Stability-based generation of scenario trees for multistage stochastic programs

H. Heitsch and W. Römisch

Humboldt-University Berlin
Institute of Mathematics
10099 Berlin, Germany

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

ISMP 2006, Rio de Janeiro, July 30–August 5, 2006

EDF Electricité de France

DFG Research Center MATHEON
Mathematics for key technologies

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 $x_t \in L_{r'}(\Omega, \mathcal{F}, I\!P; I\!R^{m_t})$ and $\xi_t \in L_r(\Omega, \mathcal{F}, I\!P; I\!R^d)$, where $r \geq 1$ and

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

Then nonanticipativity may be expressed as

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

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

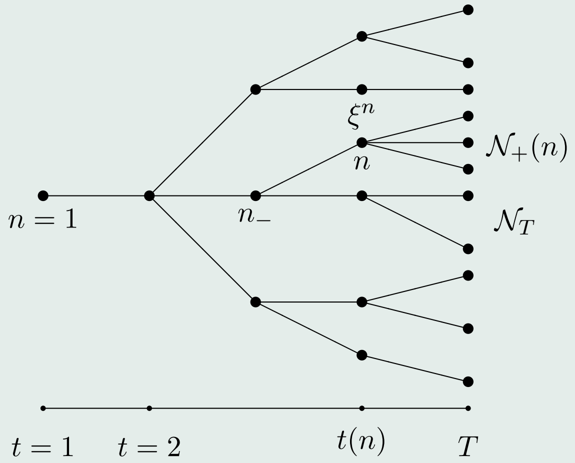
i.e., as a subspace constraint, by using the conditional expectations $I\!E[\cdot|\mathcal{F}_t(\xi)]$.

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

$\longrightarrow$ infinite-dimensional optimization problem

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, $\mathrm{path}(n) = \{1, \ldots, n_-, n\}$, $t(n) := |\mathrm{path}(n)|$, $\mathcal{N}_+(n)$ set of successors to n, $\mathcal{N}_T := \{n \in \mathcal{N} : \mathcal{N}_+(n) = \emptyset\}$ set of leaves, $\mathrm{path}(n)$, $n \in \mathcal{N}_T$, 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)

Questions:

- Under which conditions and in which sense do multistage models behave stable with respect to perturbations of ξ ?
- Can such stability results be used to generate (multivariate) scenario trees ?

Dynamic programming

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

$$\min\left\{\int_\Xi f(x_1,\xi)P(d\xi) : x_1 \in X_1\right\},$$

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

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

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

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

$\longrightarrow$The integrand f depends on the probability measure $I\!P$ and, thus, also on the probability distribution $P = I\!P \circ \xi^{-1}$ of ξ in a nonlinear way ! Hence, earlier approaches to stability fail !

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

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

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 $\alpha \geq 0$,

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

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

The following conditions are imposed:

(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) The optimal values $v(\tilde{\xi})$ are finite for all $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, I\!P; I\!R^s)$ 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 $L_{r'}(\Omega, \mathcal{F}, I\!P; I\!R^m)$ 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 $\|\tilde{\xi} - \xi\|_r \leq \delta$.

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

Theorem: (Heitsch/Römisch/Strugarek, SIAM J. Opt. 2006)
Let (A1), (A2) and (A3) be satisfied, $r > 1$ and X_1 be bounded. Then there exist positive constants L and δ such that

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

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

Assume that technology matrices are non-random and the solution x^* of the original problem is unique.
If $(\xi^{(n)})$ is a sequence in $\times_{t=1}^{T} L_r(\Omega, \mathcal{F}_t(\xi), I\!P; I\!R^s)$ such that

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

converge to 0 and if $(x^{(n)})$ is a sequence of solutions of the approximate problems, then the sequence $(x^{(n)})$ converges to x^* with respect to the weak topology in $L_{r'}$.

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

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

Remark:

The continuity property of infima in the Theorem 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:

Simple examples show that the filtration distance is indispensable for the stability result to hold.

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) Construction of 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

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

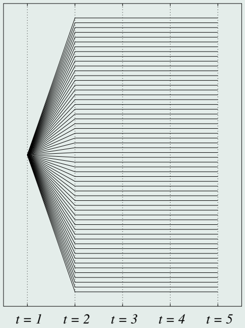
$$\|\xi - \xi_{\mathrm{tr}}\|_r \qquad \text{and} \qquad D_{\mathrm{f}}(\xi, \xi_{\mathrm{tr}})$$

and, thus,

$$|v(\xi) - v(\xi_{\mathrm{tr}})|$$

are small.

