Scenario tree modelling for multistage stochastic programs

H. Heitsch and W. Römisch

Humboldt-University Berlin

Institute of Mathematics

10099 Berlin, Germany

Abstract

An important issue for solving multistage stochastic programs consists in the approximate representation of the (multivariate) stochastic input process in the form of a scenario tree. In this paper, forward and backward approaches are developed for generating scenario trees out of an initial fan of individual scenarios. Both approaches are motivated by the recent stability result in [15] for optimal values of multistage stochastic programs. They are based on upper bounds for the two relevant ingredients of the stability estimate, namely, the probabilistic and the filtration distance, respectively. These bounds allow to control the process of recursive scenario reduction [13] and branching. Numerical experience is reported for constructing multivariate scenario trees in electricity portfolio management.

Key Words: Stochastic programming, multistage, stability, L_r-distance, filtration, scenario tree, scenario reduction.

2000 MSC: 90C15

1 Introduction

Multiperiod stochastic programs are often used to model practical decision processes over time and under uncertainty, e.g., in finance, production, energy and logistics. Their inputs are multivariate stochastic processes $\{\xi_t\}_{t=1}^T$ defined on some probability space $(\Omega, \mathcal{F}, I\!\!P)$ and with ξ_t taking values in some $I\!\!R^d$. The decision x_t at t belonging to $I\!\!R^{m_t}$ is assumed to be *nonanticipative*, i.e., to depend only on $(\xi_1, \ldots, \xi_t)$. This property is equivalent to the measurability of x_t with respect to the σ-field $\mathcal{F}_t \subseteq \mathcal{F}$, which is generated by $\xi^t := (\xi_1, \ldots, \xi_t)$. Clearly, we have $\mathcal{F}_t \subseteq \mathcal{F}_{t+1}$ for $t = 1, \ldots, T-1$. Since at time $t = 1$ the input is known, we assume that $\mathcal{F}_1 = \{\emptyset, \Omega\}$ and, without loss of generality, that $\mathcal{F}_T = \mathcal{F}$.

The multiperiod stochastic program is assumed to be of the form

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

where the subsets X_t of $\mathbb{R}^{m_t}$ are nonempty and polyhedral, the cost coefficients $b_t(\xi_t)$ belong to $\mathbb{R}^{m_t}$, the right-hand sides $h_t(\xi_t)$ are in $\mathbb{R}^{n_t}$, $A_{t,0}$ are fixed (n_t, m_t)-matrices and $A_{t,1}(\xi_t)$ are (n_t, m_{t-1})-matrices, respectively. We assume that $b_t(\cdot)$, $h_t(\cdot)$ and $A_{t,1}(\cdot)$ depend affinely linearly on ξ_t covering the situation that some of the components of b_t and h_t, and of the elements of $A_{t,1}$ are random.

While the first and third groups of constraints in (1) have to be satisfied pointwise with probability 1, the second group, the measurability, *filtration* or *information* constraints, are functional and non-pointwise at least if $T > 2$ and $\mathcal{F}_1 \subsetneqq \mathcal{F}_t \subsetneqq \mathcal{F}_T$ for some $1 < t < T$. In the latter case (1) is called *multistage*. The presence of such qualitatively different constraints constitutes the origin of both the theoretical and computational challenges of multistage models.

The main computational approach to multistage stochastic programs consists in approximating the stochastic process $\xi = \{\xi_t\}_{t=1}^T$ by a process having finitely many scenarios exhibiting tree structure and starting at a fixed element ξ_1 of $\mathbb{R}^d$. This leads to linear programming models that are very large scale in most cases and can be solved by decomposition methods that exploit specific structures of the model. We refer to [32, Chapter 3] for a recent survey.

Presently, there exist several approaches to generate scenario trees for multistage stochastic programs (see [4] for a survey of ideas and methods until 2000). They are based on several different principles. We mention here (i) bound-based constructions [1, 8], (ii) Monte Carlo-based schemes [2, 33, 34] or Quasi Monte Carlo-based methods [22, 21], (iii) EVPI-based sampling and reduction within decomposition schemes [3], (iv) the moment-matching principle [17, 18], (v) probability metric based approximations [10, 11, 16, 23]. Many of them require to prescribe the tree structure and offer different strategies for selecting scenarios. We also mention the importance of evaluating the quality of scenario trees and of a postoptimality analysis [4, 19].

In the present paper we study and extend the scenario tree generation technique of [10, 11]. Its idea is to start with a good initial approximation of the underlying stochastic input process ξ consisting of a fan $\hat{\xi}$ of individual scenarios. These scenarios might be obtained by sampling or resampling techniques based on parametric or non-parametric stochastic models of ξ. Starting from $\hat{\xi}$, a tree ξ_{tr} is constructed by deleting and bundling scenarios recursively. While the recursive method described in [10, 11] works backward in time, a forward method was recently proposed in [14]. The aim of the paper is twofold: (i) For both (backward and forward) tree generation techniques we derive error estimates for the L_r-distance $\|\hat{\xi} - \xi_{\text{tr}}\|_r$. (ii) Upper bounds are obtained for the filtration distance of $\hat{\xi}$ and ξ_{tr}, which allow to recover the filtration structure of the original input process ξ approximately. The use of the filtration distance together with the selection of $r \geq 1$ for the L_r-distance is motivated by the recent stability result in [15] for multistage models. In this way, a (stability) theory-based heuristic is developed which generates a scenario tree that approximates the probability distribution and the filtration structure of ξ simultaneously.

The backward and forward tree generation methods were implemented and tested on real-life data in several practical applications, namely, for generating passenger demand scenario trees in airline revenue management [20] and for load-price scenario trees in electricity portfolio management [6]. Incorporating the filtration distance into the backward or forward tree generation schemes has not been tested so far.

Section 2 contains some prerequisites on distances of probability distributions and random vectors, and a short introduction to scenario reduction. Section 3 records the main stability result of [15], which provides the basis of our tree constructions. Section 4 contains the main results of our paper, in particular, the tree generation algorithms and error estimates in terms of L_r- and filtration distances, respectively. In Section 5 we discuss some numerical experience on backward and forward generation of load-inflow scenario trees based on realistic data. Numerical results of a variant of the forward tree construction with integrated filtration distance estimate are also presented.

2 Distances and scenario reduction

In earlier works on quantitative stability of stochastic programs without information constraints, probability metrics for measuring the distance of probability distributions played a major role [25, 30]. In particular, distances given in terms of Monge-Kantorovich mass transportation problems became relevant. They are of the form

$$\inf\Big\{\int_{\Xi\times\Xi} c(\xi,\tilde{\xi})\eta(d\xi,d\tilde{\xi}) : \eta\in\mathcal{P}(\Xi\times\Xi),\ \pi_1\eta=P,\ \pi_2\eta=Q\Big\}, \tag{2}$$

where Ξ is a closed subset of some Euclidean space, π_1 and π_2 denote the projections onto the first and second components, respectively, c is a nonnegative, symmetric and continuous cost function, and P and Q belong to a set $\mathcal{P}_c(\Xi)$ of probability measures on Ξ, which is chosen such that all occurring integrals are finite. Two types of cost functions have been used in stability analysis [5, 31], namely,

$$c(\xi,\tilde{\xi}) := \|\xi-\tilde{\xi}\|^r \quad (\xi,\tilde{\xi}\in\Xi) \tag{3}$$

and

$$c(\xi,\tilde{\xi}) := \max\{1,\|\xi-\xi_0\|^{r-1},\|\tilde{\xi}-\xi_0\|^{r-1}\}\|\xi-\tilde{\xi}\| \quad (\xi,\tilde{\xi}\in\Xi) \tag{4}$$

for some $r\geq 1$, $\xi_0\in\Xi$ and a seminorm or a norm $\|\cdot\|$ in the Euclidean space containing Ξ. In both cases, the set $\mathcal{P}_c(\Xi)$ may be chosen as the set $\mathcal{P}_r(\Xi)$ of all probability measures on Ξ having absolute moments of order r. The cost (3) leads to L_r*-minimal metrics* ℓ_r [27], which are defined by

$$\ell_r(P,Q) := \inf\left\{\int_{\Xi\times\Xi}\|\xi-\tilde{\xi}\|^r\eta(d\xi,d\tilde{\xi})\,|\eta\in\mathcal{P}(\Xi\times\Xi),\ \pi_1\eta=P,\ \pi_2\eta=Q\right\}^{\frac{1}{r}} \tag{5}$$

and sometimes also called Wasserstein metrics of order r [9]. The mass transportation problem (2) with cost (4) defines the Monge-Kantorovich functional $\hat{\mu}_r$ [24, 26]. A variant of the functional $\hat{\mu}_r$ appears if, in its definition (2), the conditions $\eta\in\mathcal{P}(\Xi\times\Xi), \pi_1\eta=P,\ \pi_2\eta=Q$ are replaced by η, which is a finite measure on $\Xi\times\Xi$, such that $\pi_1\eta-\pi_2\eta=P-Q$. The corresponding functionals $\mathring{\mu}_r$ turn out to be metrics on $\mathcal{P}_r(\Xi)$. They are called Fortet-Mourier metrics of order r [24]. The convergence of sequences of probability measures with respect to both metrics ℓ_r and $\mathring{\mu}_r$ is equivalent to their weak convergence and the convergence of their r-th order absolute moments.

For stochastic programs containing information constraints the situation is different. Examples (e.g., [15, Example 2.6]) show that a stability analysis based only on distances of probability distributions may fail. In the recent paper [15] quantitative stability of multistage stochastic programs (1) is proved with respect to the sum of two distances, namely, the norm

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

in $L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ with $s := Td$ for the Ξ-valued random inputs and the so-called information or *filtration distance*. The latter is defined in terms of the norm $\|\cdot\|_{r'}$ with r' depending on r. Its precise definition is given in Section 3.

Let ξ and $\tilde{\xi}$ be random vectors on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with probability distributions P and Q. Since the probability distribution $\bar{\eta}$ of the pair $(\xi, \tilde{\xi})$ of two Ξ-valued random vectors is feasible for the minimization problem (5), we have

$$\ell_r(P, Q) \leq \|\xi - \tilde{\xi}\|_r. \tag{6}$$

Moreover, since an optimal solution $\eta^* \in \mathcal{P}(\Xi \times \Xi)$ of the mass transportation problem (5) always exists (cf. [24, Theorem 8.1.1]), there are a probability space and a pair of Ξ-valued random vectors, a so-called *optimal coupling*, defined on it, such that the probability distribution of the pair is just η^* (e.g., [24, Theorem 2.5.1]). Hence, equality is valid in (6) on some probability space. This fact justifies the name L_r-minimal metric for ℓ_r.

Now, let ξ and $\tilde{\xi}$ be discrete random vectors with scenarios ξ^i with probabilities p_i, $i = 1, \ldots, N$, and $\tilde{\xi}^j$ with probabilities q_j, $j = 1, \ldots, M$, respectively. Then we have

$$\ell_r^r(P, Q) = \min \Big\{ \sum_{i,j} \eta_{ij} \|\xi^i - \tilde{\xi}^j\|^r : \eta_{ij} \geq 0, \sum_i \eta_{ij} = q_j, \sum_j \eta_{ij} = p_i \Big\}, \tag{7}$$

i.e., $\ell_r^r(P, Q)$ is the optimal value of a linear transportation problem. A case of particular interest consists in the situation that $M < N$ and that the scenarios of Q form a subset $\{\xi^j\}_{j \notin J}$ of the scenario set $\{\xi^i : i = 1, \ldots, N\}$ of P. One might first wish to solve the problem of finding the best approximation of P with respect to ℓ_r by a probability measure Q_J supported by the (scenario) set $\{\xi^j\}_{j \notin J}$, i.e., to determine the minimal distance D_J and an optimal solution $\{\bar{q}_j : j \notin J\}$ such that $\ell_r(P, Q_J)$ is minimized on the simplex $\{q : q_j \geq 0, \sum_{j \notin J} q_j = 1\}$. From [5, Theorem 2] we conclude

Lemma 2.1 *Let J be a nonempty subset of $\{1, \ldots, N\}$. Then the identity*

$$D_J = \min \Big\{ \ell_r(P, Q_J) : q_i \geq 0, \sum_{i \notin J} q_i = 1 \Big\} = \Big(\sum_{j \in J} p_j \min_{i \notin J} \|\xi^i - \xi^j\|^r \Big)^{\frac{1}{r}} \tag{8}$$

*holds and the minimum is attained at $\bar{q}_i = p_i + \sum_{j \in J_i} p_j$, $i \notin J$, where $J_i := \{j \in J | i = i(j)\}$ and $i(j)$ belongs to $\arg\min_{i \notin J} \|\xi^i - \xi^j\|$ for every $j \in J$ (*optimal redistribution*).*

Let the probability space be defined by $\Omega = \{\omega_1, \ldots, \omega_N\}$, $\mathcal{F}$ be the power set of Ω and $I\!P(\omega_i) = p_i$, $i = 1, \ldots, N$. If the random vector ξ_J is defined by

$$\xi_J(\omega_i) := \begin{cases} \xi^i & , i \notin J, \\ \xi^{i(j)} & , j \in J, \end{cases}$$

where $i(j)$ is defined as in Lemma 2.1, we obtain

$$\|\xi - \xi_J\|_r^r = \sum_{j \in J} p_j \|\xi^i - \xi^{i(j)}\|^r = \sum_{j \in J} p_j \min_{i \notin J} \|\xi^i - \xi^j\|^r = D_J^r \, .$$

Hence, the distance $\ell_r(P, Q_J)$ is minimal if Q_J is the probability distribution of ξ_J. Consequently, scenario reduction with respect to the L_r-minimal distance may alternatively be considered with respect to the norm $\|\cdot\|_r$ on this specific probability space.

