to P(\{\xi \le z\}) \text{ at continuity points } z \text{ of } P(\{\xi \le \cdot\}).$$

Probability metrics on $\mathcal{P}(\Xi)$ (Monographs: Rachev 91, Rachev/Rüschendorf 98)

Metrics with ζ-structure:

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

where $\mathcal{F}$ is a suitable set of measurable functions from Ξ to $\overline{\mathbb{R}}$ and P, Q are probability measures in some set $\mathcal{P}_{\mathcal{F}}$ on which $d_{\mathcal{F}}$ is finite.

Examples (of $\mathcal{F}$): Sets of locally Lipschitzian functions on Ξ or of piecewise (locally) Lipschitzian functions.

There exist canonical sets $\mathcal{F}$ and metrics $d_{\mathcal{F}}$ for each specific class of stochastic programs!

General quantitative stability results

To simplify matters, let X be compact (otherwise, consider localizations).

$$\begin{aligned}\mathcal{F} &:= \{f_j(x,\cdot) \,:\, x \in X, j = 0,\dots,r\},\\ \mathcal{P}_{\mathcal{F}} &:= \Big\{Q \in \mathcal{P}(\Xi) : \int_\Xi \inf_{x\in X} f_j(x,\xi)Q(d\xi) > -\infty,\\ &\qquad\qquad \sup_{x\in X}\int_\Xi f_j(x,\xi)Q(d\xi) < \infty,\ j=0,\dots,r\Big\},\end{aligned}$$

and the probability (semi-) metric on $\mathcal{P}_{\mathcal{F}}$:

$$d_{\mathcal{F}}(P,Q) = \sup_{x\in X}\max_{j=0,\dots,r}\Big|\int_\Xi f_j(x,\xi)(P-Q)(d\xi)\Big|.$$

Lemma:
The functions $(x,Q) \mapsto \int_\Xi f_j(x,\xi)Q(d\xi)$ are lower semicontinuous on $X \times \mathcal{P}_{\mathcal{F}}$.

Theorem: (Rachev-Römisch 02)

If $d \geq 1$, let the function $x \mapsto \int_\Xi f_0(x,\xi)P(d\xi)$ be Lipschitz continuous on X, and, let the function

$$(x,y) \mapsto d\Big(x, \Big\{\tilde{x} \in X : \int_\Xi f_j(\tilde{x},\xi)P(d\xi) \leq y_j, j = 1,\dots,r\Big\}\Big)$$

be locally Lipschitz continuous around $(\bar{x}, 0)$ for every $\bar{x} \in S(P)$ (metric regularity condition).

Then there exist constants L, $\delta > 0$ such that

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

holds for all $Q \in \mathcal{P}_{\mathcal{F}}$ with $d_{\mathcal{F}}(P,Q) < \delta$.

Here, $\Psi_P(\eta) := \eta + \psi^{-1}(\eta)$ and $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ is given by

$$\psi(\tau) := \min\Big\{\int_\Xi f_0(x,\xi)P(d\xi) - v(P) : d(x, S(P)) \geq \tau, x \in M(P)\Big\}.$$

(Proof by appealing to general perturbation results see Klatte 94 and Rockafellar/Wets 98.)

Convex case and $r := 0$:

Assume that $f_0(\cdot, \xi)$ is convex on $\mathbb{R}^m$ for each $\xi \in \Xi$.

Theorem: (Römisch-Wets 07)

Then there exist constants L, $\bar{\varepsilon} > 0$ such that

$$d\!l_\infty(S_\varepsilon(P), S_\varepsilon(Q)) \leq \frac{L}{\varepsilon} d_{\mathcal{F}}(P, Q)$$

for every $\varepsilon \in (0, \bar{\varepsilon})$ and $Q \in \mathcal{P}_{\mathcal{F}}$ such that $d_{\mathcal{F}}(P, Q) < \varepsilon$.

Here, $d\!l_\infty$ is the Pompeiu-Hausdorff distance of nonempty closed subsets of $\mathbb{R}^m$, i.e.,

$$d\!l_\infty(C, D) = \inf\{\eta \geq 0 : C \subseteq D + \eta\mathbb{B}, D \subseteq C + \eta\mathbb{B}\}.$$

(Proof using a perturbation result see Rockafellar/Wets 98)

The (semi-) distance $d_{\mathcal{F}}$ plays the role of a minimal probability metric implying quantitative stability.

Furthermore, the result remains valid when bounding $d_{\mathcal{F}}$ from above by another distance and when reducing the set $\mathcal{P}_{\mathcal{F}}$ to a subset on which this distance is defined and finite.

Idea: Enlarge $\mathcal{F}$, but maintain the analytical (e.g., (dis)continuity) properties of $f_j(x, \cdot)$, $j = 0, \ldots, r$!

This idea may lead to well-known probability metrics, for which a well developed theory is available !

Example: (Fortet-Mourier-type metrics)

We consider the following classes of locally Lipschitz continuous functions (on Ξ)

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

where $H : \mathbb{R}_+ \to \mathbb{R}_+$ is nondecreasing, $H(0) = 0$. The corresponding distances are

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

so-called Fortet-Mourier-type 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: $H(t) := t^{p-1}$ for $p \ge 1$ leading to the notation $\mathcal{F}_p$, $\mathcal{P}_p(\Xi)$ and ζ_p, respectively.

(Convergence with respect to ζ_p means weak convergence of the probability measures and convergence of the p-th order moments (Rachev 91))

Stability of two-stage models

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

where $\Phi(u,t)$ is the optimal value function of the second-stage problem $\min\{\langle u,y\rangle : Wy=t, y\in Y\}$. We set

$$f_0(x,\xi)=\langle c,x\rangle+\Phi(q(\xi),h(\xi)-T(\xi)x)$$

if it is finite.

Proposition:

Assume (A1) and (A2). Then there exist $\hat{L}>0$ such that

$$\begin{aligned}|f_0(x,\xi)-f_0(x,\tilde{\xi})| &\le \hat{L}\max\{1,\|\xi\|,\|\tilde{\xi}\|\}\|\xi-\tilde{\xi}\|\\ |f_0(x,\xi)-f_0(\tilde{x},\xi)| &\le \hat{L}\max\{1,\|\xi\|^2\}\|x-\tilde{x}\|\end{aligned}$$

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

Theorem:

Assume (A1)–(A3) and let X be compact. Then there exist $L > 0$, $\bar{\varepsilon}$, $\delta > 0$ such that

$$\begin{aligned} |v(P) - v(Q)| &\leq L\zeta_2(P, Q), \\ S(Q) &\subseteq S(P) + \Psi_P(L\zeta_2(P, Q))\mathbb{B}, \\ \mathbb{d}_\infty(S_\varepsilon(P), S_\varepsilon(Q)) &\leq \frac{L}{\varepsilon}\zeta_2(P, Q), \end{aligned}$$

whenever Q satisfies $\zeta_2(P, Q) < \delta$, $\varepsilon \in (0, \bar{\varepsilon}]$,
$\Psi_P(\eta) := \eta + \psi^{-1}(\eta)$ and

$$\psi(\tau) := \min\left\{ \int_\Xi f_0(x, \xi) P(d\xi) - v(P) : d(x, S(P)) \geq \tau, x \in X \right\}.$$

Note ψ has quadratic growth (near 0) in a number of cases (Schultz 94) and linear growth if P is discrete.

Two-stage mixed-integer models

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

where Φ is given by

$$\Phi(u,t) := \inf \left\{ \langle u_1, y \rangle + \langle u_2, \bar{y} \rangle : Wy + \bar{W}\bar{y} \leq t, y \in \mathbb{Z}^{\hat{m}}, \bar{y} \in \mathbb{R}^{\bar{m}} \right\}.$$

The quantitative stability result holds with respect to the distance $\zeta_{2,\mathrm{ph}}$ on $\mathcal{P}_2(\Xi)$:

$$\zeta_{2,\mathrm{ph}}(P,Q) := \sup \left\{ \left| \int_B f(\xi)(P-Q)(d\xi) \right| \,\middle|\, \begin{array}{l} f \in \mathcal{F}_2(\Xi) \\ B \in \mathcal{B}_{\mathrm{ph}}(\Xi) \end{array} \right\}$$

Here, $\mathcal{B}_{\mathrm{ph}}(\Xi)$ is a class of polyhedral subsets of Ξ and $\mathcal{F}_2(\Xi)$ contains all functions $f : \Xi \to \mathbb{R}$ such that

$$|f(\xi)| \leq \max\{1, \|\xi\|^2\}, \; |f(\xi) - f(\tilde{\xi})| \leq \max\{1, \|\xi\|, \|\tilde{\xi}\|\} \|\xi - \tilde{\xi}\|$$

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

(Römisch/Vigerske 08)

Chance constrained models

$$\min\{\langle c,x\rangle : x\in X, P(\{\xi\in\Xi : T(\xi)x\ge h(\xi)\})\ge p\},$$

where $c\in\mathbb{R}^m$, X and Ξ are polyhedra in $\mathbb{R}^m$ and $\mathbb{R}^s$, respectively, $p\in(0,1)$, $P\in\mathcal{P}(\Xi)$, and the right-hand side $h(\xi)\in\mathbb{R}^d$ and the (d,m)-matrix $T(\xi)$ are affine functions of ξ.

By specifying the general (semi-) distance we obtain

$$\begin{aligned} d_{\mathcal{F}}(P,Q) &:= \sup_{x\in X}\max_{j=0,1}\left|\int_\Xi f_j(x,\xi)(P-Q)(d\xi)\right| \\ &= \sup_{x\in X}|P(H(x))-Q(H(x))|, \end{aligned}$$

where $f_0(\xi,x)=\langle c,x\rangle$, $f_1(\xi,x)=p-\mathbb{1}_{H(x)}(\xi)$ and $H(x)=\{\xi\in\Xi : T(\xi)x\ge h(\xi)\}$ (polyhedral subsets of Ξ).

Hence, the quantitative stability result holds with respect to polyhedral discrepancies:

$$\alpha_{\mathrm{ph}}(P,Q)=\sup_{B\in\mathcal{B}_{\mathrm{ph}}(\Xi)}|P(B)-Q(B)|$$

Quantitative Stability of multistage models

We consider the linear multistage model

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

where X_1 is bounded polyhedral and X_t, $t = 2,\ldots,T$, are polyhedral cones, the vectors $b_t(\cdot)$, $h_t(\cdot)$ and $A_{t,1}(\cdot)$ are affine functions of ξ_t, and $\xi = (\xi_t)_{t=1}^{T}$ a stochastic process with

$$\mathcal{F}_t(\xi) = \sigma(\xi_1,\ldots,\xi_t) \ (t = 1,\ldots,T).$$

Let F denote the objective function defined on $L_{r'}(\Omega,\mathcal{F},\mathbb{P};\mathbb{R}^s) \times L_r(\Omega,\mathcal{F},\mathbb{P};\mathbb{R}^m) \to \mathbb{R}$ by

$$F(x,\xi) := \mathbb{E}\Big[\sum_{t=1}^{T}\langle b_t(\xi_t), x_t\rangle\Big],$$

where $r \geq 1$ and

$$r' := \begin{cases} \frac{r}{r-1} & , \text{ if only costs are random} \\ r & , \text{ if only right-hand sides are random} \\ 2 & , \text{ if costs and right-hand sides are random} \\ \infty & , \text{ if all technology matrices are random and } r = T. \end{cases}$$

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

$$\mathcal{X}(\xi) := \{x \in L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m) : x_1 \in X_1, x_t \in \mathcal{X}_t(x_{t-1}; \xi_t)\}$$

the set of feasible elements with input ξ.
Then the multistage stochastic program may be rewritten as

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

where $\mathcal{N}_{r'}(\xi)$ is the nonanticipativity subspace of $L_{r'}$.

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

$$S_\varepsilon(\xi) := \{x \in \mathcal{X}(\xi) \cap \mathcal{N}_{r'}(\xi) : F(x,\xi) \leq v(\xi) + \varepsilon\}$$
$$S(\xi) := S_0(\xi)$$

denote the ε-approximate solution set and the solution set of the stochastic program with input ξ.

