Stability-based scenario tree modelling

H. Heitsch and W. Römisch

Humboldt-University Berlin
Institute of Mathematics
10099 Berlin, Germany

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

Mini-Workshop *Stochastic Programming*, Humboldt-University Berlin, March 14, 2007

EDF Electricité de France

DFG Research Center MATHEON
Mathematics for key technologies

Bundesministerium für Bildung und Forschung

Multistage stochastic programs

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

Multistage stochastic program:

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

where X_t are nonempty and polyhedral sets, $A_{t,0}$ are fixed recourse matrices and $b_t(\cdot)$, $h_t(\cdot)$ and $A_{t,1}(\cdot)$ are affine functions depending on ξ_t, where ξ varies in a polyhedral set Ξ.

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

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

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

Then nonanticipativity may be expressed as

$$x \in \mathcal{N}_{r'}(\xi)$$

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

i.e., as a subspace constraint, by using the conditional expectations $\mathbb{E}[\cdot | \mathcal{F}_t(\xi)]$.

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

$\longrightarrow$ infinite-dimensional optimization problem

Quantitative Stability

Let us introduce some notations. 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 $F(\xi, x) := \mathbb{E}[\sum_{t=1}^{T} \langle b_t(\xi_t), x_t \rangle]$, 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

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

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

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

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

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

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

The following conditions are imposed:

(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 $t = 2, \dots, T$ and any $x_1 \in X_1$, $x_\tau \in \mathcal{X}_\tau(x_{\tau-1}; \tilde{\xi}_\tau)$, $\tau = 2, \dots, t-1$, the set $\mathcal{X}_t(x_{t-1}; \tilde{\xi}_t)$ is nonempty (relatively complete recourse locally around ξ).

(A3) The optimal values $v(\tilde{\xi})$ are finite for all $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ with $\|\tilde{\xi} - \xi\|_r \leq \delta$ and 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 $L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ such that $S_\alpha(\tilde{\xi})$ is nonempty and contained in B for all $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ with $\|\tilde{\xi} - \xi\|_r \leq \delta$.

Norm in L_r: $\|\xi\|_r := \left(\sum_{t=1}^{T} \mathbb{E}[\|\xi_t\|^r] \right)^{\frac{1}{r}}$

Theorem:
Let (A1), (A2) and (A3) be satisfied, $r \geq 1$ and X_1 be bounded. Then there exist positive constants L, $\bar{\varepsilon}$ and δ such that the estimates

$$\begin{aligned} |v(\xi) - v(\tilde{\xi})| &\leq L(\|\xi - \tilde{\xi}\|_r + D_{\mathrm{f}}^*(\xi, \tilde{\xi})) \\ d\!l_\infty(S_\varepsilon(\xi), S_\varepsilon(\tilde{\xi})) &\leq \frac{\bar{L}}{\varepsilon}(\|\xi - \tilde{\xi}\|_r + D_{\mathrm{f}}^*(\xi, \tilde{\xi})) \end{aligned}$$

hold for any $\varepsilon \in (0, \bar{\varepsilon})$ and for all $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ with $\|\tilde{\xi} - \xi\|_r \leq \delta$, where for the latter estimate it is required that $S(\xi)$ and $S(\tilde{\xi})$ are nonempty and $r' \in [1, \infty)$.

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

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

and $d\!l_\infty$ the Pompeiu-Hausdorff distance of closed subsets of $L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$.

Remark 1: (strengthening (A3))
If in (A3) it is even required that the set B is bounded in L_∞, the $L_{r'}$-unit ball in the definition of D_{f}^* may be replaced by the L_∞-unit ball. The latter distance is (much) smaller and easier to estimate !

Remark 2: (weak convergence of solutions)
If (A1), (A2) and (A3) are satisfied, X_1 is bounded, $r' \in (1, \infty)$ and $(\xi^{(n)})$ is a sequence in $L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ converging to ξ in L_r and with respect to D_{f}^*, any sequence of solutions $(x^{(n)})$ with $x^{(n)} \in S(\xi^{(n)})$ contains a subsequence that converges weakly in $L_{r'}$ to some element of $S(\xi)$. Hence, the convergence result for approximate solution sets is much stronger than for solution sets.

Remark 3: (two-stage case)
For the two-stage situation $T = 2$, the stability result extends earlier work, since it states stability results on approximate second-stage solution sets. In earlier work, the stability of (approximate) first-stage solution sets was considered under an inf-boundedness condition for first-stage decisions (weaker than (A3)).