Using the explicit formula (8), the optimal reduction problem for a scenario index set J with prescribed cardinality $|J| = N - n$ from P is given by the combinatorial optimization model

$$\min \Big\{ D_J = \sum_{j \in J} p_j \min_{i \notin J} \|\xi^i - \xi^j\|^r : J \subset \{1, ..., N\}, |J| = N - n \Big\}. \tag{9}$$

For the two extremal cases $n = N - 1$ and $n = 1$ the problem (9) is of the form

$$\min_{l \in \{1,\ldots,N\}} p_l \min_{i \neq l} \|\xi^l - \xi^i\|^r \quad (n = N-1) \quad \text{and} \quad \min_{u \in \{1,\ldots,N\}} \sum_{\substack{j=1 \\ j \neq u}}^{N} p_j \|\xi^u - \xi^j\|^r \quad (n = 1),$$

and easily solvable. Their solutions $J = \{l^*\}$ and $J = \{1, \ldots, N\} \setminus \{u^*\}$ arise as the result of two different processes: Backward reduction and forward selection. Both process ideas may be extended and lead to the following two heuristics for finding approximate solutions of (9). Their results are the index sets $J^{[N-n]}$ and $J^{[n]}$, respectively, of deleted scenarios and have cardinality $N - n$.

Algorithm 2.2 *(Backward reduction)*

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

Step [i]: $l_i \in \arg\min_{l \notin J^{[i-1]}} \sum_{k \in J^{[i-1]} \cup \{l\}} p_k \min_{j \notin J^{[i-1]} \cup \{l\}} \|\xi^k - \xi^j\|^r$,

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

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

Algorithm 2.3 *(Forward selection)*

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

Step [i]: $u_i \in \arg\min_{u \in J^{[i-1]}} \sum_{k \in J^{[i-1]} \setminus \{u\}} p_k \min_{j \notin J^{[i-1]} \setminus \{u\}} \|\xi^k - \xi^j\|^r$,

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

Step [n+1]: *Optimal redistribution.*

These heuristics were studied in [13] for different cost functions c. There it is shown that both algorithms exhibit polynomial complexity. Although the algorithms do not lead to optimality in general, the performance evaluation of their implementations in [13] is very encouraging.

3 Stability of multistage models

Here, we record the main result of the recent paper [15]. We assume that the stochastic input process ξ belongs to the Banach space $L_r(\Omega, \mathcal{F}, I\!P; I\!R^s)$ and $r \geq 1$. The multistage model (1) is regarded as an optimization problem in the space $L_{r'}(\Omega, \mathcal{F}, I\!P; I\!R^m)$ with $m = \sum_{t=1}^T m_t$ and endowed with the norm

$$\|x\|_{r'} := \left(\sum_{t=1}^{T} I\!E[\|x_t\|^{r'}] \right)^{\frac{1}{r'}} \quad (1 \leq r' < \infty) \quad \text{or} \quad \|x\|_\infty := \max_{t=1,\ldots,T} \operatorname{ess\,sup} \|x_t\|,$$

where the number r' is defined by

$$r' := \begin{cases} \frac{r}{r-1} & , \text{ if only costs are random} \\ r & , \text{ if only right-hand sides are random} \\ r = 2 & , \text{ if only costs and right-hand sides are random} \\ \infty & , \text{ if all technology matrices are random and } r = T. \end{cases} \tag{10}$$

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

$$\mathcal{X}(\xi) := \{x = (x_1, x_2, \ldots, x_T) \in \times_{t=1}^T L_{r'}(\Omega, \mathcal{F}_t, I\!P; I\!R^{m_t}) | x_1 \in X_1, x_t \in \mathcal{X}_t(x_{t-1}; \xi_t)\}$$

the set of feasible elements of (1) with input Ξ. Then the multistage stochastic program (1) may be rewritten as

$$\min\{F(\xi, x) \,:\, x \in \mathcal{X}(\xi)\}. \tag{11}$$

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

$$l_\alpha(F(\xi, \cdot)) := \{x \in \mathcal{X}(\xi) : F(\xi, x) \leq v(\xi) + \alpha\}$$

denote the *α-level set* of the stochastic program (11) with input ξ.

The following conditions are imposed on (11):

(A1) 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).

(A2) The optimal value $v(\xi)$ of (11) is finite 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 $\|\tilde{\xi} - \xi\|_r \leq \delta$.

(A3) $\xi \in L_r(\Omega, \mathcal{F}, I\!P; I\!R^s)$ for some $r \geq 1$.

The following stability result for optimal values of multistage stochastic programs is proved as [15, Theorem 2.1]. Its main observation is that the optimal value of a multistage model depends continuously on the stochastic input process if both its probability distribution and its filtration are approximated with respect to the L_r-distance and the filtration distance defined by (13), respectively.

Theorem 3.1 *Let (A1), (A2) and (A3) be satisfied and X_1 be bounded. Then there exist positive constants L, α and δ such that the estimate*

$$|v(\xi) - v(\tilde{\xi})| \leq L(\|\xi - \tilde{\xi}\|_r + D_{\mathrm{f}}(\xi, \tilde{\xi})) \tag{12}$$

holds for all random elements $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, I\!\!P; I\!\!R^s)$ with $\|\tilde{\xi} - \xi\|_r \leq \delta$. Here, $D_{\mathrm{f}}(\xi, \tilde{\xi})$ denotes the filtration distance *of ξ and $\tilde{\xi}$ defined by*

$$D_{\mathrm{f}}(\xi, \tilde{\xi}) := \sup_{\varepsilon \in (0,\alpha]} \inf_{\substack{x \in l_\varepsilon(F(\xi,\cdot)) \\ \tilde{x} \in l_\varepsilon(F(\tilde{\xi},\cdot))}} \sum_{t=2}^{T-1} \max\{\|x_t - I\!\!E[x_t|\tilde{\mathcal{F}}_t]\|_{r'}, \|\tilde{x}_t - I\!\!E[\tilde{x}_t|\mathcal{F}_t]\|_{r'}\}, \tag{13}$$

where $\mathcal{F}_t$ and $\tilde{\mathcal{F}}_t$ denote the σ-fields genarated by ξ^t and $\tilde{\xi}^t$, and $I\!\!E[\cdot|\mathcal{F}_t]$ and $I\!\!E[\cdot|\tilde{\mathcal{F}}_t]$, $t = 1, \ldots, T$, the corresponding conditional expectations, respectively.

An example in [15] shows that the filtration distance D_{f} is indispensable for Theorem 3.1 to hold. The filtration distance of two stochastic processes vanishes if their filtrations coincide, in particular, if the model is two-stage (i.e., $T = 2$). If solutions of (11) with inputs ξ and $\tilde{\xi}$ exist, the filtration distance is of the simplified form

$$D_{\mathrm{f}}(\xi, \tilde{\xi}) = \inf_{\substack{x \in l_0(F(\xi,\cdot)) \\ \tilde{x} \in l_0(F(\tilde{\xi},\cdot))}} \sum_{t=2}^{T-1} \max\{\|x_t - I\!\!E[x_t|\tilde{\mathcal{F}}_t]\|_{r'}, \|\tilde{x}_t - I\!\!E[\tilde{x}_t|\mathcal{F}_t]\|_{r'}\}. \tag{14}$$

For example, solutions of (11) exist if Ω is finite or if $1 < r' < \infty$ implying that the spaces $L_{r'}$ are finite-dimensional or reflexive Banach spaces (hence, the level sets are compact or weakly sequentially compact).

Theorem 3.1 is valid for any choice of the underlying probability space such that there exists a version of ξ with its probability distribution. The right-hand side of (12) is minimal if the probability space is selected such that both norms $\|\cdot\|_r$ and $\|\cdot\|_{r'}$ coincide with the corresponding L_r-minimal and $L_{r'}$-minimal distances (cf. the discussion in Section 2). However, for deriving estimates of the filtration distance, specific probability spaces might be more appropriate (see Section 4.3).

4 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 $\tilde{\xi}_{\mathrm{tr}}$ such that

$$\|\xi - \tilde{\xi}_{\mathrm{tr}}\|_r \qquad \text{and} \qquad D_{\mathrm{f}}(\xi, \tilde{\xi}_{\mathrm{tr}}) \tag{15}$$

are small and, hence, the optimal values $v(\xi)$ and $v(\tilde{\xi}_{\mathrm{tr}})$ are close to each other according to Theorem 3.1. Since this problem is hardly solvable in general, we replace ξ by a (good) finitely discrete approximation. This approximation is again denoted by ξ and its scenarios by $\xi^i = (\xi_1^i, \ldots, \xi_T^i)$ with probabilities p_i, $i = 1, \ldots, N$. We assume that all scenarios coincide at the first time period $t = 1$, i.e., $\xi_1^1 = \ldots = \xi_1^N =: \xi_1^*$. Hence, they form a *fan* of invidual scenarios. Such a fan may be regarded as a scenario tree with root node at $t = 1$ having N branches at the root and consisting of $1 + (T-1)N$

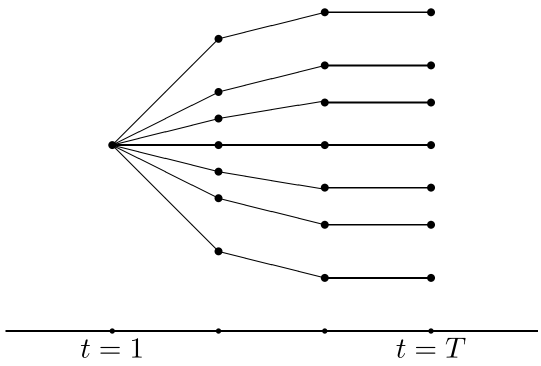

Figure 1: Example of a fan of individual scenarios with $T = 4$ and $N = 7$

nodes. If such a scenario fan is inserted into a multiperiod stochastic program (1), the model is two-stage as all σ-fields $\mathcal{F}_t$, $t = 2, \ldots, T$, coincide.

In this section we develop algorithmic procedures that produce scenario trees $\tilde{\xi}_{\mathrm{tr}}$ with root node ξ_1^*, less nodes than the original fan and that allow for constructive estimates of the L_r-norm $\|\xi - \tilde{\xi}_{\mathrm{tr}}\|_r$ and the corresponding filtration distance. Here, $r \geq 1$ is determined such that the optimal values of the underlying multistage stochastic program satisfy an estimate of the form (12) in Theorem 3.1. The idea of the algorithm consists in forming clusters of scenarios based on scenario reduction on the time horizon $\{1, \ldots, t\}$ recursively for decreasing and increasing time t, respectively.
To this end, the L_r-seminorm $\|\cdot\|_{r,t}$ on $L_r(\Omega, \mathcal{F}, I\!P; I\!R^s)$ (with $s = Td$) given by

$$\|\xi\|_{r,t} := \Big(I\!E[\|\xi\|_t^r] \Big)^{\frac{1}{r}} = \Big(\sum_{i=1}^{N} p_i \|\xi^i\|_t^r \Big)^{\frac{1}{r}} \tag{16}$$

is needed at step t. Here, we denote by $\|\cdot\|_t$ a seminorm on $I\!R^s$ that is defined by $\|\xi\|_t := \|(\xi_1, \ldots, \xi_t, 0, \ldots, 0)\|$ for each $\xi = (\xi_1, \ldots, \xi_T) \in I\!R^s$.