To determine such a scenario tree, we start with a discrete approximation ξ_{f} consisting of scenarios $\xi^i = (\xi^i_1, \ldots, \xi^i_T)$ with probabilities p_i, $i = 1, \ldots, N$. ξ_{f} is a fan of individual scenarios.



The fan ξ_{f} is chosen such that it is adapted to the filtration $(\mathcal{F}_t(\xi))_{t=1}^T$ and

$$\|\xi - \xi_{\mathrm{f}}\|_r \leq \varepsilon_{\mathrm{appr}}.$$

Algorithms are developed that generate a scenario tree ξ_{tr} by deleting and bundling scenarios of ξ_{f} (that are similar at t) such that it is also adapted to the filtration $(\mathcal{F}_t(\xi))_{t=1}^T$ and satisfies

$$\|\xi_{\mathrm{f}} - \xi_{\mathrm{tr}}\|_r \leq \varepsilon_{\mathrm{r}} \tag{1}$$

$$\inf_{x \in S(\xi_{\mathrm{f}})} \sum_{t=2}^{T-1} \|x_t - \mathbb{E}[x_t|\mathcal{F}_t(\xi_{\mathrm{tr}})]\|_{r'} \leq \varepsilon_{\mathrm{f}}. \tag{2}$$

Since it holds

$$D_{\mathrm{f}}(\xi, \xi_{\mathrm{tr}}) \leq \varepsilon_{\mathrm{appr}} + \inf_{x \in S(\xi_{\mathrm{f}})} \sum_{t=2}^{T-1} \|x_t - \mathbb{E}[x_t|\mathcal{F}_t(\xi_{\mathrm{tr}})]\|_{r'},$$

if ξ_{f} is sufficiently close to ξ, we obtain in case $\varepsilon_{\mathrm{appr}} + \varepsilon_{\mathrm{r}} \leq \delta$ that

$$|v(\xi) - v(\xi_{\mathrm{tr}})| \leq L(2\varepsilon_{\mathrm{appr}} + \varepsilon_{\mathrm{r}} + \varepsilon_{\mathrm{f}}).$$

(1) Forward tree generation

Let scenarios ξ^i with probabilities p_i, $i = 1, \ldots, N$, fixed root $\xi^*_1 \in I\!R^d$, $r \geq 1$, 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 $\hat{\xi}^1 := \xi_{\mathrm{f}}$ and $\mathcal{C}_1 = \{I = \{1, \ldots, N\}\}$.

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

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

where $i^k_t(j) \in I^k_t$ such that

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

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

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

such that

$$\|\hat{\xi}^t - \hat{\xi}^{t-1}\|_{r,t} \leq \varepsilon_t.$$

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 $\xi_{\rm tr}$ having K_T scenarios $\xi^k_{\rm tr}$ such that $\xi^k_{{\rm tr},t} := \xi_t^{\alpha_t(i)}$ with probabilities $\pi^i_T = \sum\limits_{j \in C_T^k} p_j$ if $i \in C_T^k$, $k = 1, \ldots, K_T$, $t = 2, \ldots, T$.

Proposition: $\|\xi_{\rm f} - \xi_{\rm tr}\|_r \leq \sum\limits_{t=2}^{T} \varepsilon_t \leq \varepsilon_{\rm r}$.