Generation of scenario trees

(i) In most practical situations scenarios ξ^i with known probabilities p_i, $i = 1, \ldots, N$, can be generated, e.g., simulation scenarios from (parametric or nonparametric) statistical models of ξ or (nearly) optimal quantizations of the probability distribution of ξ.

(ii) Generating a scenario tree out of the scenarios ξ^i with probabilities p_i, $i = 1, \ldots, N$,.

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, Mirkov/Pflug 06; Gröwe-Kuska/Heitsch/Römisch 01, 03, Heitsch/Römisch 05).

Survey: Dupačová/Consigli/Wallace 00

Generating scenario trees

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

$$\|\xi - \xi_{\text{tr}}\|_r \quad \text{and} \quad D_{\text{f}}^*(\xi, \xi_{\text{tr}})$$

are small and, hence, the optimal values $v(\xi)$ and $v(\xi_{\text{tr}})$, and the approximate solution sets $S_\varepsilon(\xi)$ and $S_\varepsilon(\xi_{\text{tr}})$ are close to each other according to the stability result.

The idea is to start with a good initial approximation $\hat{\xi}$ of ξ having a finite number of scenarios $\xi^i = (\xi_1^i, \ldots, \xi_T^i) \in \mathbb{R}^{Td}$ with probabilities $p_i > 0$, $i = 1, \ldots, N$, and common root, i.e., $\xi_1^1 = \ldots = \xi_1^N =: \xi_1^*$. These scenarios might be obtained by quantization techniques or by sampling or resampling techniques based on parametric or nonparametric stochastic models of ξ.

In the following we assume that

$$\|\xi - \hat{\xi}\|_r \; + \; D_{\mathrm{f}}^*(\xi, \hat{\xi}) \; \leq \; \varepsilon$$

holds for some given (initial) tolerance $\varepsilon > 0$.

For example, the condition may be satisfied for D_{f}^* with respect to the unit ball in L_∞ instead of $L_{r'}$ for any tolerance $\varepsilon > 0$ if $\hat{\xi}$ is obtained by sampling from a finite set with sufficently large sample size.

Forward tree generation

Let scenarios ξ^i with probabilities p_i, $i = 1, \ldots, N$, fixed root $\xi_1^* \in \mathbb{R}^d$, $r \geq 1$, and tolerances ε_{r}, ε_t, $t = 2, \ldots, T$, be given such that $\sum_{t=2}^{T} \varepsilon_t \leq \varepsilon_{\mathrm{r}}$.

Step 1: Set $\hat{\xi}^1 := \hat{\xi}$ and $\mathcal{C}_1 = \{I = \{1, \ldots, N\}\}$.

Step t: Let $\mathcal{C}_{t-1} = \{C_{t-1}^1, \ldots, C_{t-1}^{K_{t-1}}\}$. Determine disjoint index sets I_t^k and J_t^k of remaining and deleted scenarios such that $I_t^k \cup J_t^k = C_{t-1}^k$, a mapping $\alpha_t : I \to I$

$$\alpha_t(j) = \begin{cases} i_t^k(j) & , j \in J_t^k,\ k = 1, \ldots, K_{t-1}, \\ j & , \text{otherwise}, \end{cases}$$

where $i_t^k(j) \in I_t^k$ such that

$$i_t^k(j) \in \arg\min_{i \in I_t^k} |\hat{\xi}^{t-1,i} - \hat{\xi}^{t-1,j}|_t,$$

a stochastic process $\hat{\xi}^t$

$$\hat{\xi}^{t,i}_\tau = \begin{cases} \xi_\tau^{\alpha_\tau(i)} & , \tau \le t, \\ \xi_\tau^i & , \text{otherwise}, \end{cases}$$

such that

$$\|\hat{\xi}^t - \hat{\xi}^{t-1}\|_{r,t}^r = \sum_{k=1}^{K_{t-1}} \sum_{j \in J_t^k} p_j \min_{i \in I_t^k} |\xi_t^i - \xi_t^j|^r \le \varepsilon_t^r.$$

Set $I_t := \cup_{k=1}^{K_{t-1}} I_t^k$ and $\mathcal{C}_t := \{\alpha_t^{-1}(i) : i \in I_t^k,\ k = 1, \ldots, K_{t-1}\}$.