4.1 Backward tree construction

Setting $\bar{\xi}^{T+1} := \xi$, recursive scenario reduction on $\{1, \ldots, t\}$ for decreasing t leads to stochastic processes $\bar{\xi}^t$ having scenarios $\{\bar{\xi}^{t,i} := \xi^i\}_{i \in I_t}$ with $I_t \subset I := \{1, \ldots, N\}$ and increasing cardinality $|I_t|$. We obtain a chain of index sets

$$I_1 = \{i_*\} \subseteq I_2 \subseteq \cdots \subseteq I_{t-1} \subseteq I_t \subseteq \cdots \subseteq I_T \subseteq I_{T+1} := I$$

and denote the index set of deleted scenarios at t by $J_t := I_{t+1} \setminus I_t$ for each $t = 1, \ldots, T$. The probabilities π_t^i of the scenarios $\bar{\xi}^{t,i}$ for $i \in I_t$ are set by $\pi_{T+1}^i := p_i$ for $i \in I_{T+1}$ and further defined according to the optimal redistribution rule (see Lemma 2.1) for the norm $\|\cdot\|_t$, i.e.,

$$\pi_t^i = \pi_{t+1}^i + \sum_{j \in J_{t,i}} \pi_{t+1}^j \quad (i \in I_t), \tag{17}$$

where

$$J_t = \bigcup_{i \in I_t} J_{t,i}, \quad J_{t,i} := \{j \in J_t : i = i_t(j)\} \quad \text{and} \quad i_t(j) \in \arg\min_{i \in I_t} \|\xi^i - \xi^j\|_t. \tag{18}$$

At time t we obtain the scenario clusters $\bar{I}_{t,i} := \{i, j : j \in J_{t,i}\}$ for each $i \in I_t$ that form a partition of I_T, i.e., $I_T = \cup_{i \in I_t} \bar{I}_{t,i}$. The cardinality of $\bar{I}_{t,i}$ corresponds to the branching degree of scenario i at t. If $|\bar{I}_{t,i}| = 1$, i.e., $J_{t,i} = \emptyset$, scenario i will not branch at t. Lemma 2.1 also implies

$$\|\bar{\xi}^{t+1} - \bar{\xi}^t\|_{r,t}^r = \sum_{j \in J_t} \pi_{t+1}^j \min_{i \in I_t} \|\xi^i - \xi^j\|_t^r \tag{19}$$

for $t = 1, \ldots, T$. The final scenario tree $\tilde{\xi}_{\mathrm{tr}}$ consists of $|I_T|$ scenarios $\tilde{\xi}^j$ with probabilities π_T^j for $j \in I_T$. Each of its components $\tilde{\xi}_t^j$ is a node of degree $|\bar{I}_{t,j}| = 1 + |J_{t,j}|$ with probability π_t^j and belongs to the set $\{\xi_t^i\}_{i \in I_t}$. The corresponding index $i \in I_t$ is given by $i = \alpha_t(j)$, where the index mappings $\alpha_t : I \to I_t$ are defined recursively by setting α_{T+1} to be the identity and

$$\alpha_t(j) := \begin{cases} i_t(\alpha_{t+1}(j)) & , \alpha_{t+1}(j) \in J_t, \\ \alpha_{t+1}(j) & , \text{otherwise}, \end{cases} \tag{20}$$

for $j \in I$ and $t = T, \ldots, 1$. We obtain the following estimate for the L_r-distance of ξ and $\tilde{\xi}_{\mathrm{tr}}$.

Theorem 4.1 *Let the stochastic process ξ with fixed initial node ξ_1^*, scenarios ξ^i and probabilities p_i, $i = 1, \ldots, N$, be given. Let $\tilde{\xi}_{\mathrm{tr}}$ be the stochastic process with scenarios $\tilde{\xi}^i = (\xi_1^*, \xi_2^{\alpha_2(i)}, \ldots, \xi_t^{\alpha_t(i)}, \ldots, \xi_T^i)$ and probabilities π_T^i for $i \in I_T$. Then we have the estimate*

$$\|\xi - \tilde{\xi}_{\mathrm{tr}}\|_r \leq \sum_{t=2}^{T} \left(\sum_{j \in J_t} \pi_{t+1}^j \min_{i \in I_t} \|\xi^i - \xi^j\|_t^r \right)^{\frac{1}{r}}. \tag{21}$$

Proof: Let $\hat{\xi}^\tau$ be the stochastic process having scenarios $\hat{\xi}^{\tau,i}$ and probabilities π_T^i for $i \in I_T$, where

$$\hat{\xi}_t^{\tau,i} := \begin{cases} \xi_t^{\alpha_t(i)} & , t \geq \tau, \\ \xi_t^{\alpha_\tau(i)} & , t < \tau, \end{cases}$$

for $\tau = 1, \ldots, T$. The processes $\hat{\xi}^\tau$ are illustrated in Figure 2, where $\hat{\xi}^\tau$ corresponds to the $(T - \tau + 2)$-th picture for $\tau = 2, \ldots, T$. According to the above constructions we have $\hat{\xi}^T = \bar{\xi}^T$ and $\hat{\xi}^1 = \tilde{\xi}_{\mathrm{tr}}$. Next we show for $t = 1, \ldots, T-1$ that

$$\|\hat{\xi}^{t+1} - \hat{\xi}^t\|_r = \|\bar{\xi}^{t+1} - \bar{\xi}^t\|_{r,t}. \tag{22}$$

We have

$$\|\hat{\xi}^{t+1} - \hat{\xi}^t\|_r = \sum_{i \in I_T} \pi_T^i \|\hat{\xi}^{t+1,i} - \hat{\xi}^{t,i}\|^r. \tag{23}$$

Since the final $T-t$ components of the elements $\hat{\xi}^{t+1,i}$ and $\hat{\xi}^{t,i}$ are identical, the norm $\|\cdot\|$ may be replaced by the seminorm $\|\cdot\|_t$ in (23). Moreover, since the first t components of $\hat{\xi}^{t+1,i}$ and $\hat{\xi}^{t,i}$ are $\xi_\tau^{\alpha_{t+1}(i)}$ and $\xi_\tau^{\alpha_t(i)}$, respectively, $\tau = 1, \ldots, t$, we have

$$\sum_{i \in I_T} \pi_T^i \|\hat{\xi}^{t+1,i} - \hat{\xi}^{t,i}\|^r = \sum_{i \in I_T} \pi_T^i \|\xi^{\alpha_{t+1}(i)} - \xi^{\alpha_t(i)}\|_t^r.$$

Since $\alpha_t(j) = \alpha_{t+1}(j)$ holds for $\alpha_{t+1}(j) \notin J_t$ (see (20)), we obtain

$$\sum_{i \in I_T} \pi_T^i \|\xi^{\alpha_{t+1}(i)} - \xi^{\alpha_t(i)}\|_t^r = \sum_{\substack{i \in I_T \\ \alpha_{t+1}(i) \in J_t}} \pi_T^i \|\xi^{\alpha_{t+1}(i)} - \xi^{\alpha_t(i)}\|_t^r.$$

With (20) and (19) the latter sum may be rewritten as

$$\begin{aligned}
\sum_{\substack{i \in I_T \\ \alpha_{t+1}(i) \in J_t}} \pi_T^i \|\xi^{\alpha_{t+1}(i)} - \xi^{\alpha_t(i)}\|_t^r &= \sum_{j \in J_t} \sum_{\substack{k \in I_T \\ \alpha_{t+1}(k) = j}} \pi_T^k \|\xi^{\alpha_{t+1}(k)} - \xi^{\alpha_t(k)}\|_t^r \\
&= \sum_{j \in J_t} \Bigg(\sum_{\substack{k \in I_T \\ \alpha_{t+1}(k) = j}} \pi_T^k \Bigg) \|\xi^j - \xi^{i_t(j)}\|_t^r \\
&= \sum_{j \in J_t} \pi_{t+1}^j \|\xi^j - \xi^{i_t(j)}\|_t^r = \|\bar{\xi}^{t+1} - \bar{\xi}^t\|_{r,t}^r.
\end{aligned}$$

Hence, the proof of (22) for $t = 1, \ldots, T$ is complete.
Finally, we prove (21) by applying repeatedly the triangle inequality for $\|\cdot\|_r$, using (22) and the identities $\xi = \bar{\xi}^{T+1}$, $\hat{\xi}^T = \bar{\xi}^T$ and $\hat{\xi}^1 = \xi_{tr}$.

$$\begin{aligned}
\|\xi - \tilde{\xi}_{\mathrm{tr}}\|_r &\leq \|\xi - \hat{\xi}^T\|_r + \|\hat{\xi}^T - \tilde{\xi}_{\mathrm{tr}}\|_r \\
&\leq \|\bar{\xi}^{T+1} - \bar{\xi}^T\|_r + \sum_{k=1}^{T-1} \|\hat{\xi}^{T-k+1} - \hat{\xi}^{T-k}\|_r \\
&= \sum_{k=0}^{T-1} \|\bar{\xi}^{T-k+1} - \bar{\xi}^{T-k}\|_{r,T-k} \\
&= \sum_{t=2}^{T} \|\bar{\xi}^{t+1} - \bar{\xi}^t\|_{r,t}\,,
\end{aligned}$$

where for $t = 1$ the summand vanishes. Together with the representation (19) of $\|\cdot\|_{r,t}$, the proof is complete. $\square$

The preceding result allows to estimate the quality of scenario trees that are generated by the backward tree construction algorithm. For example, if the tree structure is *stagewise fixed*, say, to decreasing numbers $N_t \leq N$ as t decreases from T to 1, the algorithm selects almost best possible candidates for deletion and Theorem 4.1 allows to estimate the quality of the tree. In addition, the estimate (21) provides the possibility to quantify the relative error at time t and, hence, to modify the structure. If the tree structure is *free*, the following flexible algorithm allows to generate a variety of scenario trees satisfying a given accuracy *tolerance* with respect to the L_r-distances.

Algorithm 4.2 (backward tree construction)
Let N 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, ε_t, $t = 2, \ldots, T$, be given such that $\sum_{t=2}^{T} \varepsilon_t \leq \varepsilon$.

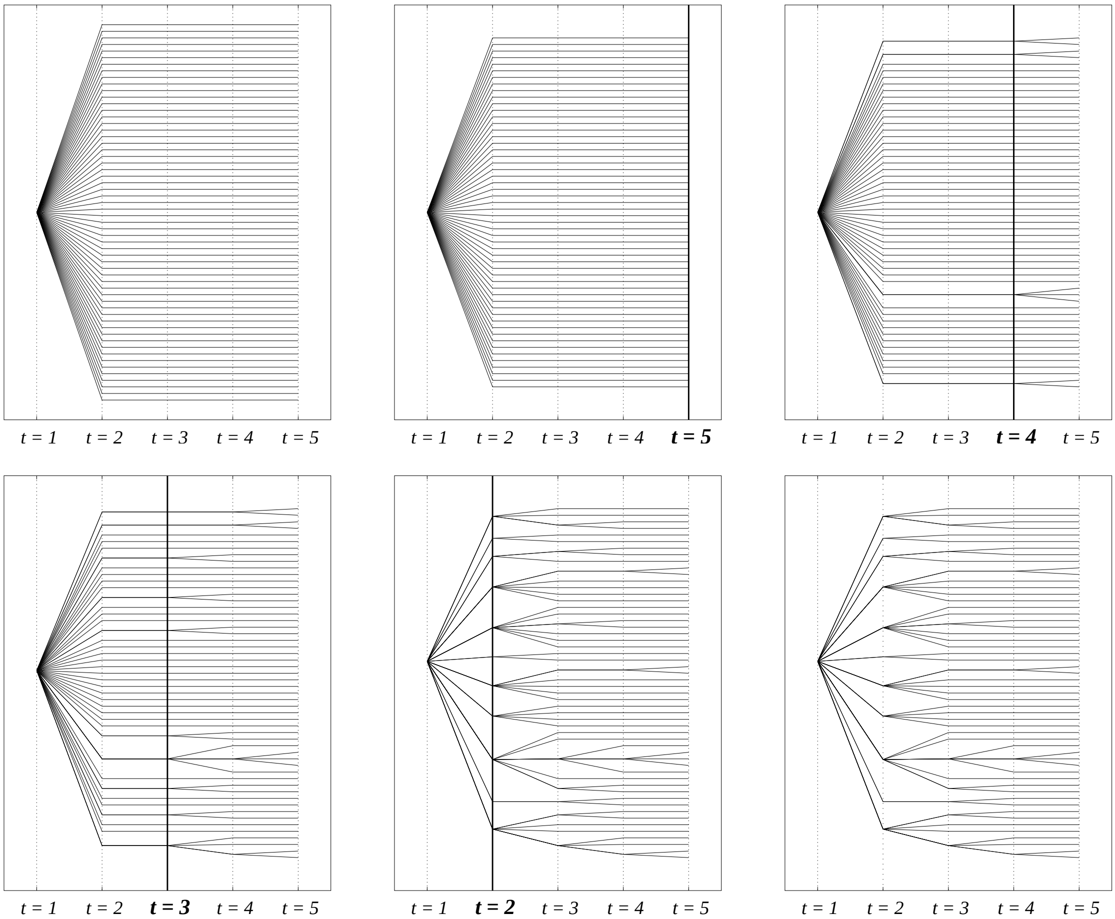

Figure 2: Illustration of the backward tree construction for an example including T=5 time periods starting with a scenario fan containing N=58 scenarios

Step 0: *Set* $\bar{\xi}^{T+1} := \xi$ *and* $I_{T+1} = \{1, \ldots, N\}$. *Determine an index set* $I_T \subseteq I_{T+1}$ *and a stochastic process* $\bar{\xi}^T$ *with* $|I_T|$ *scenarios such that* $\|\bar{\xi}^{T+1} - \bar{\xi}^T\|_r \leq \varepsilon_T$.