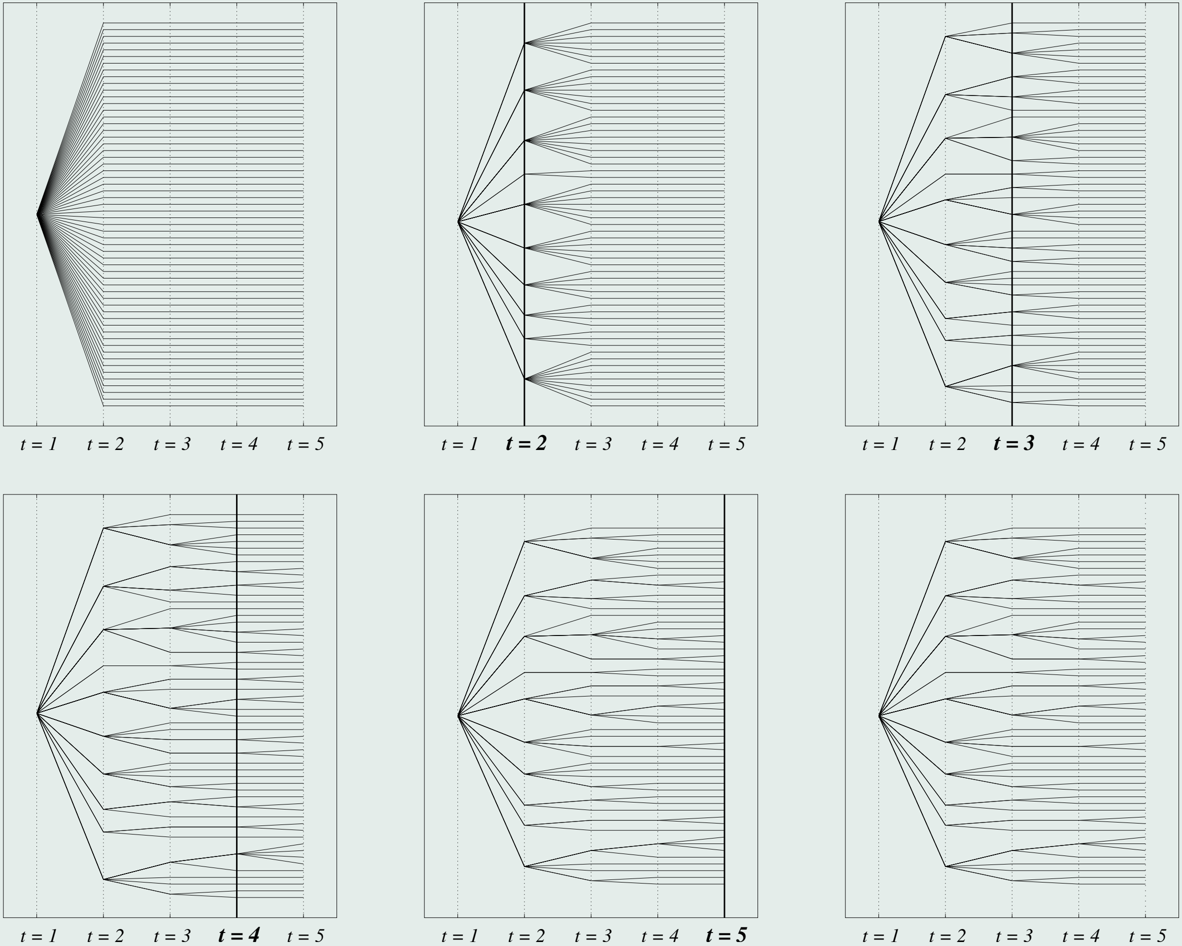


Illustration of the forward tree construction for an example including T=5 time periods starting with a scenario fan containing N=58 scenarios

<Start Animation>

(2) Bounding approximate filtration distances

Aim: $$\Delta(\xi_{\mathrm{f}},\xi_{\mathrm{tr}}) := \inf_{x\in S(\xi_{\mathrm{f}})} \sum_{t=2}^{T-1} \|x_t - \mathbb{E}[x_t|\mathcal{F}_t(\xi_{\mathrm{tr}})]\|_{r'} \leq \varepsilon_{\mathrm{f}}$$

Two possibilities:

(i) Estimates in terms of some solutions with input ξ_{f}, which would require to solve a two-stage model.

(ii) Estimates in terms of the input ξ_{f}.

Proposition:
Let (A2) and (A3) be satisfied and X_1 be bounded. Assume that the technology matrices $A_{t,1}$ are non-random, $1 \leq r' < \infty$ and ξ_{f} is sufficiently close to ξ. Then there exists a constant $\hat{L} > 0$ such that

$$\Delta(\xi_{\mathrm{f}},\xi_{\mathrm{tr}}) \leq \hat{L}\Big(\sum_{i\in I_2}\sum_{j\in I_{2,i}} p_j|\xi^j - \xi^i|^{r'}\Big)^{\frac{1}{r'}}$$

Condition: $$\sum_{i\in I_2}\sum_{j\in I_{2,i}} p_j|\xi^j - \xi^i|^{r'} \leq \varepsilon_{\mathrm{f}}^{r'}$$

Numerical experience

We consider the electricity portfolio management of a municipal power company. Data was available on the electrical load demand and on electricity prices at the market place EEX.