Step T+1: Let $\mathcal{C}_T = \{C_T^1, \ldots, C_T^{K_T}\}$. Construct a stochastic process ξ_{tr} having K_T scenarios ξ_{tr}^k such that $\xi_{\mathrm{tr},t}^k := \xi_t^{\alpha_t(i)}$ with probabilities $\pi_T^i = \sum_{j \in C_T^k} p_j$ if $i \in C_T^k$, $k = 1, \ldots, K_T$, $t = 2, \ldots, T$.

Theorem: $\|\hat{\xi} - \xi_{\mathrm{tr}}\|_r \le \sum_{t=2}^{T} \varepsilon_t \le \varepsilon_{\mathrm{r}}$.

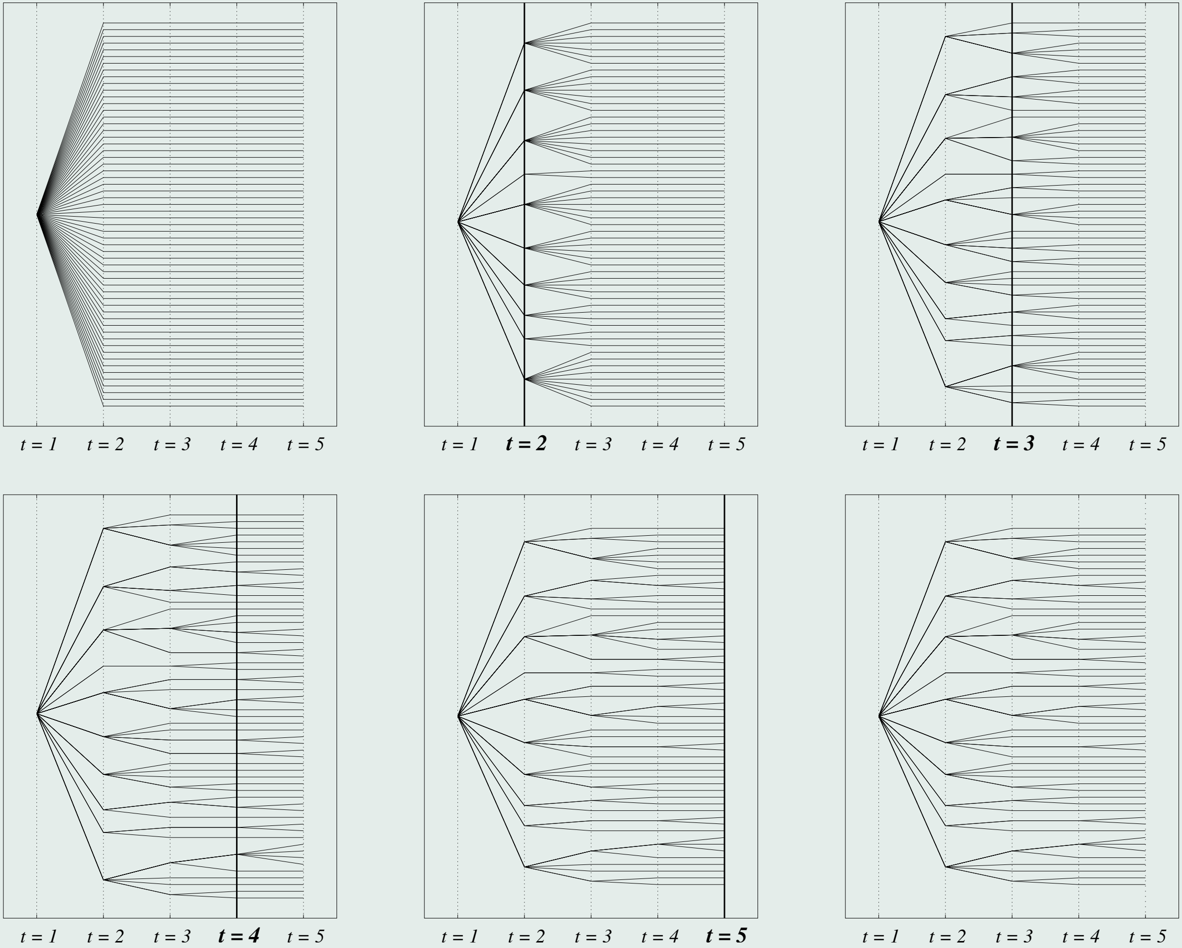


Illustration of the forward tree generation for an example including T=5 time periods starting with a scenario fan containing N=58 scenarios