Step t: *Determine an index set* $I_{T-t} \subseteq I_{T-t+1}$ *and a stochastic process* $\bar{\xi}^{T-t}$ *with* $|I_{T-t}|$ *scenarios such that* $\|\bar{\xi}^{T-t+1} - \bar{\xi}^{T-t}\|_{r,T-t} \leq \varepsilon_{T-t}$.

Step T-1: *Construct the stochastic process* $\tilde{\xi}_{\mathrm{tr}}$ *having* $|I_T|$ *scenarios* $\tilde{\xi}^j$, $j \in I_T$, *such that* $\tilde{\xi}^j_t := \xi_t^{\alpha_t(j)}$, $t = 1, \ldots, T$, *where* $\alpha_t(\cdot)$ *is defined by (20).*

While the first picture in Figure 2 illustrates the original fan ξ, the second one corresponds to the situation after the reduction Step 0 and the third, fourth and fifth one to the Steps 1–3, respectively. The final picture corresponds to the final Step 4 and illustrates the scenario tree $\tilde{\xi}_{\mathrm{tr}}$.

Corollary 4.3 *Let a stochastic process* ξ *with fixed initial node* ξ_1^*, *scenarios* ξ^i *and probabilities* p_i, $i = 1, \ldots, N$, *be given. If* $\tilde{\xi}_{\mathrm{tr}}$ *is constructed according to Algorithm 4.2, we have*

$$\|\xi - \tilde{\xi}_{\mathrm{tr}}\|_r \leq \sum_{t=2}^{T} \varepsilon_t \leq \varepsilon.$$

Proof: This is a direct consequence of the estimate (21) in Theorem 4.1, which reads

$$\|\xi - \tilde{\xi}_{\text{tr}}\|_r \leq \sum_{t=2}^{T} \|\bar{\xi}^{t+1} - \bar{\xi}^t\|_{r,t}. \qquad \square$$

If the Algorithm 4.2 is used to generate scenario trees in practical applications, one has to select $r > 1$ and the tolerances ε_t, $t = 1, \ldots, T$. Often there are good reasons for selecting r according to the properties of the original process ξ and the desired approximation quality of the solutions expressed by the norm $\|\cdot\|_{r'}$. The choice of the *tolerances* ε_t, however, is essentially open so far. Clearly, branching at t occurs more often if ε_t gets larger and $\varepsilon_t = 0$ leads to no branching of scenarios at time t. Some experience on selecting the tolerances is reported in Section 5.1, where the (non-vanishing) tolerances are chosen according to the exponential rule (45).

4.2 Forward tree construction

The forward selection procedure determines recursively stochastic processes $\hat{\xi}^t$ having scenarios $\hat{\xi}^{t,i}$ endowed with probabilities p_i, $i \in I := \{1, \ldots, N\}$, and partitions $\mathcal{C}_t = \{C_t^1, \ldots, C_t^{K_t}\}$ of I, i.e., such that

$$C_t^k \cap C_t^{k'} = \emptyset \quad \forall k \neq k' \quad \text{and} \quad \bigcup_{k=1}^{K_t} C_t^k = I. \tag{24}$$

The elements of such a partition $\mathcal{C}_t$ will be called (scenario) clusters. The initialization of the procedure consists in setting $\hat{\xi}^1 = \xi$, i.e., $\hat{\xi}^{1,i} = \xi^i$, $i \in I$, and $\mathcal{C}_1 = \{I\}$. At step t (with $t > 1$) every cluster C_{t-1}^k, i.e., every scenario subset $\{\hat{\xi}^{t-1,i}\}_{i \in C_{t-1}^k}$, is considered separately and subjected to scenario reduction with respect to the seminorm $\|\cdot\|_t$ as described in Section 2. This leads to index sets I_t^k and J_t^k of remaining and deleted scenarios, respectively, where

$$I_t^k \cup J_t^k = C_{t-1}^k$$

and

$$J_t^k = \bigcup_{i \in I_t^k} J_{t,i}^k, \quad J_{t,i}^k := \{j \in J_t^k : i = i_t^k(j)\} \quad \text{and} \quad i_t^k(j) \in \arg\min_{i \in I_t^k} \|\hat{\xi}^{t-1,i} - \hat{\xi}^{t-1,j}\|_t^r.$$

Next we define a mapping $\alpha_t : I \to I$ such that

$$\alpha_t(j) = \begin{cases} i_t^k(j) & , j \in J_t^k,\ k = 1, \ldots, K_{t-1}, \\ j & , \text{otherwise}. \end{cases} \tag{25}$$

Then the scenarios of the stochastic process $\hat{\xi}^t = \{\hat{\xi}_\tau^t\}_{\tau=1}^T$ are defined by

$$\hat{\xi}_\tau^{t,i} = \begin{cases} \xi_\tau^{\alpha_\tau(i)} & , \tau \leq t, \\ \xi_\tau^i & , \text{otherwise}, \end{cases} \tag{26}$$

with probabilities p_i for each $i \in I$. The processes $\hat{\xi}^t$ are illustrated in Figure 3, where $\hat{\xi}^t$ corresponds to the t-th picture for $t = 1, \ldots, T$. The partition $\mathcal{C}_t$ at time t is defined by

$$\mathcal{C}_t = \{\alpha_t^{-1}(i) : i \in I_t^k,\ k = 1, \ldots, K_{t-1}\}, \tag{27}$$

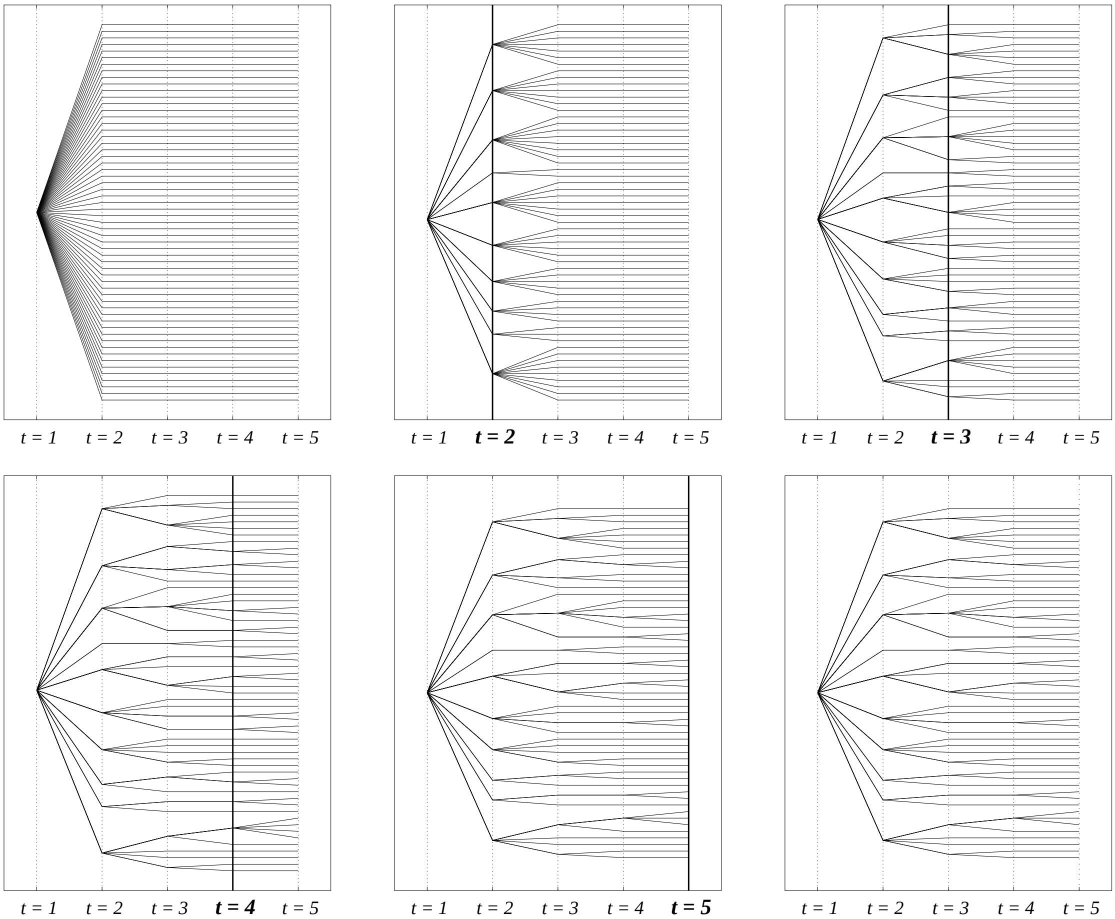

Figure 3: Illustration of the forward tree construction for an example including T=5 time periods starting with a scenario fan containing N=58 scenarios

i.e., each element of the index sets I_t^k defines a new cluster and the partition $\mathcal{C}_t$ is a refinement of the partition $\mathcal{C}_{t-1}$. The scenario sets I_t, scenario clusters $\bar{I}_{t,i}$ and cluster probabilities π_t^i in the description of the backward reduction procedure in the preceding subsection have now the form

$$I_t := \bigcup_{k=1}^{K_{t-1}} I_t^k$$

$$\bar{I}_{t,i} := \{i, j : j \in J_{t,i}^k\} = C_t^k \quad \text{and} \quad \pi_t^i = \sum_{j \in C_t^k} p_j \ \text{ if } \ i \in I_t^k \ \text{ for some } \ k = 1, \ldots, K_{t-1}.$$

The branching degree of scenario i at t coincides with the cardinality of $\bar{I}_{t,i}$.

Finally, the scenarios and their probabilities of the scenario tree $\tilde{\xi}_{\text{tr}} := \hat{\xi}^T$ are given by the structure of the final partition $\mathcal{C}_T$, i.e., they are of the form

$$\tilde{\xi}^k = (\xi_1^*, \xi_2^{\alpha_2(i)}, \ldots, \xi_t^{\alpha_t(i)}, \ldots, \xi_T^{\alpha_T(i)}) \quad \text{and} \quad \pi_T^i \quad \text{if} \ \ i \in C_T^k$$

for each $k = 1, \ldots, K_T$. Furthermore, we have the following error estimate with respect to the L_r-norm.