Assumptions:

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

(A2) There exists a $\delta > 0$ such that for any $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ 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) Assume that the optimal values $v(\tilde{\xi})$ are finite if $\|\xi - \tilde{\xi}\|_r \leq \delta$ and that the objective function F is level-bounded locally uniformly at ξ, i.e., for some $\varepsilon > 0$ there exists a bounded subset B of $L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ such that $S_\varepsilon(\tilde{\xi})$ is contained in B if $\|\tilde{\xi} - \xi\|_r \leq \delta$.

Theorem: (Heitsch/Römisch/Strugarek 06)

Let (A1) – (A3) be satisfied and X_1 be bounded.
Then there exist positive constants L and δ such that

$$|v(\xi) - v(\tilde{\xi})| \le L(\|\xi - \tilde{\xi}\|_r + d_{\mathrm{f},T-1}(\xi, \tilde{\xi}))$$

holds for all $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ with $\|\tilde{\xi} - \xi\|_r \le \delta$.
If $1 < r' < \infty$ and $(\xi^{(n)})$ converges to ξ in L_r and with respect to $d_{\mathrm{f},T}$, then any sequence $x_n \in S(\xi^{(n)})$, $n \in \mathbb{N}$, contains a subsequence converging weakly in $L_{r'}$ to some element of $S(\xi)$.

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

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

Remark:

For $T = 2$ one obtains the same result for the optimal values as in the two-stage case ! However, one obtains weak convergence of subsequences of (random) second-stage solutions, too!

Different approach by Plug/Pichler based on distances of conditional probability distributions.

Empirical or Monte Carlo approximations of stochastic programs

Given a probability distribution $P \in \mathcal{P}(\Xi)$, we consider a sequence $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ of independent, identically distributed Ξ-valued random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ having the common distribution P.

We consider the empirical measures

$$P_n(\omega) := \frac{1}{n}\sum_{i=1}^{n} \delta_{\xi_i(\omega)}$$

for every $n \in \mathbb{N}$.

Empirical or sample average approximation of stochastic programs (replacing P by $P_n(\cdot)$):

$$\min\Big\{\frac{1}{n}\sum_{i=1}^{n} f_0(\xi_i, x) : x \in X, \frac{1}{n}\sum_{i=1}^{n} f_j(\xi_i, x) \leq 0, j = 1, \ldots, r\Big\}$$

To study convergence of empirical approximations, one may use the quantitative stability results by deriving estimates of the (uniform) distances

$$d_{\mathcal{F}}(P, P_n(\cdot))$$

Tool: Empirical process theory, in particular, the size of $\mathcal{F}$ as subset of $L_p(\Xi, P)$ measured by covering numbers, where

$$\mathcal{F} = \{f_j(x, \cdot) : x \in X, j = 0, \ldots, r\}.$$

Empirical process (indexed by some class of functions):

$$\left\{ n^{\frac{1}{2}}(P_n(\cdot) - P)f = n^{-\frac{1}{2}} \sum_{i=1}^{n} \left(f(\xi_i(\cdot)) - \int_{\Xi} f(\xi) P(d\xi) \right) \right\}_{f \in \mathcal{F}}$$

Desirable estimate:

$$\mathbb{P}\left(\left\{\omega : n^{\frac{1}{2}} d_{\mathcal{F}}(P, P_n(\omega)) \geq \varepsilon\right\}\right) \leq C_{\mathcal{F}}(\varepsilon) \quad (\forall \varepsilon > 0, n \in \mathbb{N})$$

for some tail function $C_{\mathcal{F}}(\cdot)$ defined on $(0, +\infty)$ and decreasing to 0, in particular, exponential tails $C_{\mathcal{F}}(\varepsilon) = K\varepsilon^r \exp(-2\varepsilon^2)$.

If $N(\varepsilon, L_p(Q))$ denotes the minimal number of open balls $\{g : \|g - f\|_{Q,p} < \varepsilon\}$ needed to cover $\mathcal{F}$, then an estimate of the form

$$\sup_Q N(\varepsilon, L_2(Q)) \leq \left(\frac{R}{\varepsilon}\right)^r$$

for some $r, R \geq 1$ and all $\varepsilon > 0$, is needed to obtain exponential tails.

(Literature: Talagrand 94, van der Vaart/Wellner 96, van der Vaart 98)

Typical result for optimal values:

$$\mathbb{P}\left(|v(P) - v(P_n)| \geq \varepsilon\, n^{-\frac{1}{2}}\right) \leq C_{\mathcal{F}}\left(\min\{\delta, \varepsilon\, L^{-1}\}\right)$$

Such results are available for two-stage (mixed-integer) and chance constrained stochastic programs (Römisch 03).

Desirable results for optimal values: Limit theorems

$$n^{\frac{1}{2}}(v(P_n(\cdot)) - v(P)) \longrightarrow z,$$

where z is a real random variable and the convergence is *convergence in distribution*.

Such results can be derived if $\mathcal{F}$ is a Donsker class of functions. Donsker classes can also be characterized via covering numbers.

Examples for available limit theorems:

- Limit theorem for optimal values of mixed-integer two-stage stochastic programs (Eichhorn/Römisch 07).
- Limit theorem for optimal values of kth order stochastic dominance constrained stochastic programs for $k \geq 2$

  (Dentcheva/Römisch 12).

(Chapters by Shapiro and Pflug in the Handbook 2003; recent work of Shapiro, Xu and coworkers)

Scenario generation methods

Assume that we have to solve a stochastic program with a class $\mathcal{F} = \{f_j(x,\cdot) : x \in X, j = 1,\ldots,r\}$ of functions on $\Xi \subseteq \mathbb{R}^d$ and probability (semi-) metric

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

Optimal scenario generation:

For given $n \in \mathbb{N}$ and probabilities $p_i = \frac{1}{n}$, $i = 1,\ldots,n$, the best possible choice of scenarios $\xi_i \in \Xi$, $i = 1,\ldots,n$, is obtained by solving the best approximation problem

$$\min \Big\{ d_{\mathcal{F}}\big(P, \tfrac{1}{n}\textstyle\sum_{i=1}^n \delta_{\xi_i}\big); \xi_i \in \Xi, i = 1,\ldots,n \Big\}.$$

However, this is a large-scale, nonsmooth and nonconvex minimization problem (of dimension $n \cdot d$) and often extremely difficult to solve. Note that, in addition, function calls for $f_j(x,\cdot)$ are often expensive and the appropriate choice of $n \in \mathbb{N}$ is difficult.

Next we discuss 4 specific scenario generation methods for stochastic programs (*without information constraints*) based on (high-dimensional) numerical integration methods:

(a) Monte Carlo sampling from the underlying probability distribution P on $\mathbb{R}^d$ (Shapiro 03).

(b) Optimal quantization of probability distributions (Pflug-Pichler 11).

(c) Quasi-Monte Carlo methods (Koivu-Pennanen 05, Homem-de-Mello 08).

(d) Quadrature rules based on sparse grids (Chen-Mehrotra 08).

Given an integral

$$I_d(f) = \int_{\mathbb{R}^d} f(\xi)\rho(\xi)d\xi \quad \text{or} \quad I_d(f) = \int_{[0,1]^d} f(\xi)d\xi$$

a numerical integration method means

$$Q_{n,d}(f) = \frac{1}{n}\sum_{i=1}^{n} f(\xi_i).$$

Monte Carlo sampling

Monte Carlo methods are based on drawing independent identically distributed (iid) Ξ-valued random samples $\xi^1(\cdot), \ldots, \xi^n(\cdot), \ldots$ (defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$) from an underlying probability distribution P (on Ξ) such that

$$Q_{n,d}(\omega)(f) = \frac{1}{n} \sum_{i=1}^{n} f(\xi^i(\omega)),$$

i.e., $Q_{n,d}(\cdot)$ is a random functional, and it holds

$$\lim_{n\to\infty} Q_{n,d}(\omega)(f) = \int_{\Xi} f(\xi)P(d\xi) = \mathbb{E}(f) \quad \mathbb{P}\text{-almost surely}$$

for every real continuous and bounded function f on Ξ.
If P has finite moment of order $r \geq 1$, the error estimate

$$\mathbb{E}\left(\left|\frac{1}{n}\sum_{i=1}^{n} f(\xi^i(\omega)) - \mathbb{E}(f)\right|^r\right) \leq \frac{\mathbb{E}\left((f - \mathbb{E}(f))^r\right)}{n^{r-1}}$$

is valid. Hence, the mean square convergence rate is

$$\|Q_{n,d}(\omega)(f) - \mathbb{E}(f)\|_{L_2} = \sigma(f)n^{-\frac{1}{2}},$$

where $\sigma^2(f) = \mathbb{E}\left((f - \mathbb{E}(f))^2\right)$.
The latter holds without any assumption on f except $\sigma(f) < \infty$.

Advantages:
(i) MC sampling works *for (almost) all integrands*.
(ii) The machinery of probability theory is available.
(iii) The convergence *rate does not depend on* d.

Deficiencies: (Niederreiter 92)
(i) There exist 'only' *probabilistic error bounds*.
(ii) Possible regularity of the integrand *does not improve* the rate.
(iii) Generating (independent) random samples is *difficult*.

Practically, iid samples are approximately obtained by pseudo random number generators as uniform samples in $[0, 1]^d$ and later transformed to more general sets Ξ and distributions P.

Survey: L'Ecuyer 94.

Classical generators for pseudo random numbers are based on linear congruential methods. As the parameters of this method, we choose a large $M \in \mathbb{N}$ (*modulus*), a *multiplier* $a \in \mathbb{N}$ with $1 \leq a < M$ and $gcd(a, M) = 1$, and $c \in Z_M = \{0, 1, \ldots, M-1\}$. Starting with $y_0 \in Z_M$ a sequence is generated by

$$y_n \equiv a y_{n-1} + c \mod M \qquad (n \in \mathbb{N})$$

and the linear congruential pseudo random numbers are

$$\xi^n = \frac{y_n}{M} \in [0, 1).$$

Excellent pseudo random number generator: Mersenne Twister (Matsumoto-Nishimura 98).

Use only pseudo random number generators that passed a series of statistical tests, e.g., uniformity test, serial correlation test, serial test, coarse lattice structure test etc.

Optimal quantization of probability measures

Assume that the underlying stochastic program behaves stable with respect to a distance d of probability measures on $\mathbb{R}^d$.

Examples:

(a) Fortet-Mourier metric ζ_r of order r,

(b) L_r-minimal metric ℓ_r (or Wasserstein metric), i.e.

$$\ell_r(P,Q) = \inf\{(\mathbb{E}(\|\xi - \eta\|^r))^{\frac{1}{r}} : \mathcal{L}(\xi) = P,\, \mathcal{L}(\eta) = Q\}$$

Let P be a given probability distribution on $\mathbb{R}^d$. We are looking for a discrete probability measure Q_n with support

$$\mathrm{supp}(Q_n) = \{\xi^1, \ldots, \xi^n\} \quad \text{and} \quad Q_n(\{\xi^i\}) = \frac{1}{n},\ i = 1, \ldots, n,$$

such that it is the best approximation to P with respect to d, i.e.,

$$d(P, Q_n) = \min\{d(P,Q) : |\mathrm{supp}(Q)| = n, Q \text{ is uniform}\}.$$

Existence of best approximations, called optimal quantizers, and their convergence rates are well known for ℓ_r (Graf-Luschgy 00).

Note, however, $\ell_r(P, Q_n) \geq cn^{-\frac{1}{d}}$ for some $c > 0$ and all $n \in \mathbb{N}$.