<Start Animation>

Convergence

Theorem:
Let (A1), (A2) and (A3) be satisfied with $r' \in [1, \infty)$ and $\mathcal{X}_1$ be bounded. Let $L > 0$ and $\delta > 0$ be the constants appearing in the stability result and let $\|\xi - \hat{\xi}\|_r < \delta$.
If $(\varepsilon_r^{(n)})$ is a sequence tending to 0 such that the corresponding tolerances $\varepsilon_t^{(n)}$ in the forward tree generation algorithm are nonincreasing for all $t = 2, \ldots, T$, the corresponding sequence $(\xi_{\mathrm{tr}}^{(n)})$ has the property

$$\limsup_{n \to \infty} |v(\xi) - v(\xi_{\mathrm{tr}}^{(n)})| \leq L\varepsilon,$$

where $\varepsilon > 0$ is an initial tolerance such that

$$\|\xi - \hat{\xi}\|_r + D_{\mathrm{f}}^*(\xi, \hat{\xi}) \leq \varepsilon$$

Numerical experience

We consider a yearly mean-risk optimization model for electricity portfolios of a German municipal electricity company which consist of the own (thermal) electricity production, the spot market contracts, supply contracts and electricity futures. Stochasticity enters the model via the electricity demand, heat demand and spot prices (at EEX). Our approach for generating input scenarios in form of a scenario tree consists in developing a (multivariate) statistical model for all stochastic components and in using the forward tree generation algorithm started with a finite number of yearly demand-price scenarios ξ^i with probabilities $p_i = \frac{1}{N}$, $i = 1, \ldots, N$, which are simulated from the statistical model.

The statistical model corresponds to ξ and the finite number of scenarios are assumed to form the process $\hat{\xi}$. In our test series we started with $N = 100$ sample scenarios for a one year time horizon with hourly discretization. Due to the fact that electricity future products can only be traded monthly, branching was allowed only at the end of each month. Scenario trees were generated by the

Components	Horizon	Scenarios	Time steps	Nodes
3 (trivariate)	1 year	100	8 760	875 901

Table 1: Dimension of simulated input scenarios

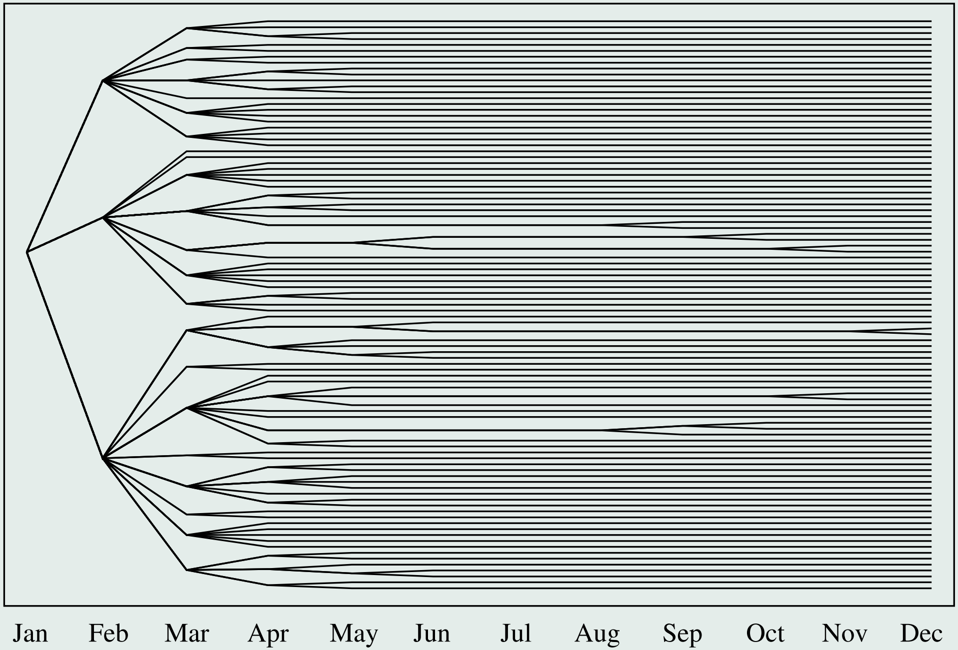
forward tree generation algorithm for $r = r' = 2$ and different relative reduction levels $\varepsilon_{\mathrm{r,rel}}$. The relative levels are given by

$$\varepsilon_{\mathrm{r,rel}} := \frac{\varepsilon}{\varepsilon_{\max}} \qquad \text{and} \qquad \varepsilon_{\mathrm{rel},t} := \frac{\varepsilon_t}{\varepsilon_{\max}},$$