Theorem 4.4 *Let the stochastic process ξ with fixed initial node ξ_1^*, scenarios ξ^i and probabilities p_i, $i = 1, \dots, N$, be given. Let $\tilde{\xi}_{\mathrm{tr}}$ be the stochastic process with scenarios $\tilde{\xi}^k = (\xi_1^*, \xi_2^{\alpha_2(i)}, \dots, \xi_t^{\alpha_t(i)}, \dots, \xi_T^{\alpha_T(i)})$ and probabilities π_T^k if $i \in C_T^k$, $k = 1, \dots, K_T$. Then we have the estimate*

$$\|\xi - \tilde{\xi}_{\mathrm{tr}}\|_r \leq \sum_{t=2}^{T} \left(\sum_{k=1}^{K_{t-1}} \sum_{j \in J_t^k} p_j \min_{i \in I_t^k} \|\xi_t^i - \xi_t^j\|^r \right)^{\frac{1}{r}}. \tag{28}$$

Proof: We recall that $\hat{\xi}^1 = \xi$ and $\hat{\xi}^T = \tilde{\xi}_{\mathrm{tr}}$ and obtain

$$\|\xi - \tilde{\xi}_{\mathrm{tr}}\|_r \leq \sum_{t=2}^{T} \|\hat{\xi}^t - \hat{\xi}^{t-1}\|_r,$$

using the triangle inequality of $\|\cdot\|_r$. Since the scenarios of $\hat{\xi}^t$ and $\hat{\xi}^{t-1}$ coincide on $\{t+1, \dots, T\}$, the latter estimate may be rewritten as

$$\|\xi - \tilde{\xi}_{\mathrm{tr}}\|_r \leq \sum_{t=2}^{T} \|\hat{\xi}^t - \hat{\xi}^{t-1}\|_{r,t}. \tag{29}$$

By definition of $\hat{\xi}^t$ and $\hat{\xi}^{t-1}$ we have $\hat{\xi}_\tau^{t,i} = \hat{\xi}_\tau^{t-1,i}$ for all $\tau = 1, \dots, t-1$. Hence, we obtain

$$\begin{aligned}
\|\hat{\xi}^t - \hat{\xi}^{t-1}\|_{r,t}^r &= \sum_{i=1}^{N} p_i \|\hat{\xi}^{t,i} - \hat{\xi}^{t-1,i}\|_t^r = \sum_{k=1}^{K_{t-1}} \sum_{j \in C_{t-1}^k} p_j \|\hat{\xi}_t^{t,j} - \hat{\xi}_t^{t-1,j}\|^r \\
&= \sum_{k=1}^{K_{t-1}} \sum_{j \in C_{t-1}^k} p_j \|\xi_t^{\alpha_t(j)} - \xi_t^j\|^r = \sum_{k=1}^{K_{t-1}} \sum_{j \in J_t^k} p_j \|\xi_t^{i_t^k(j)} - \xi_t^j\|^r \\
&= \sum_{k=1}^{K_{t-1}} \sum_{j \in J_t^k} p_j \min_{i \in I_t^k} \|\xi_t^i - \xi_t^j\|^r,
\end{aligned}$$

using, in addition, the partition property (24) and the definitions (25) of the mappings α_t and i_t^k. Inserting the latter result into (29) completes the proof. □

The error estimate in Theorem 4.4 is very similar to that in Theorem 4.1. Both estimates allow to quantify the relative error of the t-th construction step. As in the previous section, we provide a flexible algorithm that allows to generate a variety of scenario trees satisfying a given approximation tolerance with respect to the L_r-distance.

Algorithm 4.5 (forward tree construction)
Let N scenarios ξ^i with probabilities p_i, $i = 1, \dots, N$, fixed root $\xi_1^* \in I\!R^d$ and probability distribution P, $r \geq 1$, and tolerances ε, ε_t, $t = 2, \dots, T$, be given such that $\sum_{t=2}^{T} \varepsilon_t \leq \varepsilon$.

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

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 *such that* $I_t^k \cup J_t^k = C_{t-1}^k$, *the mapping* $\alpha_t(\cdot)$ *according to (25) and a stochastic process* $\hat{\xi}^t$ *having* N *scenarios* $\hat{\xi}^{t,i}$ *with probabilities* p_i *according to (26) and such that* $\|\hat{\xi}^t - \hat{\xi}^{t-1}\|_{r,t} \leq \varepsilon_t$. *Set* $\mathcal{C}_t = \{\alpha_t^{-1}(i) : i \in I_t^k,\ k = 1, \ldots, K_{t-1}\}$.

Step T+1: *Let* $\mathcal{C}_T = \{C_T^1, \ldots, C_T^{K_T}\}$. *Construct a stochastic process* $\tilde{\xi}_{\mathrm{tr}}$ *having* K_T *scenarios* $\hat{\xi}^k$ *such that* $\hat{\xi}_t^k := \xi_t^{\alpha_t(i)}$ *if* $i \in C_T^k$, $k = 1, \ldots, K_T$, $t = 1, \ldots, T$.

While the first picture in Figure 3 illustrates the original fan ξ, the second, third, fourth and fifth ones correspond to the situation after the Steps 2–5. The final picture corresponds to Step 6 and illustrates the scenario tree $\tilde{\xi}_{\mathrm{tr}}$.

Corollary 4.6 *Let a stochastic process* ξ *with fixed initial node* ξ_1^*, *scenarios* ξ^i *and probabilities* p_i, $i = 1, \ldots, N$, *be given. If* $\tilde{\xi}_{\mathrm{tr}}$ *is constructed by Algorithm 4.5, we have*

$$\|\xi - \tilde{\xi}_{\mathrm{tr}}\|_r \leq \sum_{t=2}^{T} \varepsilon_t \leq \varepsilon.$$

Proof: This is a direct consequence of (29). □

When using Algorithm 4.5, the selection of $r > 1$ should be done according to the same reasons as mentioned at the end of Section 4.1. The choice of the tolerances ε_t, however, is different. Here, it is suggested to choose nonincreasing ε_t, $t = 2, \ldots, T$. The smaller ε_t is, the more branchings occur at t. Some experience on selecting the tolerances is provided by the rule (46) in Section 5.2.

4.3 Estimating filtration distances

Let ξ be the (discrete) approximation of the original stochastic process and $\tilde{\xi} = \tilde{\xi}_{\mathrm{tr}}$ be the process obtained by means of one of the tree construction approaches in Sections 4.1 and 4.2, respectively. So far we are able to estimate the first ingredient $\|\xi - \tilde{\xi}_{\mathrm{tr}}\|_r$ of the stability estimate (12) in Theorem 3.1. Here, we derive estimates for the second ingredient $D_{\mathrm{f}}(\xi, \tilde{\xi}_{\mathrm{tr}})$ and develop strategies for controlling the tree generation process by bounding both distances.

Next, we consider two stochastic processes ξ and $\tilde{\xi}$ given in the form of scenario trees. We assume that conditions (A1) and (A2) of Section 3 are satisfied and derive estimates for the bound

$$D_{\mathrm{f}}(\xi, \tilde{\xi}) \leq \begin{cases} \sum\limits_{t=2}^{T-1} \max\left\{ \mathbb{E}[\|x_t - \mathbb{E}[x_t|\tilde{\mathcal{F}}_t]\|^{r'}], \mathbb{E}[\|\tilde{x}_t - \mathbb{E}[\tilde{x}_t|\mathcal{F}_t]\|^{r'}] \right\}^{\frac{1}{r'}}, & 1 \leq r' < \infty \\ \sum\limits_{t=2}^{T-1} \max\left\{ \|x_t - \mathbb{E}[x_t|\tilde{\mathcal{F}}_t]\|_\infty, \|\tilde{x}_t - \mathbb{E}[\tilde{x}_t|\mathcal{F}_t]\|_\infty \right\} & , r' = \infty \end{cases} \tag{30}$$

of the filtration distance of ξ and $\tilde{\xi}$, respectively, defined by (13). Here, x and $\tilde{x}$ are solutions of (11) with inputs ξ and $\tilde{\xi}$, respectively, and r' is defined by (10).

To this end, we assume that $\xi = \{\xi_t\}_{t=1}^T$ and $\tilde{\xi} = \{\tilde{\xi}_t\}_{t=1}^T$ are defined on the probability space $(\Omega, \mathcal{F}, I\!P)$ with $\Omega = \{\omega_1, \ldots, \omega_N\}$, $\mathcal{F}$ denoting the power set of Ω and $I\!P(\omega_i) = p_i$, $i = 1, \ldots, N$. Let $\mathcal{I}_t$ and $\tilde{\mathcal{I}}_t$ denote the index set of realizations of ξ_t and $\tilde{\xi}_t$, respectively. Furthermore, let $\mathcal{E}_t$ and $\tilde{\mathcal{E}}_t$ denote families of nonempty elements of $\mathcal{F}_t$ and $\tilde{\mathcal{F}}_t$, respectively, that form partitions of Ω and generate the corresponding σ-fields. We set $E_{ts} := \{\omega \in \Omega : (\xi_1(\omega), \ldots, \xi_t(\omega)) = (\xi_1^s, \ldots, \xi_t^s)\}$, $s \in \mathcal{I}_t$, and $\tilde{E}_{ts} := \{\omega \in \Omega : (\tilde{\xi}_1(\omega), \ldots, \tilde{\xi}_t(\omega)) = (\tilde{\xi}_1^s, \ldots, \tilde{\xi}_t^s)\}$, $s \in \tilde{\mathcal{I}}_t$. For the t-th summand of the bound (30) we introduce the notation

$$D_t(\xi, \tilde{\xi}) := \begin{cases} \max\left\{ I\!E[\|x_t - I\!E[x_t|\tilde{\mathcal{F}}_t]\|^{r'}], I\!E[\|\tilde{x}_t - I\!E[\tilde{x}_t|\mathcal{F}_t]\|^{r'}] \right\}^{\frac{1}{r'}} & , 1 \le r' < \infty \\ \max\left\{ \|x_t - I\!E[x_t|\tilde{\mathcal{F}}_t]\|_\infty, \|\tilde{x}_t - I\!E[\tilde{x}_t|\mathcal{F}_t]\|_\infty \right\} & , r' = \infty \end{cases}$$

and obtain for $1 \le r' < \infty$

$$\begin{aligned} D_t(\xi, \tilde{\xi})^{r'} &= \max\Big\{ \sum_{i=1}^N p_i \|x_t(\omega_i) - I\!E[x_t|\tilde{\mathcal{F}}_t](\omega_i)\|^{r'}, \sum_{i=1}^N p_i \|\tilde{x}_t(\omega_i) - I\!E[\tilde{x}_t|\mathcal{F}_t](\omega_i)\|^{r'} \Big\} \\ &= \max\Bigg\{ \sum_{s \in \tilde{\mathcal{I}}_t} \sum_{\omega_i \in \tilde{E}_{ts}} p_i \Bigg\| x_t(\omega_i) - \frac{\sum\limits_{\omega_j \in \tilde{E}_{ts}} p_j x_t(\omega_j)}{\sum\limits_{\omega_j \in \tilde{E}_{ts}} p_j} \Bigg\|^{r'}, \\ &\qquad \sum_{s \in \mathcal{I}_t} \sum_{\omega_i \in E_{ts}} p_i \Bigg\| \tilde{x}_t(\omega_i) - \frac{\sum\limits_{\omega_j \in E_{ts}} p_j \tilde{x}_t(\omega_j)}{\sum\limits_{\omega_j \in E_{ts}} p_j} \Bigg\|^{r'} \Bigg\} \end{aligned}$$

For $r' = \infty$ we have

$$\begin{aligned} D_t(\xi, \tilde{\xi}) &= \max\Big\{ \max_{i=1,\ldots,N} \|x_t(\omega_i) - I\!E[x_t|\tilde{\mathcal{F}}_t](\omega_i)\|, \max_{i=1,\ldots,N} \|\tilde{x}_t(\omega_i) - I\!E[\tilde{x}_t|\mathcal{F}_t](\omega_i)\| \Big\} \\ &= \max\Bigg\{ \max_{s \in \tilde{\mathcal{I}}_t} \max_{\omega_i \in \tilde{E}_{ts}} \Bigg\| x_t(\omega_i) - \frac{\sum\limits_{\omega_j \in \tilde{E}_{ts}} p_j x_t(\omega_j)}{\sum\limits_{\omega_j \in \tilde{E}_{ts}} p_j} \Bigg\|, \\ &\qquad \max_{s \in \mathcal{I}_t} \max_{\omega_i \in E_{ts}} \Bigg\| \tilde{x}_t(\omega_i) - \frac{\sum\limits_{\omega_j \in E_{ts}} p_j \tilde{x}_t(\omega_j)}{\sum\limits_{\omega_j \in E_{ts}} p_j} \Bigg\| \Bigg\}. \end{aligned}$$

Now, we return to the special case considered in this paper that ξ is a fan of individual scenarios $\{\xi^i : i = 1, \ldots, N\}$ and $\tilde{\xi} = \tilde{\xi}_{\text{tr}}$ a scenario tree with I_T scenarios. Hence, we have $\mathcal{F}_t = \mathcal{F}$ for $t = 2, \ldots, T$ and the second item of the maximum defining $D_t(\mathcal{F}_t, \tilde{\mathcal{F}}_t)^{r'}$ vanishes. Using the notation of Section 4 we denote by I_t again the index set of realizations of the scenario tree $\tilde{\xi}_{\text{tr}}$ at time t, by $\bar{I}_{t,i} = \{i\} \cup I_{t,i}$, $i \in I_t$ the scenario clusters at t, by π_t^i the (node) probability of $\tilde{\xi}_t^i$, i.e., $\pi_t^i = \sum_{j \in \bar{I}_{t,i}} p_j$, and by p_j the probability of scenario ξ^j for $j = 1, \ldots, N$. Since $\omega_j \in \tilde{E}_{ts}$ is equivalent to $j \in \bar{I}_{t,s}$, we obtain

$$D_t(\xi, \tilde{\xi})^{r'} = \sum_{i \in I_t} \sum_{j \in \bar{I}_{t,i}} p_j \Big\| x_t^j - \frac{1}{\pi_t^i} \sum_{k \in \bar{I}_{t,i}} p_k x_t^k \Big\|^{r'} \; (1 \le r' < \infty) \tag{31}$$

$$D_t(\xi,\tilde{\xi}) \;=\; \max_{i\in I_t}\max_{j\in \bar{I}_{t,i}} \Big\| x_t^j - \frac{1}{\pi_t^i}\sum_{k\in \bar{I}_{t,i}} p_k x_t^k \Big\| \quad (r'=\infty), \tag{32}$$

where x_t^i, $i=1,\ldots,N$, are the t-th components of the solution scenarios of the two-stage models with input ξ (having scenarios ξ^j, $j=1,\ldots,N$). Starting from (31), (32) the following estimates are valid.