In general, the function

$$\Psi_d(\xi^1, \ldots, \xi^n) := d\Big(P, \frac{1}{n}\sum_{i=1}^{n} \delta_{\xi^i}\Big)$$

$$\Psi_{\ell_r}(\xi^1, \ldots, \xi^n) = \Big(\int_{\mathbb{R}^d} \min_{i=1,\ldots,n} \|\xi - \xi^i\|^r P(d\xi)\Big)^{\frac{1}{r}}$$

is nonconvex and nondifferentiable on $\mathbb{R}^{dn}$.
Hence, the global minimization of Ψ_d is not an easy task.

Algorithmic procedures for minimizing Ψ_{ℓ_r} globally may be based on stochastic gradient algorithms, stochastic approximation methods and stochastic branch-and-bound techniques (e.g. Pflug 01, Hochreiter-Pflug 07, Pagés 97, Pagés et al 04).

Asymptotically optimal quantizers can be determined explicitly in a number of cases (Pflug-Pichler 11).

Quasi-Monte Carlo methods

The basic idea of Quasi-Monte Carlo (QMC) methods is to replace random samples in Monte Carlo methods by deterministic points that are uniformly distributed in $[0,1]^d$. The latter property may be defined in terms of the so-called star-discrepancy of $\xi^1,\ldots,\xi^n$

$$D_n^*(\xi^1,\ldots,\xi^n) := \sup_{\xi\in[0,1]^d}\left|\lambda^d([0,\xi)) - \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{[0,\xi)}(\xi^i)\right|,$$

by calling a sequence $(\xi^i)_{i\in\mathbb{N}}$ uniformly distributed in $[0,1]^d$

$$D_n^*(\xi^1,\ldots,\xi^n) \to 0 \quad \text{for} \quad n\to\infty\,.$$

A **classical result** due to Roth 54 states

$$D_n^*(\xi^1,\ldots,\xi^n) \geq B_d \frac{(\log n)^{\frac{d-1}{2}}}{n}$$

for some constant B_d and all sequences (ξ^i) in $[0,1]^d$.

Classical convergence results:

Theorem: (Proinov 88)
If the real function f is continuous on $[0,1]^d$, then there exists $C > 0$ such that

$$|Q_{n,d}(f) - I_d(f)| \leq C\omega_f\big(D_n^*(\xi^1,\ldots,\xi^n)^{\frac{1}{d}}\big),$$

where $\omega_f(\delta) = \sup\{|f(\xi) - f(\tilde{\xi})| : \|\xi - \tilde{\xi})\| \leq \delta,\ \xi, \tilde{\xi} \in [0,1]^d\}$ is the modulus of continuity of f.

Theorem: (Koksma-Hlawka 61)
If f is of bounded variation $V_{\mathrm{HK}}(f)$ in the sense of Hardy and Krause, it holds

$$\big|I_d(f) - Q_{n,d}(f)\big| \leq V_{\mathrm{HK}}(f) D_n^*(\xi^1,\ldots,\xi^n)\,.$$

for any $n \in \mathbb{N}$ and any $\xi^1,\ldots,\xi^n \in [0,1]^d$.
There exist sequences (ξ^i) in $[0,1]^d$ such that

$$D_n^*(\xi^1,\ldots,\xi^n) = O(n^{-1}(\log n)^{d-1}),$$

however, the constant depends on the dimension d.

First general construction: (Sobol 69, Niederreiter 87)

Elementary subintervals E in base b:

$$E = \prod_{j=1}^{d} \left[\frac{a_j}{b^{d_j}}, \frac{a_j + 1}{b^{d_j}} \right),$$

with $a_i, d_i \in \mathbb{Z}_+, 0 \le a_i < d_i, i = 1, \ldots, d$.

Let $m, t \in \mathbb{Z}_+$, $m > t$.
A set of b^m points in $[0,1]^d$ is a (t, m, d)-**net** in base b if every elementary subinterval E in base b with $\lambda^d(E) = b^{t-m}$ contains b^t points.

A sequence (ξ^i) in $[0,1]^d$ is a (t, d)-**sequence** in base b if, for all integers $k \in \mathbb{Z}_+$ and $m > t$, the set

$$\{\xi^i : kb^m \le i < (k+1)b^m\}$$

is a (t, m, d)-net in base b.

Specific sequences: Faure, Sobol', Niederreiter and Niederreiter-Xing sequences (Lemieux 09, Dick-Pillichshammer 10).

Recent development: Scrambled (t, m, d)-nets, where the digits are randomly permuted (Owen 95).

Second general construction: (Korobov 59, Sloan-Joe 94)

Lattice rules: Let $g \in \mathbb{Z}^d$ and consider the lattice points

$$\left\{\xi^i = \left\{\frac{i}{n} g\right\} : i = 1, \ldots, n\right\},$$

where $\{z\}$ is defined componentwise and is the *fractional part* of $z \in \mathbb{R}_+$, i.e., $\{z\} = z - \lfloor z \rfloor \in [0, 1)$.
The generator g is chosen such that the lattice rule has good convergence properties.

Such lattice rules may achieve better convergence rates $O(n^{-k+\delta})$, $k \in \mathbb{N}$, for smooth integrands.

Recent development: Randomized lattice rules.

Randomly shifted lattice points:

$$\left\{\xi^i = \left\{\frac{i}{n}g + \triangle\right\} : i = 1, \ldots, n\right\},$$

where $\triangle$ is uniformly distributed in $[0,1]^d$.

There is a component-by-component construction algorithm for g such that for some constant $C(\delta)$ and all $0 < \delta \leq \frac{1}{2}$ the optimal convergence rate

$$e(Q_{n,d}) \leq C(\delta) n^{-1+\delta} \quad (n \in \mathbb{N})$$

is achieved if the integrand f belongs to the tensor product Sobolev space

$$\mathbb{F}_d = W_2^{(1,\ldots,1)}([0,1]^d) = \bigotimes_{i=1}^{d} W_2^1([0,1])$$

equipped with a weighted norm. Since the space $\mathbb{F}_d$ is a kernel reproducing Hilbert space, a well developed technique for estimating the quadrature error can be used.

(Hickernell 96, Sloan/Wožniakowski 98, Sloan/Kuo/Joe 02, Kuo 03)

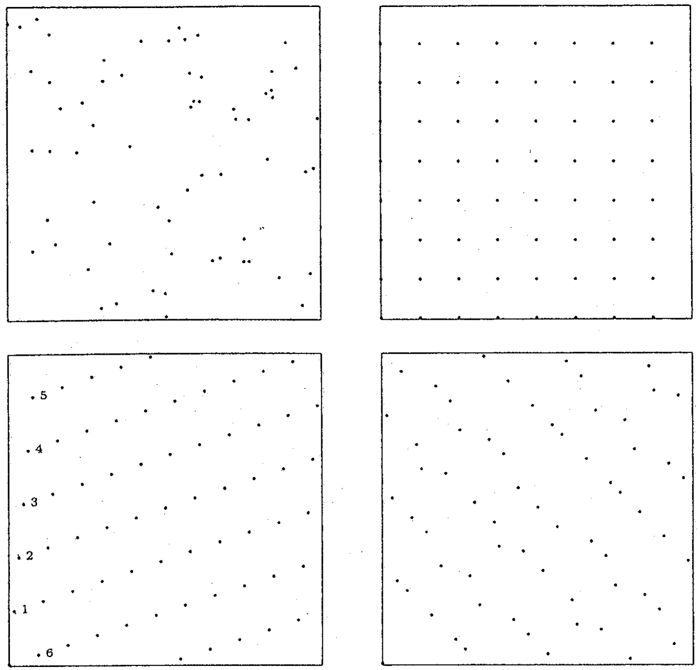


Fig. 5.3 Four different point sets with $n = 64$: random (top left), rectangular grid (top right), Korobov lattice (bottom left), and Sobol' (bottom right).

Home Page
Title Page
Contents
Page 35 of 77
Go Back
Full Screen
Close
Quit

Is QMC efficient in stochastic programming ?

Problem: Typical integrands in linear stochastic programming are not of bounded variation in the HK sense and nonsmooth and, hence, do not belong to the relevant function space $\mathbb{F}_d$ in general.

Idea: Study the ANOVA decomposition and efficient dimension of two-stage integrands.

ANOVA-decomposition of f:

$$f = \sum_{u \subseteq D} f_u \,,$$

where $f_\emptyset = I_d(f) = I_D(f)$ and recursively

$$f_u = I_{-u}(f) + \sum_{v \subseteq u} (-1)^{|u|-|v|} I_{u-v}(I_{-u}(f)) \,,$$

where I_{-u} means integration with respect to ξ_j in $[0,1]$, $j \in D \setminus u$ and $D = \{1, \ldots, d\}$. Hence, f_u is essentially as smooth as $I_{-u}(f)$ and does not depend on ξ^{-u}.

We set $\sigma^2(f) = \|f - I_d(f)\|_{L_2}^2$ and have

$$\sigma^2(f) = \|f\|_{L_2}^2 - (I_d(f))^2 = \sum_{\emptyset \neq u \subseteq D} \|f_u\|_{L_2}^2 .$$

The superposition dimension d_s of f is the smallest $d_s \in \mathbb{N}$ with

$$\sum_{|u| \leq d_s} \|f_u\|_{L_2}^2 \geq (1 - \varepsilon)\sigma^2(f) \quad \text{(where } \varepsilon \in (0,1) \text{ is small).}$$

Then

$$\|f - \sum_{|u| \leq d_s} f_u\|_{L_2}^2 \leq \varepsilon \sigma^2(f).$$

Result:

All ANOVA terms f_u, $u \subset D$, $u \neq D$, of integrands in two-stage stochastic programming belong to C^∞ if the underlying marginal densities belong to $C_b^\infty(\mathbb{R})$ and certain geometric condition is satisfied (Heitsch/Leovey/Römisch 12).

Hence, after reducing the efficient superposition dimension of f such that (at least) $d_s \leq d - 1$ holds, QMC methods should have optimal rates.

Quadrature rules with sparse grids

Again we consider the unit cube $[0,1]^d$ in $\mathbb{R}^d$. Let nested sets of grids in $[0,1]$ be given, i.e.,

$$\Xi^i = \{\xi_1^i, \ldots, \xi_{m_i}^i\} \subset \Xi^{i+1} \subset [0,1] \quad (i \in \mathbb{N}),$$

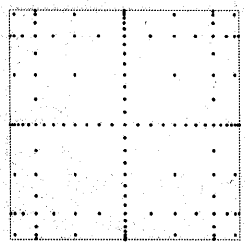
for example, the dyadic grid

$$\Xi^i = \left\{ \frac{j}{2^i} : j = 0, 1, \ldots, 2^i \right\}.$$

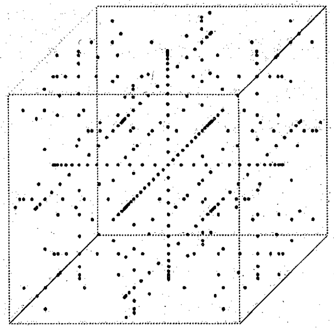
Then the point set suggested by Smolyak (Smolyak 63)

$$H(n,d) := \bigcup_{\sum_{j=1}^d i_j = n} \Xi^{i_1} \times \cdots \times \Xi^{i_d} \qquad (n \in \mathbb{N})$$

is called a sparse grid in $[0,1]^d$. In case of dyadic grids in $[0,1]$ the set $H(n,d)$ consists of all d-dimensional dyadic grids with product of mesh size given by $\frac{1}{2^n}$.



(a) $d = 2$



(b) $d = 3$

The corresponding tensor product quadrature rule for $n \geq d$ on $[0,1]^d$ with respect to the Lebesgue measure λ^d is of the form

$$Q_{n,d}(f) = \sum_{n-d+1 \leq |\mathbf{i}| \leq n} (-1)^{n-|\mathbf{i}|} \binom{d-1}{n-|\mathbf{i}|} \sum_{j_1=1}^{m_{i_1}} \cdots \sum_{j_d=1}^{m_{i_d}} f(\xi_{j_1}^{i_1}, \ldots, \xi_{j_d}^{i_d}) \prod_{l=1}^{d} a_{j_l}^{i_l},$$

where $|\mathbf{i}| = \sum_{j=1}^{d} i_j$ and the coefficients a_j^i ($j = 1, \ldots, m_i$, $i = 1, \ldots, d$) are weights of one-dimensional quadrature rules.