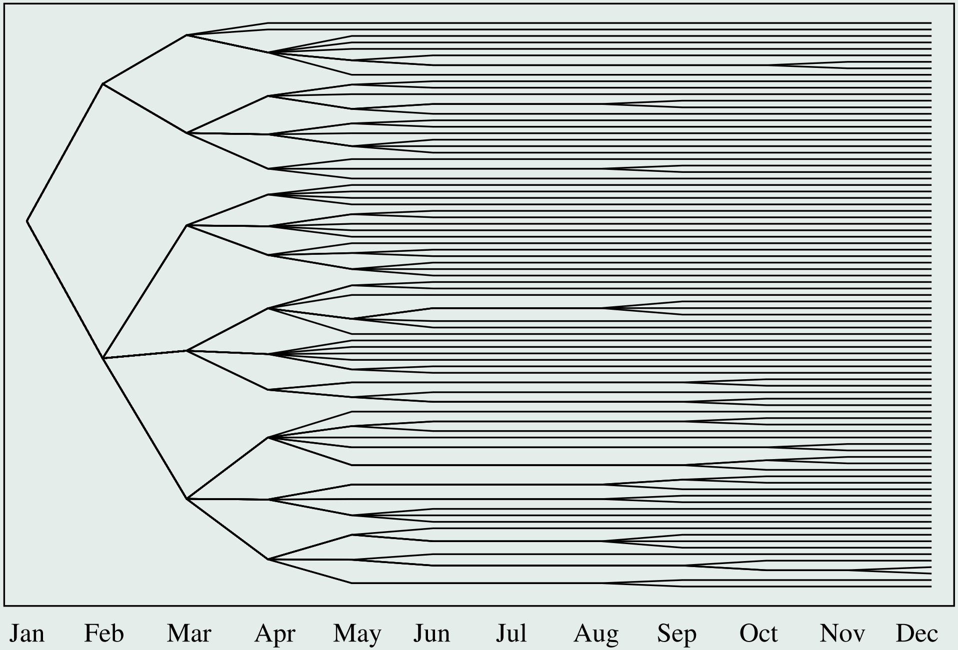
where $\varepsilon_{\max}$ is given as the maximum of the best possible L_r-distance of $\hat{\xi}$ and of one of its scenarios endowed with unit mass. The individual tolerances ε_t at branching points were chosen such that

$$\varepsilon_t^r = \frac{\varepsilon^r}{T}\left[1 + \overline{q}\left(\frac{1}{2} - \frac{t}{T}\right)\right], \quad t = 2, \ldots, T, \quad r = 2,$$

where $\overline{q} \in [0, 1]$ is a parameter that affects the branching structure of the constructed trees. For the test runs we used $\overline{q} = 0.2$ which results in a slightly decreasing sequence ε_t. All test runs were performed on a PC with a 3 GHz Intel Pentium CPU and 1 GByte main memory.



Yearly demand-price scenario tree with reduction level $\varepsilon_{\mathrm{r,rel}} = 0.4$



Yearly demand-price scenario tree with reduction level $\varepsilon_{\mathrm{r,rel}} = 0.55$

$\varepsilon_{r,rel}$	Scenarios		Nodes		Stages	Time (sec)
	initial	tree	initial	tree		
0.20	100	100	875 901	775 992	4	24.53 s
0.25	100	100	875 901	752 136	5	24.54 s
0.30	100	100	875 901	719 472	7	24.55 s
0.35	100	97	875 901	676 416	8	24.61 s
0.40	100	98	875 901	645 672	10	24.64 s
0.45	100	96	875 901	598 704	10	24.75 s
0.50	100	95	875 901	565 800	9	24.74 s
0.55	100	88	875 901	452 184	10	24.75 s
0.60	100	87	875 901	337 728	11	25.89 s

Table 2: Numerical results for yearly demand-price scenario trees

Future research

- Extension of scenario reduction algorithms to mixed-integer (linear) two-stage stochastic programs using discrepancy distances,
- Extension of the stability result for optimal values to mixed-integer (linear) multistage stochastic programs (in terms of (extended) discrepancy distances),
- Development of stability-based forward tree generation algorithms for mixed-integer (linear) multistage stochastic programs.