Theorem 4.7 *Let (A1) and (A2) be satisfied. Let the stochastic process ξ have scenarios ξ^i with probabilities p_i, $i=1,\ldots,N$, and $\tilde{\xi}_{\rm tr}$ be a scenario tree with index set I_t of realizations and scenario clusters $\bar{I}_{t,i}$ at t. Then the distance of their filtrations allows the estimates*

$$D_{\rm f}(\xi,\tilde{\xi}) \;\le\; \begin{cases} \sum\limits_{t=2}^{T-1}\Big(\sum\limits_{i\in I_t}\sum\limits_{j\in \bar{I}_{t,i}} p_j(\pi_t^i)^{r'-1}\max\limits_{k\in \bar{I}_{t,i}}\|x_t^k - x_t^j\|^{r'}\Big)^{\frac{1}{r'}} & ,\ 1\le r'<\infty \\ \sum\limits_{t=2}^{T-1}\max\limits_{i\in I_t}\max\limits_{j,k\in \bar{I}_{t,i}}\|x_t^k - x_t^j\| & ,\ r'=\infty \end{cases} \tag{33}$$

$$D_{\rm f}(\xi,\tilde{\xi}) \;\le\; K\begin{cases} \sum\limits_{t=2}^{T-1}\Big(\sum\limits_{i\in I_t}\sum\limits_{j\in I_{t,i}} p_j\|x_t^j - x_t^i\|^{r'}\Big)^{\frac{1}{r'}} & ,\ 1\le r'<\infty \\ \sum\limits_{t=2}^{T-1}\max\limits_{i\in I_t}\max\limits_{j\in I_{t,i}}\|x_t^j - x_t^i\| & ,\ r'=\infty \end{cases} \tag{34}$$

for any solution x of (11) with input ξ and some constant $K>0$.

Proof: Let x be a solution of (11) with input ξ, which exists according to (A1) and (A2). The proof is carried out for the case $1\le r'<\infty$. In case $r'=\infty$ the estimates follow by immediate modifications. To derive (33), we start from (31) and obtain

$$\begin{aligned} D_t(\xi,\tilde{\xi})^{r'} &= \sum_{i\in I_t}\sum_{j\in \bar{I}_{t,i}} \frac{p_j}{\pi_t^i}\Big\|\sum_{k\in \bar{I}_{t,i}} p_k(x_t^k - x_t^j)\Big\|^{r'} \\ &\le \sum_{i\in I_t}\sum_{j\in \bar{I}_{t,i}} \frac{p_j}{\pi_t^i}\Big(\sum_{k\in \bar{I}_{t,i}} p_k\|x_t^k - x_t^j\|\Big)^{r'} \\ &\le \sum_{i\in I_t}\sum_{j\in \bar{I}_{t,i}} \frac{p_j}{\pi_t^i}(\pi_t^i)^{r'}\max_{k\in \bar{I}_{t,i}}\|x_t^k - x_t^j\|^{r'} \end{aligned}$$

For the second estimate we consider for any $i\in I_t$ the index $\alpha_t(i)$ defined by (20) and (25), respectively. Starting again from (31) we get

$$\begin{aligned} D_t(\xi,\tilde{\xi})^{r'} &\le \sum_{i\in I_t}\sum_{j\in \bar{I}_{t,i}} p_j\Big(\|x_t^j - x_t^{\alpha_t(i)}\| + \sum_{k\in \bar{I}_{t,i}}\frac{p_k}{\pi_t^i}\|x_t^{\alpha_t(i)} - x_t^k\|\Big)^{r'} \\ &\le \bar{K}\sum_{i\in I_t}\sum_{j\in \bar{I}_{t,i}} p_j\Big(\|x_t^j - x_t^{\alpha_t(i)}\|^{r'} + \sum_{k\in \bar{I}_{t,i}}\Big(\frac{p_k}{\pi_t^i}\Big)^{r'}\|x_t^{\alpha_t(i)} - x_t^k\|^{r'}\Big) \\ &\le \bar{K}\sum_{i\in I_t}\Big(\sum_{j\in \bar{I}_{t,i}} p_j\|x_t^j - x_t^{\alpha_t(i)}\|^{r'} + \pi_t^i\frac{(\pi_t^i)^{r'-1}}{(\pi_t^i)^{r'}}\sum_{k\in \bar{I}_{t,i}} p_k\|x_t^{\alpha_t(i)} - x_t^k\|^{r'}\Big) \end{aligned}$$

$$
\begin{aligned}
&= 2\bar{K}\sum_{i\in I_t}\sum_{j\in \bar{I}_{t,i}} p_j\|x_t^j - x_t^{\alpha_t(i)}\|^{r'}\\
&= 2\bar{K}\sum_{i\in I_t}\sum_{j\in I_{t,i}} p_j\|x_t^j - x_t^{i}\|^{r'},
\end{aligned}
$$

where the identity $\alpha_t(i) = i$ for each $i \in I_t$ is used in the final step and $\bar{K} > 0$ is some constant depending on r' and the maximum of the cardinalities of $I_{t,i}$. □

The upper bound on the filtration distance contains only nontrivial summands at pairs (t, i), where $J_{t,i} \neq \emptyset$ or, in other words, where scenario i branches at time t. The relevant contribution of such pairs (t, i) to the filtration bound amounts to

$$
b_{t,i} := \begin{cases} \sum\limits_{j\in I_{t,i}} p_j\|x_t^j - x_t^i\|^{r'} & , 1 \le r' < \infty \\ \max\limits_{j\in I_{t,i}} \|x_t^j - x_t^i\| & , r' = \infty \end{cases} \tag{35}
$$

Hence, the term (35) has to be controlled during the whole (backward or forward) tree construction process. Roughly speaking, this means that scenario i may branch at t only if $b_{t,i}$ is sufficiently small. More precisely, both Algorithms 4.2 and 4.5 should be extended by incorporating the condition

$$
B_t := \begin{cases} \sum\limits_{i\in I_t} b_{t,i} \le \varepsilon_{\mathrm{f},t}^{r'} & , 1 \le r' < \infty \\ \max\limits_{i\in I_t} b_{t,i} \le \varepsilon_{\mathrm{f},t} & , r' = \infty \end{cases} \tag{36}
$$

with some filtration tolerance $\varepsilon_{\mathrm{f},t}$ into their step t for $t = 2, \ldots, T-1$. Unfortunately, the solution process $\{x_t\}_{t=1}^T$ of the two-stage model is hardly available in general and only available at certain extra cost. Hence, reducing scenarios of the fan ξ with respect to $\|\cdot\|_r$ (cf. Section 2), computing the solution of (11) with reduced input fan and estimating the bounds B_t might be a suitable alternative.

Nevertheless, it is of considerable interest to derive bounds on the filtration distance that are based on input information only. This requires to derive estimates for the distance of any two scenarios of some solution x of (11) with input ξ. To this end, we assume that only costs and right-hand sides are random in (11). We consider the scenario-based stochastic program with input ξ

$$
\min\left\{ \langle b_1(\xi_1^*), x_1\rangle + \sum_{i=1}^N p_i \sum_{t=2}^T \langle b_t(\xi_t^i), x_t^i\rangle \;\middle|\; \begin{array}{l} x_t^i \in X_t, t = 1, \ldots, T, \\ A_{t,0}x_t^i + A_{t,1}x_{t-1}^i = h_t(\xi_t^i), \\ t = 2, \ldots, T, i = 1, \ldots, N \end{array} \right\}, \tag{37}
$$

which is indeed two-stage and represents a linear program. Furthermore, the minimization decomposes into first- and second-stage variables leading to the following form of the two-stage program (37)

$$
\min\Big\{ \langle b_1(\xi_1^*), x_1\rangle + \sum_{i=1}^N p_i \Phi(\xi^i, x_1) \mid x_1 \in X_1 \Big\}, \tag{38}
$$

where the optimal value function Φ is extended real-valued and defined on $\Xi \times X_1$ by

$$\Phi(z, x_1) := \inf \Big\{ \sum_{t=2}^{T} \langle b_t(z_t), x_t \rangle \mid x_t \in X_t, A_{t,0}x_t + A_{t,1}x_{t-1} = h_t(z_t), t = 2, \ldots, T \Big\} \quad (39)$$

for any pair $(z, x_1) \in \Xi \times X_1$. Exploiting Lipschitz stability properties of solutions to the linear program on the right-hand side of (39) together with ideas of the proof of Theorem 4.7 leads to the following estimate of the filtration distance.

Theorem 4.8 *Assume that only costs and right-hand sides are random in (11) and that (A1) and (A2) are satisfied. Then there exists a constant $\hat{L} > 0$ such that the filtration distance allows the estimate*

$$D_{\mathrm{f}}(\xi, \tilde{\xi}) \leq \hat{L} \begin{cases} \Big(\sum_{i \in I_2} \sum_{j \in I_{2,i}} p_j \|\xi^j - \xi^i\|^{r'} \Big)^{\frac{1}{r'}} & , 1 \leq r' < \infty \\ \max_{i \in I_2} \max_{j \in I_{2,i}} \|\xi^j - \xi^i\| & , r' = \infty \, . \end{cases} \quad (40)$$

Proof: Due to (A1) and (A2) there exists a solution x^* of (11) with input ξ. Let us consider the parametric linear program

$$\min \Big\{ \sum_{t=2}^{T} \langle b_t(z_t), x_t \rangle \mid x_t \in X_t,\ A_{t,0}x_t + A_{t,1}x_{t-1} = h_t(z_t),\ t = 2, \ldots, T \Big\}, \quad (41)$$

with parameter $z \in \Xi$, where $x_1 = x_1^*$ is fixed. Let $S(z)$ denote the solution set of the linear program (41). Conditions (A1) and (A2) imply that $S(\xi^j)$ is nonempty for every $j = 1, \ldots, N$. Since the functions $b_t(\cdot)$ and $h_t(\cdot)$, $t = 2, \ldots, N$, are affinely linear, $\operatorname{dom} S$ is convex polyhedral (cf. [35]) and the multifunction S is polyhedral (cf. [28]). Hence, S is locally upper Lipschitz continuous at each $z \in \operatorname{dom} S$ [28, Proposition 1] and, thus, it is upper Lipschitz continuous on each bounded subset B of $\operatorname{dom} S$. Setting $B := \operatorname{conv}\{\xi^1, \ldots, \xi^N\}$ there exists a constant $\bar{L} > 0$ such that

$$S(\xi) \subseteq S(\xi^j) + \bar{L}\|\xi - \xi^j\| \quad (42)$$

holds for all $\xi \in B$ and $j = 1, \ldots, N$. For each $i \in I_2$ and some $x^i \in S(\xi^i)$ we select elements $x^j \in S(\xi^j)$ according to (42) such that

$$\|x^i - x^j\| \leq \bar{L}\|\xi^i - \xi^j\|$$

for each $j \in \bar{I}_{2,i}$. Then, the stochastic process x having scenarios $(x_1^*, x_2^i, \ldots, x_T^i)$ with probabilities p_i, $i = 1, \ldots, N$, is a solution of (11) with input ξ. With the solution x at hand we are ready to estimate the filtration distance by using a similar idea as in the proof of Theorem 4.7. This time we use the indices $\alpha_2(i)$ instead of $\alpha_t(i)$ and obtain in case $1 \leq r' < \infty$ that

$$D_{\mathrm{f}}(\xi, \tilde{\xi}) \quad \leq \quad \hat{K} \Big(\sum_{t=2}^{T-1} D_t(\xi, \tilde{\xi})^{r'} \Big)^{\frac{1}{r'}}$$

$$\leq \hat{K}\bar{K}\Big(\sum_{t=2}^{T-1}\sum_{i\in I_t}\sum_{j\in \bar{I}_{t,i}} p_j \|x_t^j - x_t^{\alpha_2(i)}\|^{r'}\Big)^{\frac{1}{r'}}$$

$$= \hat{K}\bar{K}\Big(\sum_{t=2}^{T-1}\sum_{i\in I_2}\sum_{j\in \bar{I}_{2,i}} p_j \|x_t^j - x_t^{i}\|^{r'}\Big)^{\frac{1}{r'}}$$

$$= \hat{K}\bar{K}\Big(\sum_{i\in I_2}\sum_{j\in \bar{I}_{2,i}} p_j \|x^j - x^{i}\|^{r'}\Big)^{\frac{1}{r'}}$$

$$\leq \hat{K}\bar{K}\bar{L}\Big(\sum_{i\in I_2}\sum_{j\in I_{2,i}} p_j \|\xi^j - \xi^{i}\|^{r'}\Big)^{\frac{1}{r'}},$$

where $\hat{K} := (T-2)^{\frac{1}{r}}$ and $\bar{K}$ is the same constant as in the proof of Theorem 4.7. Setting $\hat{L} := \hat{K}\bar{K}\bar{L}$ completes the proof for $1 \leq r' < \infty$. Immediate modifications of the above proof lead to the desired estimate for $r' = \infty$. □