Even if the one-dimensional weights are positive, some of the weights w_i may become negative. Hence, an interpretation as discrete probability measure is no longer possible.

Theorem: (Bungartz-Griebel 04)
If f belongs to $\mathbb{F}_d = W_2^{(r,\ldots,r)}([0,1]^d)$, it holds

$$\left| \int_{[0,1]^d} f(\xi) d\xi - \sum_{i=1}^{n} w_i f(\xi^i) \right| \leq C_{r,d} \|f\|_d \frac{(\log n)^{(d-1)(r+1)}}{n^r}.$$

Scenario reduction

Assume that a two-stage stochastic program behaves stable with respect to a Fortet-Mourier metric ζ_r for some $r \geq 1$.

Proposition: (Rachev-Rüschendorf 98)
If Ξ is bounded, ζ_r may be reformulated as transportation problem

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

where $\hat{c}_r$ is a metric (reduced cost) with $\hat{c}_r \leq c_r$ and given by

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

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

Best approximation given a scenario set J:

The best approximation of P with respect to ζ_r by such a distribution Q exists and is denoted by Q^*. It has the distance

$$D_J := \zeta_r(P, Q^*) = \min_{Q} \zeta_r(P, Q) = \sum_{i \in J} p_i \min_{j \notin J} \hat{c}_r(\xi^i, \xi^j)$$

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 $j(i) \in \arg\min_{j \notin J} \hat{c}_r(\xi^i, \xi^j)$, $\forall i \in J$

(optimal redistribution) (Dupačová-Gröwe-Römisch 03).

For mixed-integer two-stage stochastic programs the relevant distance is a polyhedral discrepancy. In that case, the new weights have to be determined by linear programming (Henrion-Küchler-Römisch 08, 09).

Determining the optimal index set J with prescribed cardinality $N - n$ is a clustering problem, thus, a combinatorial optimization problem of n-median type:

$$\min \{D_J : J \subset \{1, ..., N\}, |J| = N - n\}$$

Hence, the problem of finding the optimal set J for deleting scenarios is $\mathcal{NP}$-hard and polynomial time algorithms are not available in general.

Development fast heuristics starting from $n = 1$ or $n = N - 1$.

Fast reduction heuristics

Starting point ($n = N - 1$): $\min\limits_{l \in \{1,\dots,N\}} p_l \min\limits_{j \neq l} \hat{c}_r(\xi^l, \xi^j)$

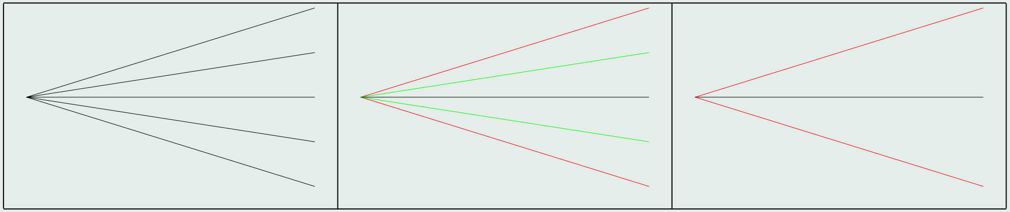
Algorithm 1: (Backward reduction)

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

Step [i]: $l_i \in \arg\min\limits_{l \notin J^{[i-1]}} \sum\limits_{k \in J^{[i-1]} \cup \{l\}} p_k \min\limits_{j \notin J^{[i-1]} \cup \{l\}} \hat{c}_r(\xi^k, \xi^j).$

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

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

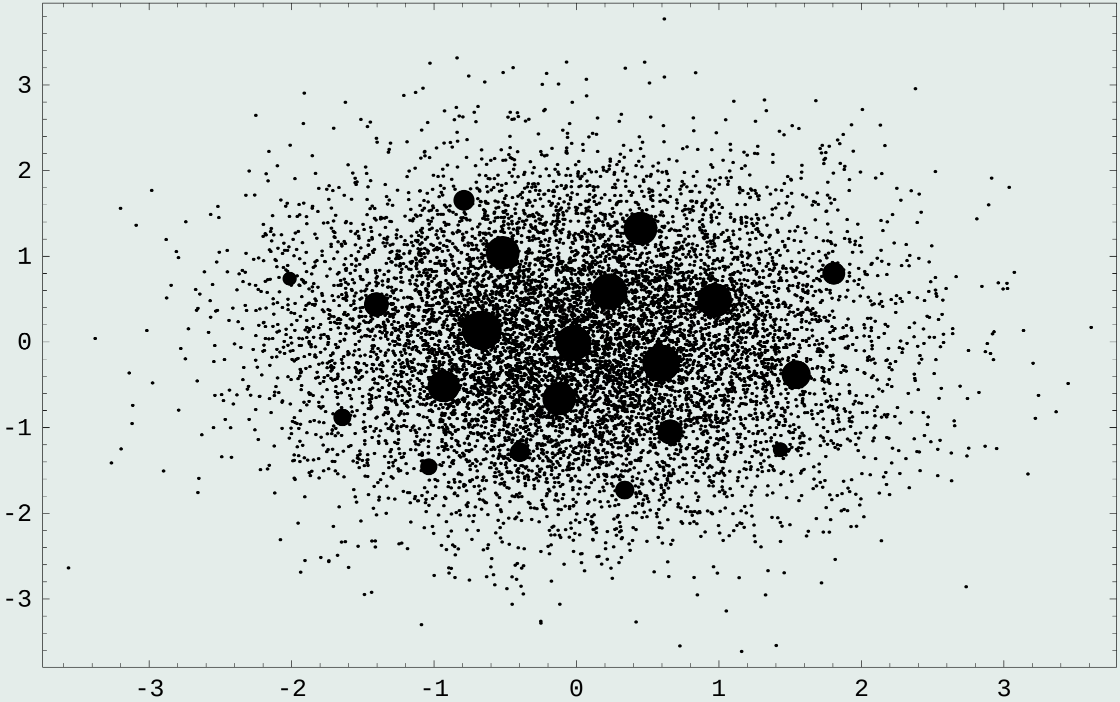
$$J^{[i]} := J^{[i-1]} \setminus \{u_i\}\,.$$

Step [n+1]: Optimal redistribution.

(Heitsch-Römisch 03, 07)

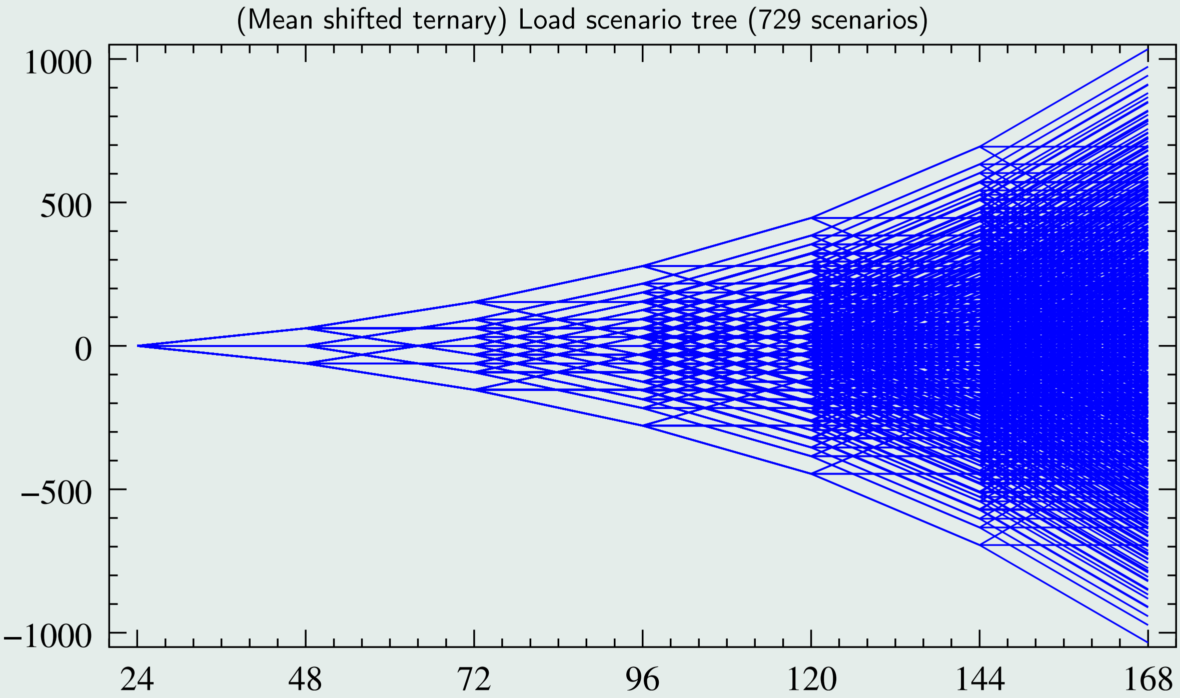
Home Page
Title Page
Contents
◀◀ ▶▶
◀ ▶

Go Back
Full Screen
Close
Quit



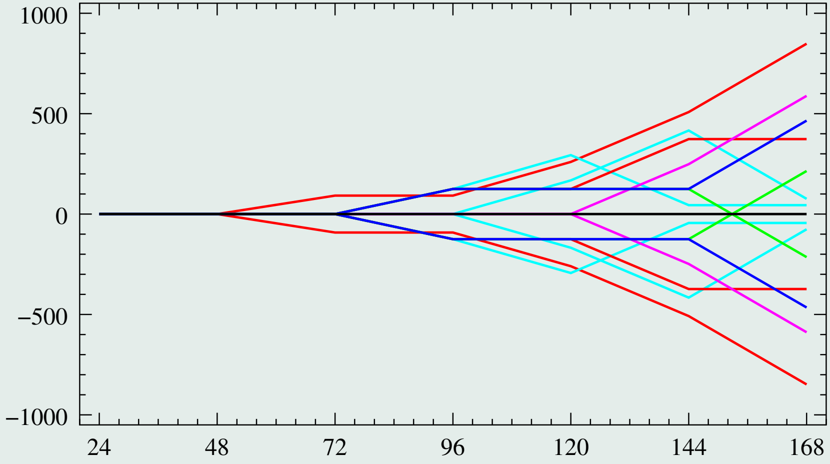
Scenario reduction w.r.t. ℓ_1 from N=10 000 MC samples of $N(0, I)$ in $\mathbb{R}^2$ to $n = 20$. The diameters of the circles are proportional to their probabilities

Example: (Electrical load scenario tree)

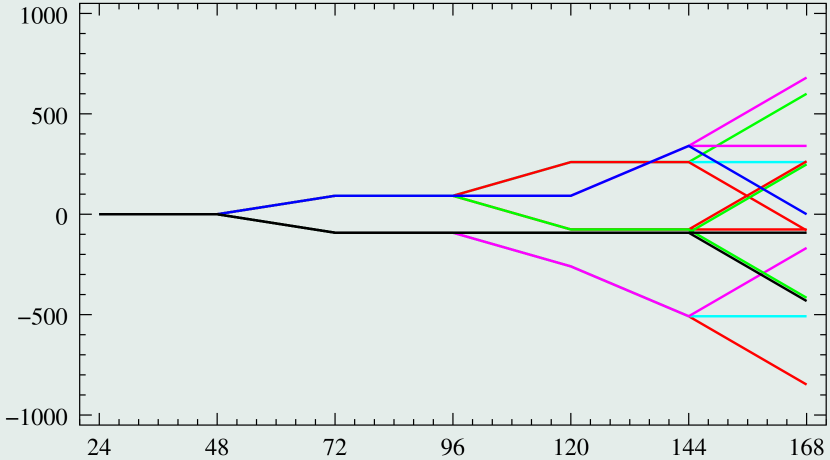


<Start Animation>

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



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



Generation of scenario trees

In multistage stochastic programs the decisions x satisfy the information constraint that x_t is measurable with respect to $\mathcal{F}_t = \sigma(\xi_1, \ldots, \xi_t)$, $t = 1, \ldots, T$. The increase of the σ-fields $\mathcal{F}_t$ w.r.t. t is reflected by approximating the underlying stochastic process $\xi = (\xi_t)_{t=1}^T$ by scenarios forming a scenario tree.