A multivariate statistical model is developed for the yearly demand-price process ξ that allowed to generate yearly demand-price scenarios ξ^i, with probabilities $p_i = \frac{1}{N}$, $i = 1, \ldots, N$.

These scenarios are assumed to form the process ξ_{f}. Branching in ξ_{tr} was allowed at most monthly. The tolerances ε_t at branching points were chosen such that

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

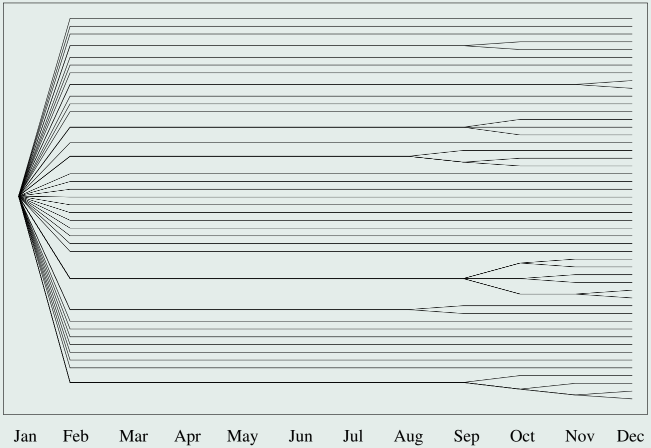
where the parameter $\overline{q} \in [0, 1]$ affects the branching structure of the constructed trees. For the test runs we used $\overline{q} = 0.6$.

The test runs were performed on a PC with a 3 GHz Intel Pentium CPU and 1 GByte main memory.

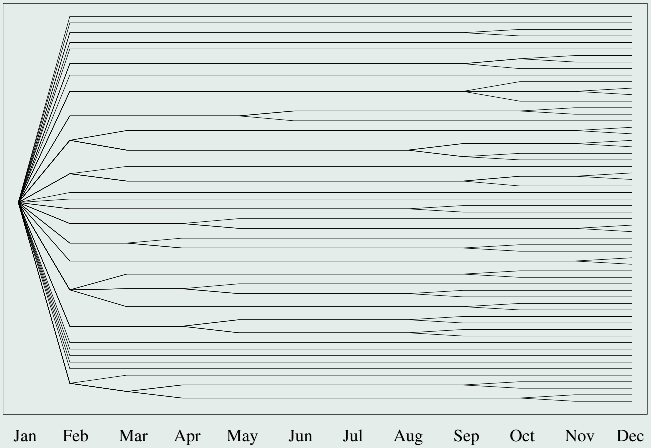
a) Forward tree construction with relative filtration tolerance $\varepsilon_{\mathrm{rel,f}} = 0.35$

b) Forward tree construction with relative filtration tolerance $\varepsilon_{\mathrm{rel,f}} = 0.45$

Yearly demand-price scenario trees with relative tolerance $\varepsilon_{\mathrm{rel,r}} = 0.25$



a) Forward tree construction with relative filtration tolerance $\varepsilon_{\mathrm{rel,f}} = 0.6$



b) Forward tree construction with relative filtration tolerance $\varepsilon_{\mathrm{rel,f}} = 0.7$

Yearly demand-price scenario trees with relative tolerance $\varepsilon_{\mathrm{rel,r}} = 0.5$

$\varepsilon_{\text{rel,r}}$	$\varepsilon_{\text{rel,f}}$	Scenarios	Nodes	Stages	Time (sec)
0.10	0.20	98	774 988	6	25.01
	0.30	99	774 424	6	25.05
0.15	0.25	94	719 714	12	24.97
	0.35	94	723 495	10	24.99
0.20	0.30	90	670 321	9	24.94
	0.40	90	670 478	10	24.94
0.25	0.35	85	619 296	9	24.95
	0.45	87	620 340	10	24.93
0.30	0.40	80	547 824	11	24.86
	0.50	83	567 250	11	24.91
0.35	0.45	72	482 163	11	24.94
	0.55	76	498 732	11	24.90
0.40	0.50	67	426 794	8	24.92
	0.60	71	444 060	11	24.90
0.45	0.55	60	368 380	7	24.97
	0.65	65	383 556	11	24.87
0.50	0.60	50	309 225	6	24.99
	0.70	60	319 380	11	24.88
0.55	0.65	44	247 303	6	25.00
	0.75	51	265 336	10	24.91
0.60	0.70	37	188 263	6	25.17
	0.80	45	203 321	9	24.98

Numerical results for yearly demand-price scenario trees