The theorem supports one of the standard conjectures in scenario tree generation, namely, that the first time period after the deterministic first stage plays a major role. The estimate (40) advises that every cluster $\bar{I}_{2,i}$ has to be chosen such that the term

$$b_{2,i}^* := \begin{cases} \sum\limits_{j\in I_{2,i}} p_j \|\xi^j - \xi^i\|^{r'} & , 1 \leq r' < \infty \\ \max\limits_{j\in I_{2,i}} \|\xi^j - \xi^i\| & , r' = \infty \end{cases} \tag{43}$$

is small enough. In many practical cases this condition will imply that the cardinality of $I_{2,i}$ remains relatively small and that of I_2 large. Notice that the distance of scenarios is measured with respect to the whole time horizon. The latter fact represents the main difference to the estimates of the L_r-distance in the Theorems 4.1 and 4.4. This means that scenario $i \in I_2$ should admit branching at $t = 2$ only if the distance of ξ^i and each scenario ξ^j that branches from i, i.e., $j \in I_{2,i}$, is not too large, i.e., $b_{2,i}^*$ defined in (43) is small. More precisely, both Algorithms 4.2 and 4.5 should be modified such that clusters $I_{2,i}$, $i \in I_2$, are prescribed at $t = 2$ satisfying the condition

$$B_2^* := \begin{cases} \sum\limits_{i\in I_2} b_{2,i}^* \leq \varepsilon_{\mathrm{f}}^{r'} & , 1 \leq r' < \infty \\ \max\limits_{i\in I_2} b_{2,i}^* \leq \varepsilon_{\mathrm{f}} & , r' = \infty \end{cases} \tag{44}$$

for some filtration tolerance ε_{f}. If (44) is satisfied, the further branching behavior at time periods t with $2 < t \leq T$ is controlled via the existing tests in both algorithms. Finally, we note that estimating the filtration distance via (33) in Theorem 4.7 seems to be preferable whenever a two-stage solution or a reduced version of it, i.e., an (approximate) solution of the two-stage approximation to the original problem, is available at reasonable costs.

5 Numerical experience

The Algorithms 4.2 and 4.5 have been tested on data provided by the French company Electricité de France (EDF). The data contain a finite number of scenarios representing realizations of a bivariate stochastic process whose components are electrical load and water inflow (i.e., right-hand sides of linear constraints) for a time horizon of two years. The time horizon was discretized with three time steps per day, where each time step is associated with a set of hours during which the demand does not change much. Table 1 and 2 show the discretization of the data for the time horizon of two years and provide the number of scenarios, the total number of time periods and the corresponding number of nodes of the initial scenario fan. The first node (root node) corresponds to the mean value of all scenarios at time period $t = 1$. The weekly amounts of water inflows were uniformly distributed to the corresponding time steps of the week.

Random variable	Discretization	Number time steps
electrical load	3 per day	2 184
water inflow	weekly	104

Table 1: Discretization of the two-year time horizon

	Number
scenarios	456
time periods	2 184
initial nodes	995 449

Table 2: Dimension of the initial scenario fan

To test the scenario tree construction approach, we performed test series for the Algorithms 4.2 and 4.5 to generate scenario trees such that branching is allowed at all time steps, and branching is only allowed at the beginning of a week, respectively. To measure the distances between the original and approximate probability distributions $r = 1$ and a relative tolerance $\varepsilon_{rel} := \frac{\varepsilon}{\varepsilon_{max}}$ were used in all test runs. Here, ε_{max} denotes the best possible distance between the probability distribution of the initial scenario fan and the distribution of one of its scenarios endowed with unit mass. All test runs were performed on a PC with a 3 GHz Intel Pentium CPU and 1 GByte main memory.

5.1 Results of backward tree construction

For the backward variant of scenario tree construction individual tolerances ε_t at branching points were chosen recursively such that

$$\varepsilon_T = \varepsilon \cdot (1-q), \quad q \in (0,1) \quad \text{and} \quad \varepsilon_t = q \cdot \varepsilon_{t+1}, \quad t = T-1, \ldots, 2. \tag{45}$$

According to our numerical experience a choice of $q \in (0,1)$ closer to 1 leads to a higher number of remaining scenarios and branching points (stages). Choosing q closer to 0 leads to the opposite effect. For the test runs of Algorithm 4.2 we used $q = 0.95$. The Tables 3 and 4 display the numerical results for the test series and different relative tolerances.

The second and third columns compare the sizes of the initial scenario fan and the constructed scenario tree in terms of the numbers of scenarios and nodes, respectively. The last but one column contains the number of stages, i.e., the number of time periods

ε_{rel}	Scenarios	Nodes	Stages	Time (sec)
0.10	442	584 270	151	172.86
0.20	429	371 046	150	129.11
0.30	417	268 201	146	117.42
0.40	405	193 014	135	110.83
0.50	393	140 536	115	106.30

Table 3: Results for backward tree construction without branching restriction

ε_{rel}	Scenarios	Nodes	Stages	Time (sec)
0.10	442	589 575	88	118.47
0.20	429	397 047	83	110.65
0.30	416	293 403	86	108.40
0.40	405	219 714	83	106.15
0.50	393	170 520	81	105.16

Table 4: Results for backward tree construction with weekly branching restriction

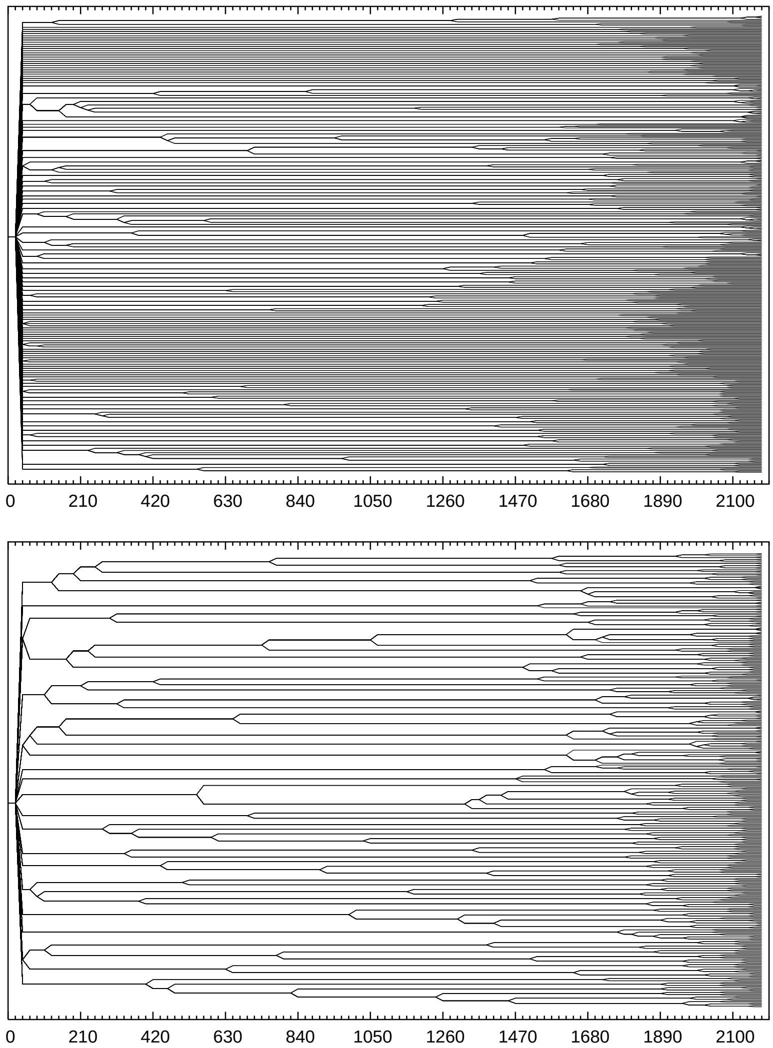

Figure 4: Scenario trees obtained with $\varepsilon_{rel} = 0.2/0.5$ and weekly branching structure

where branching occurs. The computing times for constructing the trees can be found in the last column. The computing time already contains the CPU time of (about) 100 seconds for computing the distances of scenarios, which are needed in all test runs.

It turns out that for a small relative tolerance an approximate scenario tree that contains only 50% of the original nodes can be constructed. The pictures of Figure 4 show the structure of two generated scenario trees with weekly branching structure and epsilon tolerances $\varepsilon_{rel} = 0.2$ and $\varepsilon_{rel} = 0.5$, respectively.

5.2 Results of forward tree construction

In a second series of tests, scenario trees were constructed out of the EDF data by Algorithm 4.5. In case of forward tree construction, individual tolerances ε_t at branching points were chosen such that

$$\varepsilon_t \;=\; \frac{\varepsilon}{T}\left[1+\overline{q}\left(\frac{1}{2}-\frac{t}{T}\right)\right], \quad t = 2, \ldots, T, \tag{46}$$

where $\overline{q} \in [0, 1]$ is a parameter that affects the branching structure of the constructed trees very similar to q in case of backward reduction. For the test runs we used $\overline{q} = 0.6$.

The Tables 5 and 6 provide numerical results for Algorithm 4.5. Just as before, the tables correspond to the series of tests, i.e., the first one contains results for trees without branching restriction and the second one by allowing branching only at the beginning of a week.

The numerical results illustrate that the forward variant of scenario tree construction performs as well as the backward version. Nevertheless, a comparison discloses noticeable differences. Namely, it turns out that, for small relative tolerances, the trees obtained by backward tree construction contain less nodes than in the forward case. For increasing relative tolerances the forward construction algorithm provides trees containing (much) less nodes than the backward counterpart. This is due to the fact that the error estimate (21) in Section 4.1 is derived by employing the triangle inequality extensively and, hence, is more pessimistic than (28).

Figure 5 shows the generated scenario trees with weekly branching structure for $\varepsilon_{rel} = 0.4$ and $\varepsilon_{rel} = 0.5$. For these trees it turns out that about 15% of all nodes are sufficient to guarantee 60% accuracy, while 6% of the nodes still guarantee 50% accuracy.

5.3 Tree construction and filtration distances

Finally, we want to discuss the effects of incorporating the filtration distance into the scenario tree construction approach. Here, we concentrate on the situation that only the input information is available and, hence, the upper bound (40) of the filtration distance with $r' = 1$ is employed. In our implementation, we modified Algorithm 4.5 of forward tree construction such that instead of Step 2 sets I_2 and $I_{2,i}$, $i \in I_2$, are determined which satisfy the condition (44), i.e.,

$$\sum_{i\in I_2}\sum_{j\in I_{2,i}} p_j \|\xi^j - \xi^i\| \leq \varepsilon_{\mathrm{f}}. \tag{47}$$

ε_{rel}	Scenarios	Nodes	Stages	Time (sec)
0.10	378	743 087	129	108.11
0.20	305	529 994	162	109.15
0.30	216	289 324	161	114.18
0.40	145	138 175	121	134.11
0.50	93	67 696	84	202.42

Table 5: Results for forward tree construction without branching restriction

ε_{rel}	Scenarios	Nodes	Stages	Time (sec)
0.10	389	746 613	49	106.53
0.20	300	509 103	57	106.84
0.30	228	310 653	64	107.59
0.40	163	151 809	69	109.78
0.50	92	60 501	46	119.12