Some recent approaches:

(1) Bound-based approximation methods: Kuhn 05, Casey-Sen 05.

(2) Monte Carlo-based schemes: Shapiro 03, 06.

(3) Quasi-Monte Carlo methods: Pennanen 06, 09 .

(4) Moment-matching principle: Høyland-Kaut-Wallace 03.

(5) Optimal quantization: Pagés et al. 99.

(6) Stability-based approximations: Hochreiter-Pflug 07, Pflug-Pichler 11, Heitsch-Römisch 09, 10.

Survey: Dupačová-Consigli-Wallace 00

Theoretical basis of (6):

Quantitative stability results for multi-stage stochastic programs.

(Heitsch-Römisch-Strugarek 06; Mirkov-Pflug 07, Pflug 09, Pflug-Pichler 11)

Scenario tree generation: (Heitsch-Römisch 09)

(i) Generate a number of scenarios by one of the methods discussed earlier.

(ii) Construction of a scenario tree out of these scenarios by recursive scenario reduction and bundling over time such that the optimal value stays within a prescribed tolerance.

Implementation: GAMS-SCENRED 2.0 (developed by H. Heitsch)

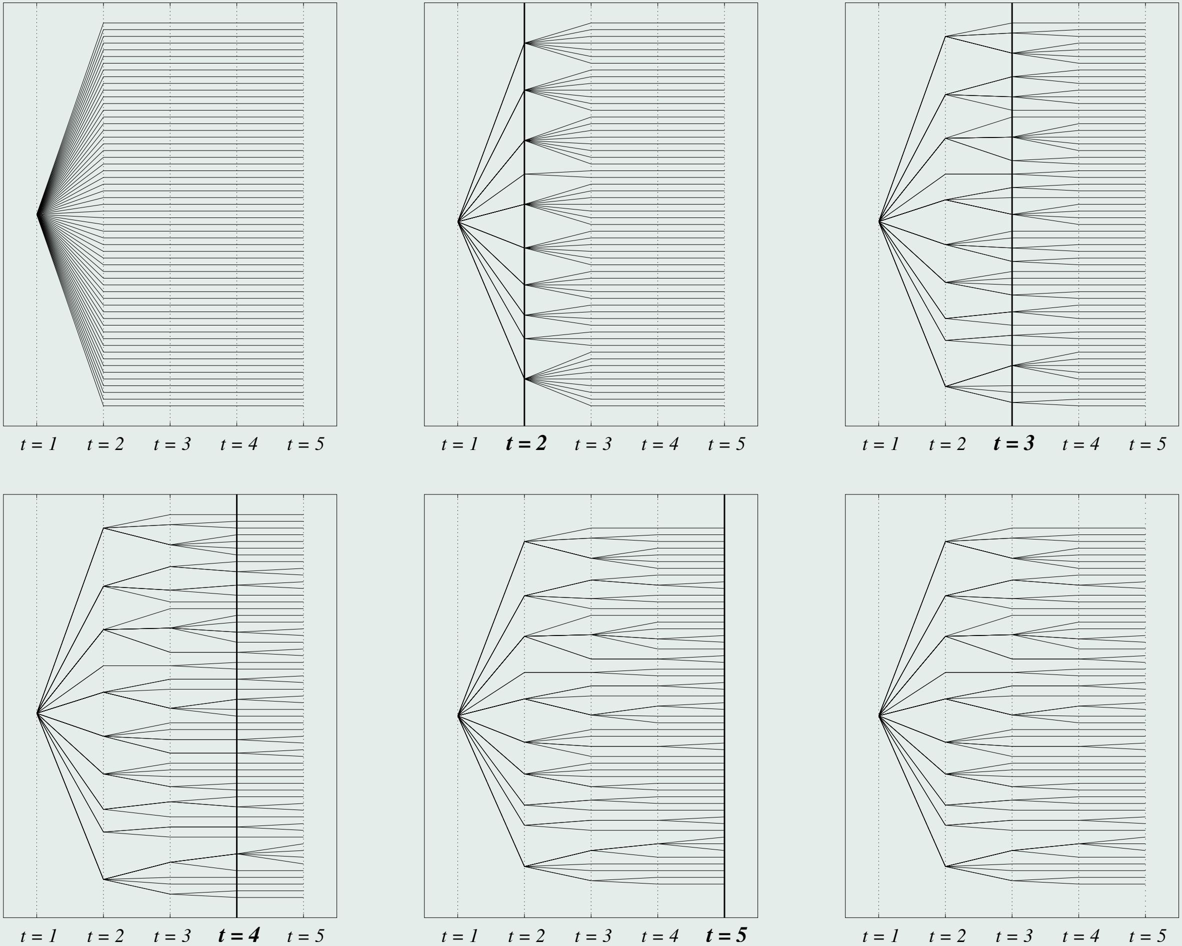
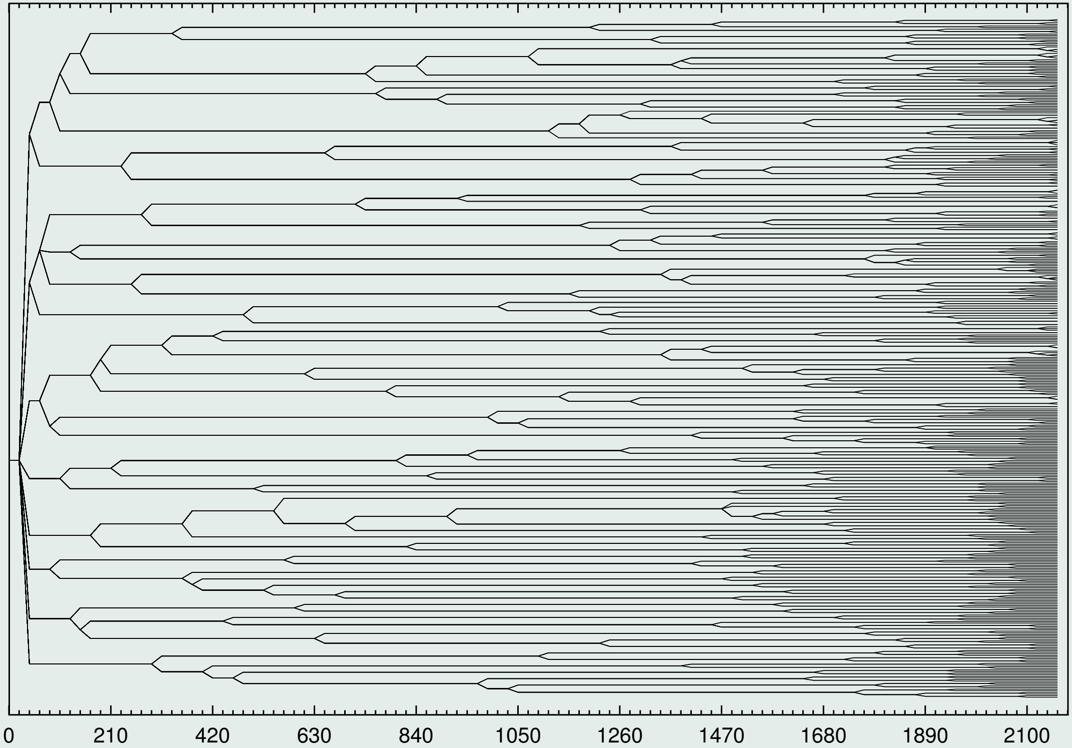


Illustration of the forward tree generation for an example including T=5 time periods starting with a scenario fan containing N=58 scenarios

<Start Animation>



Two-yearly demand-inflow scenario tree with weekly branchings for French EDF

Decomposition of (convex) stochastic programs

Direct or primal decomposition approaches:

- starting point: Benders decomposition based on both *feasibility* and *objective* cuts;
- variants: regularization to avoid an explosion of the number of cuts; nesting when applied to solve the dynamic programming equations on subtrees recursively; stochastic cuts.

Dual decomposition approaches:

(i) Scenario decomposition by Lagrangian relaxation of nonanticipativity constraints (solving the dual by bundle subgradient methods, augmented Lagrangian decomposition, splitting methods);
(ii) nodal decomposition by Lagrangian relaxation of dynamic constraints (same variants as in (i));
(iii) geographical decomposition by Lagrangian relaxation of coupling constraints (same variants as in (i)).

Mostly used for convex models: nested Benders decomposition, stochastic dual dynamic programming, stochastic decomposition and scenario decomposition. (Ruszczyński 03)

Geographical decomposition

In electricity optimization the tree representation of the multistage stochastic program often has block separable structure

$$\min\left\{ \sum_{n\in\mathcal{N}} \pi^n \sum_{i=1}^{k} \langle b^i_{t(n)}(\xi^n), x^n_i \rangle \;\middle|\; \begin{array}{l} x^n_i \in X^i_{t(n)} \\ \sum_{i=1}^k B^i_{t(n)}(\xi^n) x^n_i \geq g_{t(n)}(\xi^n) \\ A^i_{t(n),0} x^n_i + A^i_{t(n),1} x^{n-}_i = h^i_{t(n)}(\xi^n) \\ i = 1,\ldots,k,\, n \in \mathcal{N} \end{array} \right\}$$

Lagrange relaxation of coupling constraints: $L(x,\lambda) =$

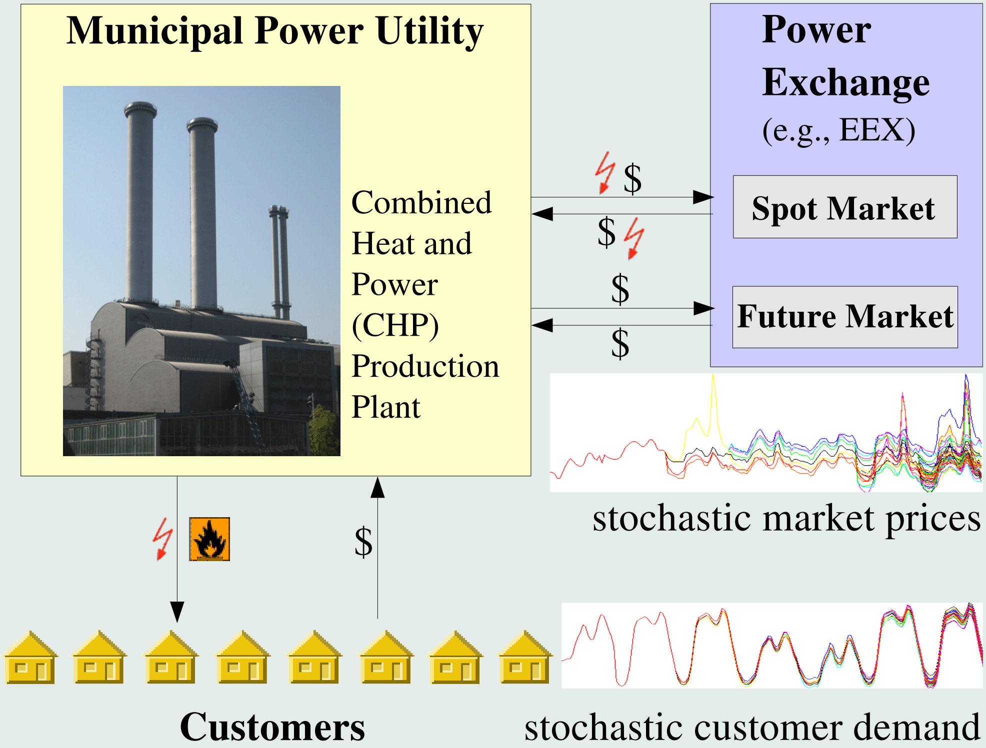
$$\sum_{n\in\mathcal{N}} \pi^n \Big(\sum_{i=1}^{k} \langle b^i_{t(n)}(\xi^n), x^n_i \rangle + \langle \lambda^n, \Big(g_{t(n)}(\xi^n) - \sum_{i=1}^{k} B^i_{t(n)}(\xi^n) x^n_i \Big) \rangle \Big)$$

The dual problem

$$\max_{\lambda \geq 0} \inf_x L(x,\lambda)$$

decomposes into k geograhical subproblems and is solved by bundle subgradient methods. For nonconvex models the duality gap is typically small allowing for Lagrangian heuristics.

Mean-Risk Electricity Portfolio Management



(Eichhorn/Römisch/Wegner 05, Eichhorn/Heitsch/Römisch 10)

We consider the electricity portfolio management of a German municipal electric power company. Its portfolio consists of the following positions:

- power production (based on company-owned thermal units),
- bilateral contracts,
- (physical) (day-ahead) spot market trading (e.g., European Energy Exchange (EEX)) and
- (financial) trading of futures.

The time horizon is discretized into hourly intervals. The underlying stochasticity consists in a multivariate stochastic load and price process that is approximately represented by a finite number of scenarios. The objective is to maximize the total expected revenue and to minimize the risk. The portfolio management model is a large scale (mixed-integer) multi-stage stochastic program.

Electricity portfolio management: statistical models and scenario trees

For the stochastic input data of the optimization model (here yearly electricity and heat demand, and electricity spot prices), a statistical model is employed. It is adapted to historical data in the following way:

- cluster classification for the intra-day (demand and price) profiles,

- 3-dimensional time series model for the daily average values (deterministic trend functions, a trivariate ARMA model for the (stationary) residual time series),

- simulation of an arbitrary number of three dimensional sample paths (scenarios) by sampling the white noise processes for the ARMA model and by adding on the trend functions and matched intra-day profiles from the clusters afterwards,

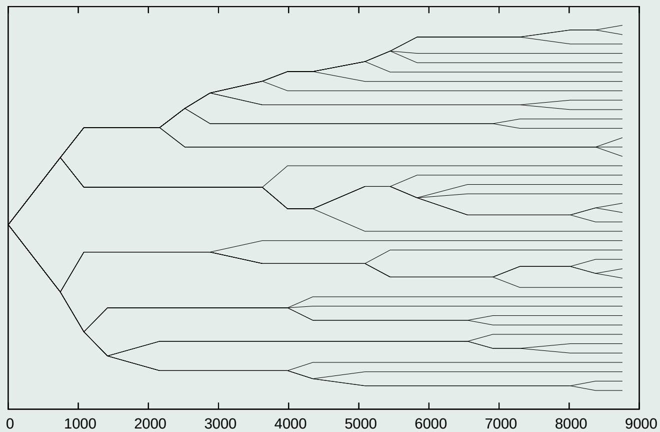
- generation of scenario trees (Heitsch-Römisch 09).

Electricity portfolio management: Results

Test runs were performed on real-life data of a German municipal power company leading to a linear program containing $T = 365 \cdot 24 = 8760$ time steps, a scenario tree with 40 demand-price scenarios (see below) with about 150.000 nodes. The objective function is of the form

$$\text{Minimize} \quad \gamma\rho(Y) - (1-\gamma)\mathbb{E}\left(\sum_{t=1}^{T} Y_t\right)$$

with a (multiperiod) risk functional ρ with risk aversion parameter $\gamma \in [0,1]$ ($\gamma = 0$ corresponds to no risk).



Single-period and multi-period risk functionals are computed for the accumulated income at $t = T$ and at the risk time steps t_j, $j = 1, \dots, J = 52$, respectively. The latter correspond to 11 pm at the last trading day of each week.

It turns out that the numerical results for the expected maximal revenue and minimal risk

$$\mathbb{E}\left(\sum_{t=1}^{T} Y_t^{\gamma *}\right) \qquad \text{and} \qquad \rho(Y_{t_1}^{\gamma *}, \dots, Y_{t_J}^{\gamma *})$$

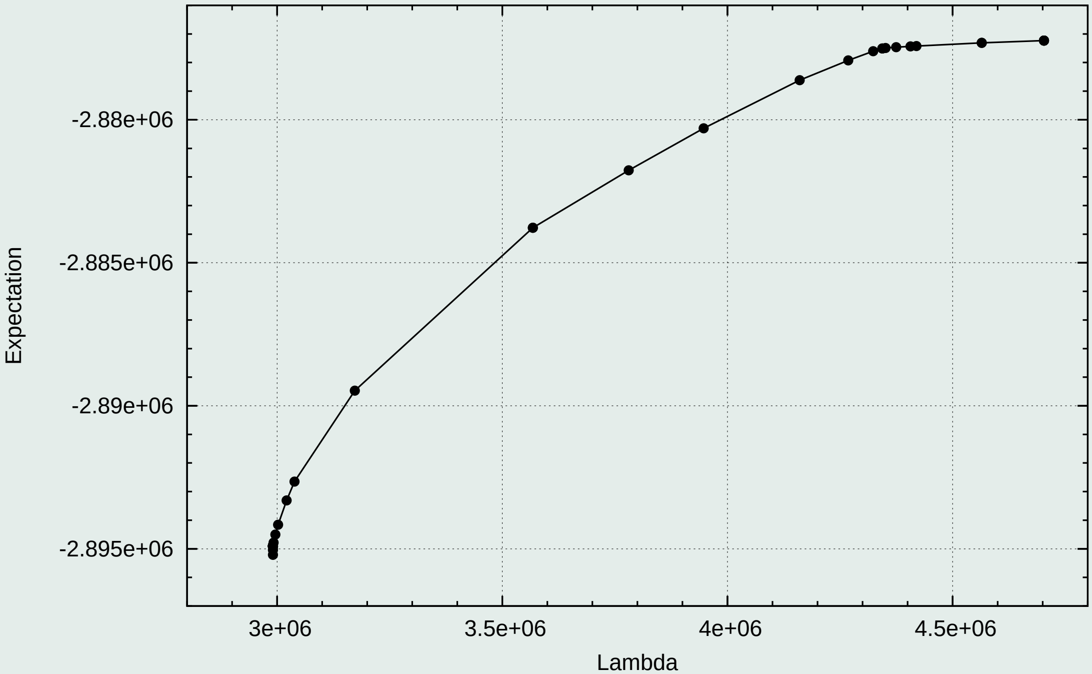
with the optimal income process $Y^{\gamma *}$ are **identical** for $\gamma \in [0.15, 0.95]$ and all risk functionals used in the test runs.

The efficient frontier

$$\gamma \mapsto \left(\rho(Y_{t_1}^{\gamma *}, \dots, Y_{t_J}^{\gamma *}), \mathbb{E}\left(\sum_{t=1}^{T} Y_t^{\gamma *}\right)\right)$$

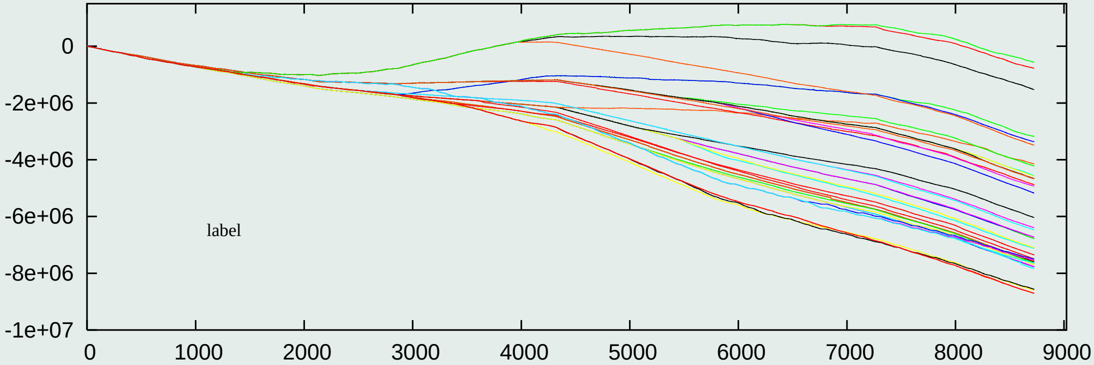
is concave for $\gamma \in [0, 1]$.

Risk aversion costs less than 1% of the expected overall revenue.

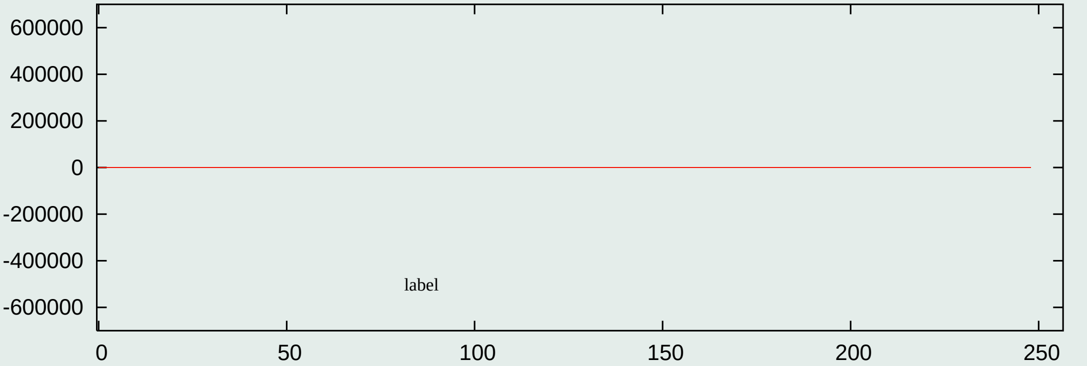


Efficient frontier with expected revenue (ordinate axis) and risk (abscissa)

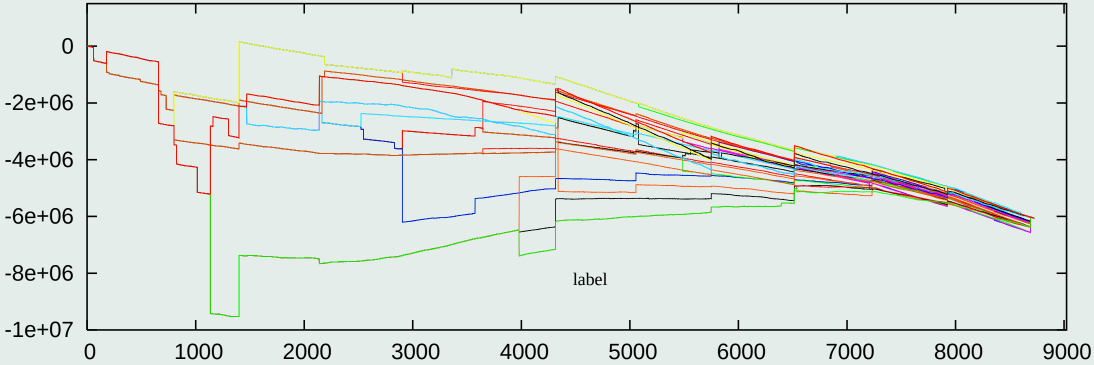
The LP is solved by CPLEX 9.1 in about 1 h running time on a 2 GHz Linux PC with 1 GB RAM.