Table 6: Results for forward tree construction with weekly branching restriction

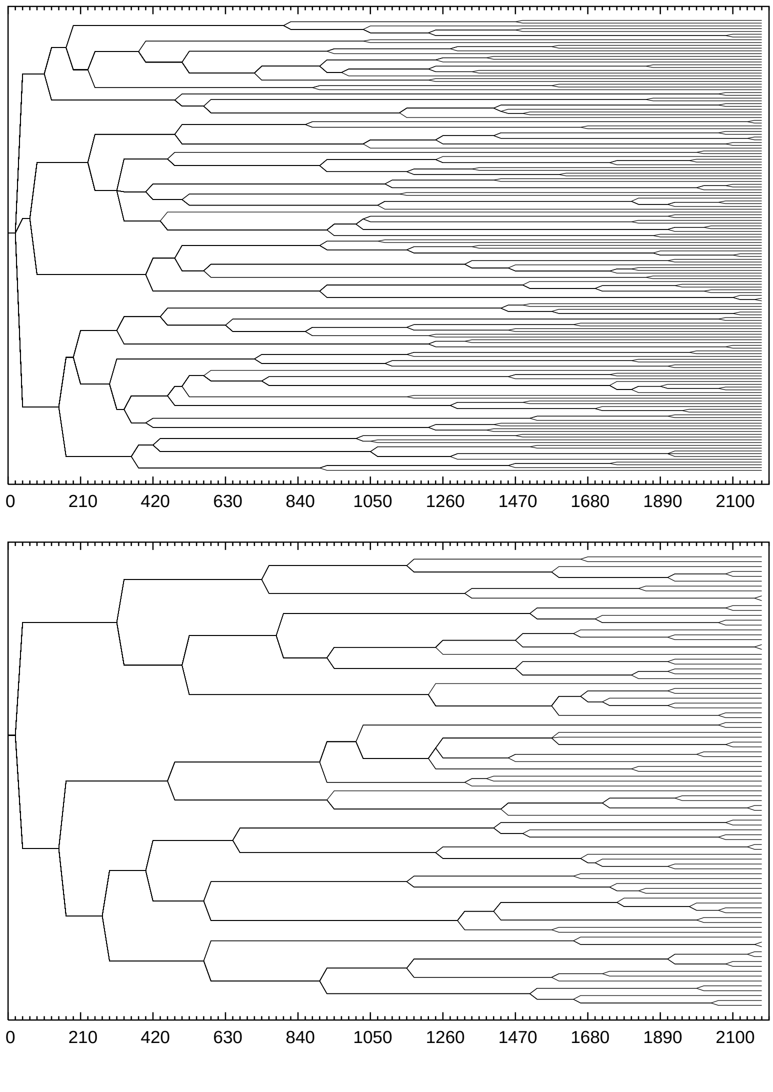

Figure 5: Scenario trees obtained with $\varepsilon_{rel} = 0.4/0.5$ and weekly branching structure

ε_{rel}	$\varepsilon_{\mathrm{f},rel}$	Scenarios	Nodes	Stages	Time (sec)
0.10	0.15	299	588 489	33	104.26
0.10	0.20	304	592 112	36	104.11
0.20	0.25	168	281 456	21	104.17
0.20	0.30	180	297 564	42	103.96
0.30	0.35	75	144 009	7	104.26
0.30	0.40	106	138 899	31	104.27
0.40	0.45	43	71 915	6	104.99
0.40	0.50	68	68 581	28	105.22
0.50	0.55	24	29 305	9	107.84
0.50	0.60	38	29 496	22	109.31

Table 7: Results of forward tree construction with incorporated filtration distance

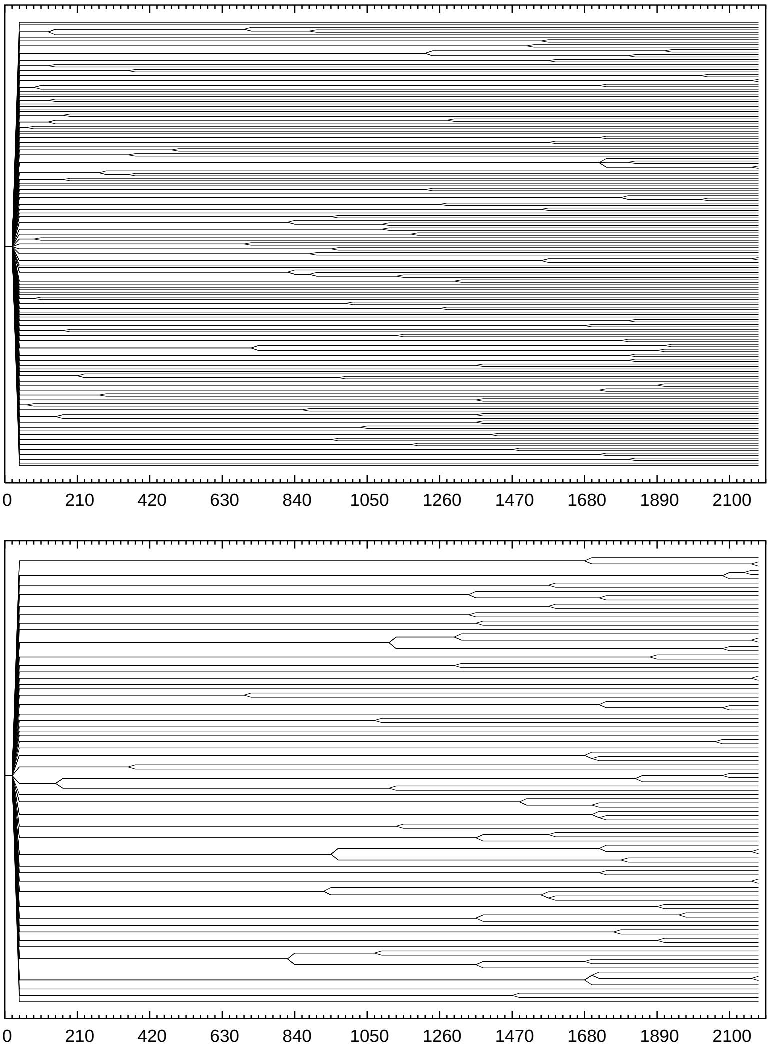

Figure 6: Scenario trees obtained with $\varepsilon_{rel}/\varepsilon_{\mathrm{f},rel} = 0.2/0.3$, and $\varepsilon_{rel}/\varepsilon_{\mathrm{f},rel} = 0.3/0.4$

Here, the filtration tolerance ε_{f} is selected to be greater than the current total tolerance ε of the forward tree construction Algorithm 4.5. Table 7 displays numerical results for different tolerances ε and ε_{f} of $\|\xi - \tilde{\xi}_{\mathrm{tr}}\|_r$ and $D_{\mathrm{f}}(\xi, \tilde{\xi}_{\mathrm{tr}})$, respectively. In both cases we used the relative tolerances $\varepsilon_{rel} = \frac{\varepsilon}{\varepsilon_{max}}$ (first column) and $\varepsilon_{\mathrm{f},rel} = \frac{\varepsilon_{\mathrm{f}}}{\varepsilon_{max}}$ (second column). The individual tolerances ε_t of the construction method were chosen by formula (46) with parameter $\bar{q} = 0.4$. Figure 6 shows the influence of the (relative) filtration tolerance on the tree structure. A comparison with Figure 5 shows that the cardinality of I_2 is clearly larger in Figure 6. This effect is due to imposing condition (47) at $t = 2$. Condition (47) also leads to a smaller number of branchings and stages.

6 Conclusions

In many applications of stochastic programming the available statistical data allows to generate a large fan of individual scenarios including their probabilities, which is considered as a good representation of the underlying stochastic process. In this paper, we developed a methodology for constructing scenario trees out of a fan of individual scenarios such that the probability distribution and the filtration structure of the original stochastic process is approximately recovered. The approximation quality is measured in terms of an L_r-distance for the input distributions and of a distance for the input filtrations. The latter is given as an $L_{r'}$-distance of optimal solutions and their conditional expectations. The use of the two different measures is advised by a stability result for multistage stochastic programs, which also suggests the choice of r and r' according to structural properties of the optimization model. Algorithms are developed and implemented that allow to construct scenario trees such that both distance measures are satisfied relative to presribed tolerances. Some computational experience is provided and discussed. Further test runs, an evaluation of the algorithms and a description of the implementation is projected for a subsequent paper.

Acknowledgement:

This work was supported by the DFG Research Center MATHEON "Mathematics for key technologies" in Berlin, the BMBF under the grants 03ROM5B3 and 03SF0312E, and a grant of EDF – Electricité de France. The authors wish to thank their colleague Nicole Gröwe-Kuska (Humboldt-University Berlin until June 2003) for the collaboration at earlier stages of this work. They extend their gratitude to C. Strugarek, J.-S. Roy, F. Turboult and other members of the OSIRIS Division at R&D of EDF for several stimulating discussions.

References

[1] Casey, M.; Sen, S.: The scenario generation algorithm for multistage stochastic linear programming, *Mathematics of Operations Research* (to appear).

[2] Chiralaksanakul, A.; Morton, D.P.: Assessing policy quality in multi-stage stochastic programming, *Stochastic Programming E-Print Series* 12–2004 (<speps.org>).

[3] Consigli, G.; Dempster, M.A.H.: Dynamic stochastic programming for dynamic asset-liability management, *Annals of Operations Research* 81 (1998), 131–161.

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

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

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

[7] Evstigneev, I.: Measurable selection and dynamic programming, *Mathematics of Operations Research* 1 (1976) 267–272.

[8] Frauendorfer, K.: Barycentric scenario trees in convex multistage stochastic programming, *Mathematical Programming* Ser. B, 75 (1996), 277–293.

[9] Givens, C.R.; Shortt, R.M.: A class of Wasserstein metrics for probability distributions, *Michigan Mathematical Journal* 31 (1984), 231–240.

[10] Gröwe-Kuska, N.; Heitsch, H.; Römisch, W.: Modellierung stochastischer Datenprozesse für Optimierungsmodelle der Energiewirtschaft, *IT-Lösungen für die Energiewirtschaft in liberalisierten Märkten*, VDI-Berichte 1647, VDI-Verlag, Düsseldorf 2001, 69–78.

[11] Gröwe-Kuska, N.; Heitsch, H.; Römisch, W.: Scenario reduction and scenario tree construction for power management problems, *IEEE Bologna Power Tech Proceedings* (Borghetti, A., Nucci, C. A., Paolone, M. eds.), 2003.

[12] Gröwe-Kuska, N.; Römisch, W.: Stochastic unit commitment in hydro-thermal power production planning, Chapter 30 in *Applications of Stochastic Programming* (Wallace, S. W., Ziemba W. T., eds.), MPS-SIAM Series in Optimization, 2005.

[13] Heitsch, H.; Römisch, W.: Scenario reduction algorithms in stochastic programming, *Computational Optimization and Applications* 24 (2003), 187–206.

[14] Heitsch, H.; Römisch, W.: Generation of multivariate scenario trees to model stochasticity in power management, *IEEE St. Petersburg Power Tech* 2005.

[15] Heitsch, H.; Römisch, W.; Strugarek, C.: Stability of multistage stochastic programs, Preprint 255, DFG Research Center Matheon "Mathematics for key technologies", 2005 and submitted to SIAM Journal on Optimization.

[16] Hochreiter, R.; Pflug, G. Ch.: Financial scenario generation for stochastic multi-stage decision processes as facility location problem, *Annals of Operations Research* (to appear).

[17] Høyland, K.; Wallace, S. W.: Generating scenario trees for multi-stage decision problems, *Management Science* **47** (2001), 295–307.

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

[19] Kaut, M.; Wallace, S.W.: Evaluation of scenario-generation methods for stochastic programming, *Stochastic Programming E-Print Series* 14-2003 (`<speps.org>`).

[20] Möller, A.; Römisch, W.; Weber, K.: A new approach to O&D revenue management based on scenario trees, *Journal of Revenue and Pricing Management* 3 (2004), 265–276.

[21] Pennanen, T.: Epi-convergent discretizations of multistage stochastic programs via integration quadratures, *Stochastic Programming E-Print Series* 19–2004 (`<speps.org>`).

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

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

[24] Rachev, S. T.: *Probability Metrics and the Stability of Stochastic Models*, Wiley, 1991.

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

[26] Rachev, S. T.; Rüschendorf, L.: *Mass Transportation Problems*, Vol. I and II, Springer, Berlin 1998.

[27] Rachev, S. T.; Schief, A.: On L_p-minimal metrics, *Probability and Mathematical Statistics* 13 (1992), 311–320.

[28] Robinson, S. M.: Some continuity properties of polyhedral multifunctions, *Mathematical Programming Study* 14 (1981), 206–214.

[29] Rockafellar, R. T.; Wets, R. J-B: *Variational Analysis*, Springer-Verlag, Berlin, 1998.

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

[31] Römisch, W.; Schultz, R.: Stability analysis for stochastic programs, *Annals of Operations Research* 30 (1991), 241–266.

[32] Ruszczyński, A.; Shapiro, A. (Eds.): *Stochastic Programming*, Handbooks in Operations Research and Management Science, Volume 10, Elsevier, Amsterdam 2003.

[33] Shapiro, A.: Inference of statistical bounds for multistage stochastic programming problems, *Math. Meth. Oper. Res.* 58 (2003), 57–68.

[34] Shapiro, A.: On complexity of multistage stochastic programs, *Operations Research Letters* (to appear).

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