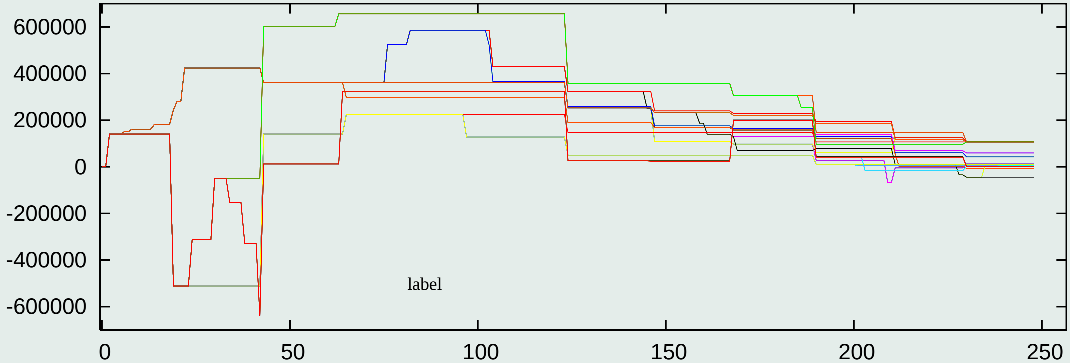
Overall revenue scenarios for $\gamma = 0$



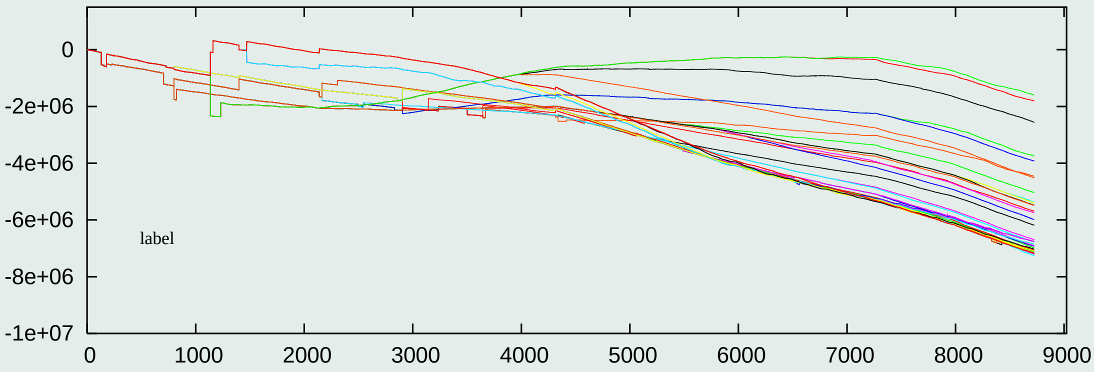
Future trading for $\gamma = 0$



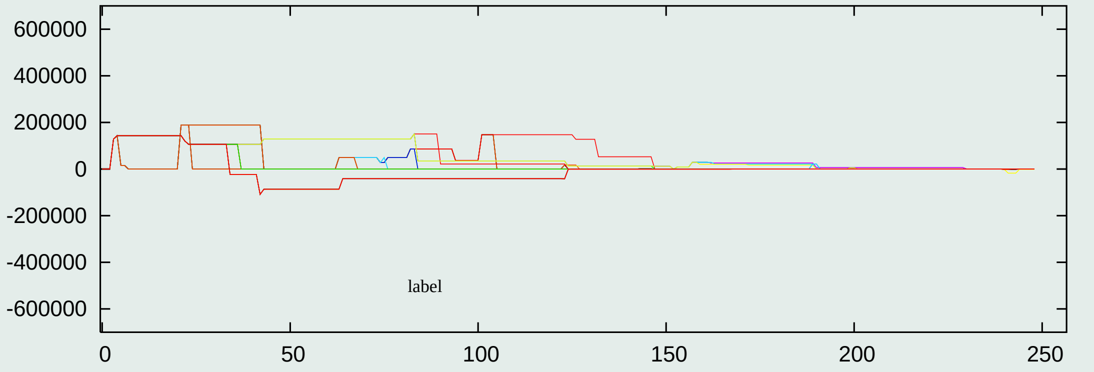
Overall revenue scenarios with $\mathbb{A}$V@R$_{0.05}$ and $\gamma = 0.9$



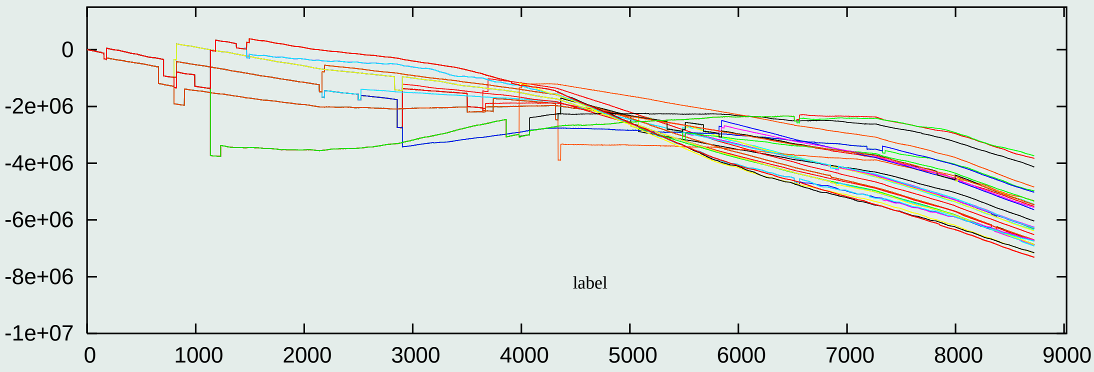
Future trading with $\mathbb{A}$V@R$_{0.05}$ and $\gamma = 0.9$



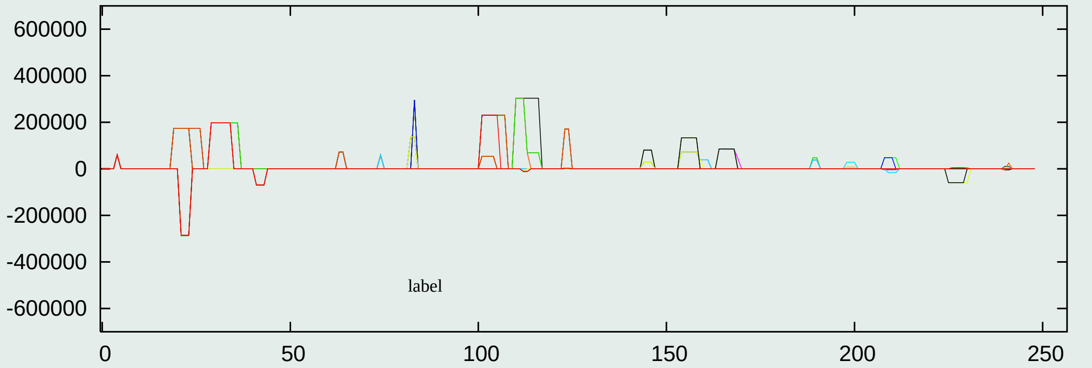
Overall revenue scenarios with ρ_s and $\gamma = 0.9$



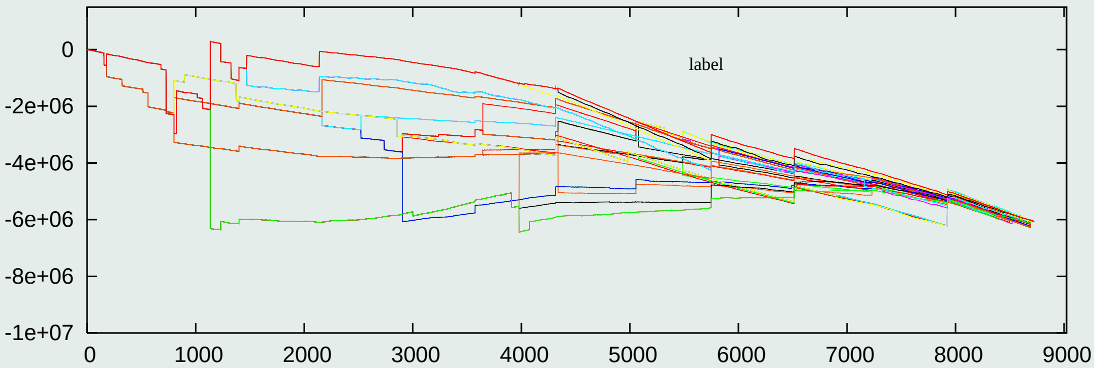
Future trading for ρ_s and $\gamma = 0.9$



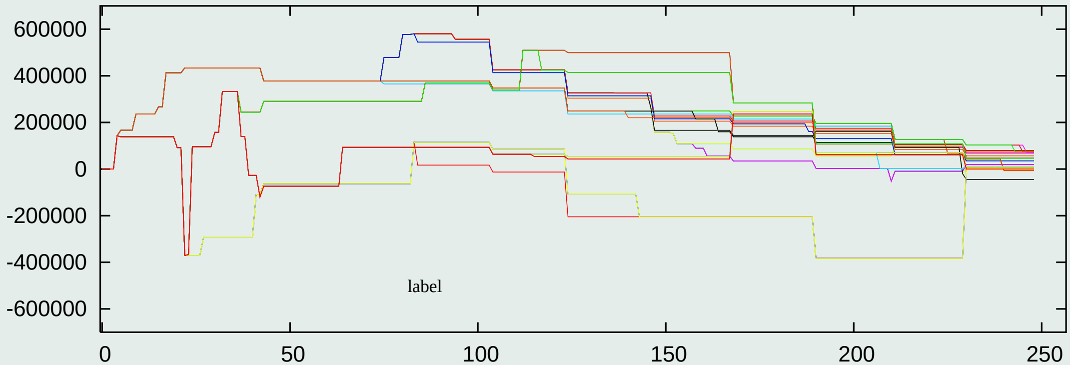
Overall revenue scenarios with ρ_a and $\gamma = 0.9$



Future trading with ρ_a and $\gamma = 0.9$



Overall revenue scenarios with ρ_m and $\gamma = 0.9$



Future trading with ρ_m and $\gamma = 0.9$

Future research directions

Stochastic Programming is still a young mathematical field and started in 1955. It has a high potential for further developments in connection with progress in other fields.

Possible research directions:

- Extending the available theory and numerical methods for chance constrained models.
- Theory and scenario generation for mixed-integer two- and multi-stage stochastic programs.
- Systematic study of methods for high-dimensional numerical integration and their use for **scenario generation**.
- Study of **conditioning** of stochastic programs and improving the understanding of "which stochastic programs are difficult to solve" (e.g. require an extremely high number of scenarios) (first attempt to conditioning by Shapiro/Homem-de-Mello/Kim 02)

- Extending theory, (quantitative) stability analysis and solution methods to more involved models like SMPECs, SGNEP (stochastic generalized Nash equilibrium problems, SEPECs (stochastic equilibrium problems with equilibrium constraints)

  (recent pioneering work by **Xu** and his coworkers)

Thank you for your patience

References:

(Note that papers related to Monte Carlo sampling, Quasi-Monte Carlo methods, optimal quantization methods, sparse grids, scenario reduction and scenario tress are listed extra below.)

Monographs:

I. M. Sobol': *Multidimensional Quadrature Formulas and Haar Functions*, Nauka, Moscow, 1969 (in Russian).

B. Bank, J. Guddat, D. Klatte, B. Kummer, K. Tammer: *Nonlinear parametric optimization*, Akademie-Verlag, Berlin, 1982.

S. T. Rachev: *Probability Metrics and the Stability of Stochastic Models*, Wiley, 1991.
H. Niederreiter: *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia, 1992.

I. H. Sloan, S. Joe: *Lattice Methods for Multiple Integration*, Clarendon Press, Oxford 1994.

A. Prekopa: *Stochastic Programming*, Kluwer, Dordrecht, 1995.

A. W. van der Vaart, J. Wellner: *Weak Convergence and Empirical Processes*, Springer, New York, 1996.

R. T. Rockafellar and R. J-B. Wets: *Variational Analysis*, Springer, Berlin, 1998.

A. W. van der Vaart: *Asymptotic Statistics*, Cambridge University Press, Cambridge 1998.

S. T. Rachev and L. Rüschendorf: *Mass Transportation Problems*, Vol. I, Springer, New York, 1998.

S. Graf, H. Luschgy: *Foundations of Quantization for Probability Distributions*, Lecture Notes in Mathematics, Vol. 1730, Springer, Berlin, 2000.

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

A. Ruszczyński and A. Shapiro (Eds.): *Stochastic Programming*, Handbooks in Operations Research and Management Science, Vol. 10, Elsevier, Amsterdam, 2003.

H. Föllmer, A. Schied: Stochastic Finance: An Introduction in Discrete Time, de Gruyter, Berlin 2004 (Second Edition 2004).

H.-J. Bungartz, M. Griebel: Sparse Grids, *Acta Numerica* (2004), 1–123.

S. W. Wallace and W. T. Ziemba (Eds.): *Applications of Stochastic Programming*, MPS-SIAM Series on Optimization, Philadelphia, 2005.

G. Ch. Pflug, W. Römisch: Modeling, Measuring and Managing Risk, World Scientific, Singapore, 2007.

C. Lemieux: *Monte Carlo and Quasi-Monte Carlo Sampling*, Springer, 2009.

A. Genz, F. Bretz: *Computation of multivariate normal and t probabilities*, Lecture Notes in Statistics Vol. 195, Springer, Berlin, 2009.

A. Shapiro, D. Dentcheva and A. Ruszczyński: *Lectures on Stochastic Programming*, MPS-SIAM Series on Optimization, Philadelphia, 2009.

J. Dick and F. Pillichshammer: *Digital Nets and Sequences*, Cambridge University Press, 2010.

Papers:

W. van Ackooj, R. Henrion, A. Möller, R. Zorgati: On joint probabilistic constraints with Gaussian coefficient matrix, *Operations Research Letters* 39 (2011), 99–102.

P. Artzner, F. Delbaen, J.-M. Eber, D. Heath: Coherent measures of risk, *Mathematical Finance* 9 (1999), 203–228.

C. F. Blair, R. G. Jeroslow: The value function of a mixed integer program I, *Discrete Mathematics* 19 (1977), 121–138.

D. Dentcheva, W. Römisch: Stability and sensitivity of stochastic dominance constrained optimization models, submitted.

D. Dentcheva, A. Ruszczyński: Stochastic optimization with dominance constraints, *SIAM Journal on Optimization* 14 (2003), 548–566.

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

A. Eichhorn, W. Römisch: Polyhedral risk measures in stochastic programming, *SIAM Journal on Optimization* 16 (2005), 69–95.

A. Eichhorn, W. Römisch: Stochastic integer programming: Limit theorems and confidence intervals, *Mathematics of Operations Research* 32 (2007), 118–135.

A. Eichhorn, W. Römisch, I. Wegner: Mean-risk optimization of electricity portfolios using multiperiod polyhedral risk measures, *IEEE St. Petersburg Power Tech* 2005.

A. Eichhorn, H. Heitsch and W. Römisch: Stochastic optimization of electricity portfolios: Scenario tree modeling and risk management, in Handbook of Power Systems, Vol. II (S. Rebennack, P.M. Pardalos, M.V.F. Pereira,, N.A. Iliadis), Springer, 2010, 405-432.

H. Heitsch, W. Römisch, C. Strugarek: Stability of multistage stochastic programs, *SIAM Journal Optimization* 17 (2006), 511–525.

R. Henrion, C. Strugarek, C.: Convexity of chance constraints with independent random variables, *Computational Optimization and Applications* 41 (2008), 263–276.

R. Hochreiter, G. Ch. Pflug: Financial scenario generation for stochastic multi-stage decision processes as facility location problem, *Annals of Operations Research* 152 (2007), 257–272.

D. Klatte: On quantitative stability for non-isolated minima, *Control and Cybernetics* 23 (1994), 183–200.

Y. Liu, H. Xu and G. Lin: Stability analysis of one stage stochastic mathematical programs with complementarity constraints, *Journal of Optimization Theory and Applications* 152 (2012), 573–555.

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

D. Ralph and H. Xu: Convergence of stationary points of sample average two stage stochastic programs: A generalized equation approach, *Mathematics of Operations Research* 36 (2011), 568–592.

R. T. Rockafellar, S. Uryasev: Conditional value-at-risk for general loss distributions, *Journal of Banking & Finance* 26 (2002), 1443–1471.

W. Römisch: Stability of stochastic programming problems, in: *Stochastic Programming* (A. Ruszczyński, A. Shapiro eds.), Handbooks in Operations Research and Management Science, Volume 10, Elsevier, Amsterdam 2003, 483–554.

W. Römisch and S. Vigerske: Quantitative stability of fully random mixed-integer two-stage stochastic programs, Optimization Letters 2 (2008), 377–388.

A. Ruszczyński: Decomposition methods, in *Stochastic Programming* (Ruszczyński, A., Shapiro, A., eds.), Elsevier, Amsterdam, 2003, 141–211.

R. Schultz: Strong convexity in stochastic programs with complete recourse, *Journal of Computational and applied Mathematics* 56 (1994), 3–22.

R. Schultz: On structure and stability in stochastic programs with random technology matrix and complete integer recourse, *Mathematical Programming* 70 (1995), 73–89.

R. Schultz: Stochastic programming with integer variables, *Mathematical Programming* 97 (2003), 285–309.

A. Shapiro: Inference of statistical bounds for multistage stochastic programming problems, *Mathematical Methods of Operations Research* 58 (2003), 57–68.

A. Shapiro: Stochastic programming with equilibrium constraints, *Journal of Optimization Theory and Applications* 128 (2006), 223–243.

A. Shapiro, T. Homem-de-Mello, J. Kim: Conditioning of convex piecewise linear stochastic programs, *Mathematical Progrogramming* 94 (2002), 1–19.

M. Talagrand, Shaper bounds for Gaussian and empirical processes, *Annals of Probability* 22 (1994), 28–76.

D. Walkup, R. J-B Wets: Lifting projections of convex polyhedra, *Pacific Journal of Mathematics* 28 (1969), 465–475.

R. J-B Wets: Stochastic programs with fixed recourse: The equivalent deterministic program, *SIAM Review* 16 (1974), 309–339.

V. M. Zolotarev: Probability metrics, *Theory of Probability and its Applications* 28 (1983), 278–302.

Monte Carlo:

P. L'Ecuyer: Uniform random number generation, *Annals of Operations Research* 53 (1994), 77–120.

M. Matsumoto, T. Nishimura: Mersenne Twister: A 623-dimensionally equidistributed uniform pseudo-random number generator, *ACM Transactions on Modeling and Computer Simulation* 8 (1998), 3–30.

A. Shapiro: Monte Carlo sampling methods, in: *Stochastic Programming* (A. Ruszczyński, A. Shapiro eds.), Handbooks in Operations Research and Management Science, Volume 10, Elsevier, Amsterdam 2003, 353–425.

Quantization:

G. Pagès: A space quantization method for numerical integration, *Journal Computational and Applied Mathematics* 89 (1997), 1–38.

G. Ch. Pflug and A. Pichler: Approximations for probability distributions and stochastic optimization problems, in: *Stochastic Optimization Methods in Finance and Energy* (M.I. Bertocchi, G. Consigli, M.A.H. Dempster eds.), Springer, 2011, 343–387.

Quasi-Monte Carlo:

T. Pennanen, M. Koivu: Epi-convergent discretizations of stochastic programs via integration quadratures, *Numerische Mathematik* 100 (2005), 141–163.

T. Homem-de-Mello: On rates of convergence for stochastic optimization problems under non-i.i.d. sampling, *SIAM Journal on Optimization* 19 (2008), 524-551.

I. H. Sloan, H. Woźniakowski: When are Quasi Monte Carlo algorithms efficient for high-dimensional integration, *Journal of Complexity* 14 (1998), 1–33.

F. Y. Kuo, I. H. Sloan: Lifting the curse of dimensionality, *Notices of the AMS* 52 (2005), 1320–1328.

X. Wang, K.-T. Fang: The effective dimension and Quasi-Monte Carlo integration, *Journal of Complexity* 19 (2003), 101–124.

F. Y. Kuo, I. H. Sloan, G. W. Wasilkowski, H. Woźniakowski: On decomposition of multivariate functions, *Mathematics of Computation* 79 (2010), 953–966.

M. Griebel, F. Y. Kuo and I. H. Sloan: The smoothing effect of the ANOVA decomposition, *Journal of Complexity* 26 (2010), 523–551.

M. Griebel, F. Y. Kuo and I. H. Sloan: The smoothing effect of integration in $\mathbb{R}^d$ and the ANOVA decomposition, *Mathematics of Computation* (to appear).

F. J. Hickernell: A generalized discrepancy and quadrature error bound, *Mathematics of Computation* 67 (1998), 299-322.

P. D. Proinov: Discrepancy and integration of continuous functions, *Journal Approximation Theory* 52 (1998), 121–131.

I. H. Sloan, F. Y. Kuo, S. Joe: On the step-by-step construction of Quasi-Monte Carlo integration rules that achieve strong tractability error bounds in weighted Sobolev spaces, *Mathematics of Com-*

putation 71 (2002), 1609-1640.

F. Y. Kuo: Component-by-component constructions achieve the optimal rate of convergence in weighted Korobov and Sobolev spaces, *Journal of Complexity* 19 (2003), 301-320.

H. Heitsch, H. Leövey and W. Römisch, Are Quasi-Monte Carlo algorithms efficient for two-stage stochastic programs?, *Stochastic Programming E-Prints* 5 (2012) (www.speps.org).

Sparse Grids:

S. A. Smolyak: Quadrature and interpolation formulas for tensor products of certain classes of functions, *Soviet Math. Dokl.* 4 (1963), 240–243.

T. Gerstner, M. Griebel: Numerical integration using sparse grids, *Numerical Algorithms* 18 (1998), 209–232.

M. Chen, S. Mehrotra: Epi-convergent scenario generation method for stochastic problems via sparse grid, *SPEPS* 7-2008.

Scenario reduction:

P. Avella, A. Sassano, I. Vasil'ev: Computational study of large-scale p-median problems, *Mathematical Programming* 109 (2007), 89–117.

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

W. Römisch and R. J-B Wets: Stability of epsilon-approximate solutions to convex stochastic programs, *SIAM Journal on Optimization* 18 (2007), 961–979.

H. Heitsch, W. Römisch: A note on scenario reduction for two-stage stochastic programs, *Operations Research Letters* 35 (2007), 731–738.

R. Henrion, C. Küchler, W. Römisch: Discrepancy distances and scenario reduction in two-stage stochastic mixed-integer programming, *Journal of Industrial and Management Optimization* 4 (2008), 363–384.

R. Henrion, C. Küchler, W. Römisch: Scenario reduction in stochastic programming with respect to discrepancy distances, *Computational Optimization and Applications* 43 (2009), 67–93.

Scenario trees:

J. Dupačová, G. Consigli, S. W. Wallace: Scenarios for multistage stochastic programs, *Annals of Operations Research* 100 (2000), 25–53.

H. Heitsch, W. Römisch: Scenario tree modeling for multistage stochastic programs, *Mathematical Programming* 118 (2009), 371–406.

H. Heitsch, W. Römisch: Scenario tree reduction for multistage stochastic programs, *Computational Management Science* 6 (2009), 117–133.

R. Hochreiter, G. Ch. Pflug: Financial scenario generation for stochastic multi-stage decision processes as facility location problem, *Annals of Operations Research* 152 (2007), 257–272.

K. Høyland, M. Kaut, S. W. Wallace: A heuristic for moment-matching scenario generation, *Computational Optimization and Applications* 24 (2003), 169–185.

T. Pennanen: Epi-convergent discretizations of multistage stochastic programs via integration quadratures, *Mathematical Programming* 116 (2009), 461–479.

R. Mirkov, G. Ch. Pflug: Tree approximations of dynamic stochastic programs, *SIAM Journal on Optimization* 18 (2007), 1082–1105.

G. Ch. Pflug: Scenario tree generation for multiperiod financial optimization by optimal discretization, *Mathematical Programming* 89 (2001), 251–271.

G. Ch. Pflug: Version independence and nested distributions in multistage stochastic optimization, *SIAM Journal on Optimization* 20 (2009), 1406–1420.

G. Ch. Pflug, A. Pichler: A distance for multistage stochastic optimization models, *SIAM Journal on Optimization* 22 (2012), 1–23.