Math. Program., Ser. A
DOI 10.1007/s10107-007-0197-2

FULL LENGTH PAPER

Scenario tree modeling for multistage stochastic programs

Holger Heitsch · Werner Römisch

Received: 23 November 2005 / Accepted: 17 September 2007

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, we develop (stability) theory-based heuristics for generating scenario trees out of an initial set of scenarios. They are based on forward or backward algorithms for tree generation consisting of recursive scenario reduction and bundling steps. Conditions are established implying closeness of optimal values of the original process and its tree approximation, respectively, by relying on a recent stability result in Heitsch, Römisch and Strugarek (SIAM J Optim 17:511–525, 2006) for multistage stochastic programs. Numerical experience is reported for constructing multivariate scenario trees in electricity portfolio management.

Keywords Stochastic programming · Multistage · Stability · L_r-distance · Filtration · Scenario tree · Scenario reduction

Mathematics Subject Classification (2000) 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 $(\Omega, \mathcal{F}, \mathbb{P})$ and with ξ_t taking values in some $\mathbb{R}^d$. The (stochastic) decision x_t at t maps from Ω to $\mathbb{R}^{m_t}$ and is assumed to be *nonanticipative*, i.e., to depend only

H. Heitsch · W. Römisch (✉)
Humboldt-University Berlin, Institute of Mathematics, 10099 Berlin, Germany
e-mail: romisch@math.hu-berlin.de

on $(\xi_1, \dots, \xi_t)$. This property is equivalent to the measurability of x_t with respect to the σ-field $\mathcal{F}_t(\xi) \subseteq \mathcal{F}$, which is generated by $\xi^t := (\xi_1, \dots, \xi_t)$. Clearly, we have $\mathcal{F}_t(\xi) \subseteq \mathcal{F}_{t+1}(\xi)$ for $t = 1, \dots, T-1$. Since at time $t = 1$ the input is known, we assume that $\mathcal{F}_1(\xi) = \{\emptyset, \Omega\}$.

The multiperiod stochastic program is assumed to be of the form

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

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 $\mathbb{R}^{n_t}$, $A_{t,0} \in \mathbb{R}^{n_t \times m_t}$ are fixed recourse matrices and $A_{t,1}(\xi_t) \in \mathbb{R}^{n_t \times m_{t-1}}$ technology matrices, respectively. We assume that costs $b_t(\cdot)$, right-hand sides $h_t(\cdot)$ and technology matrices $A_{t,1}(\cdot)$ depend affinely 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. Note that the two constraints $x_t \in X_t$ and $A_{t,0}x_t + A_{t,1}(\xi_t)x_{t-1} = h_t(\xi_t)$ mean $x_t(\omega) \in X_t$ and $A_{t,0}x_t(\omega) + A_{t,1}(\xi_t(\omega))x_{t-1}(\omega) = h_t(\xi_t(\omega))$ for $\mathbb{P}$-almost every $\omega \in \Omega$.

In addition to the pointwise constraint with probability 1, measurability, *filtration* or *information* constraints appear in (1). They are functional and non-pointwise at least if $T > 2$ and $\mathcal{F}_1(\xi) \subsetneqq \mathcal{F}_t(\xi) \subsetneqq \mathcal{F}_T(\xi)$ 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 [43, Chap. 3] for a recent survey.

Presently, there exist several approaches to generate scenario trees for multistage stochastic programs (see [8] for a survey of ideas and methods until 2000). They are based on several different principles. We mention here (1) bound-based constructions (e.g., [2,13,28]), (2) Monte Carlo-based schemes [3,44,45], (3) Quasi Monte Carlo-based discretization methods [30,31], (4) EVPI-based sampling and reduction within decomposition schemes [4,5,23], (5) the moment-matching principle [25–27], (6) probability metric based approximations [16,17,24,32]. Many of them require to prescribe the tree structure and offer different strategies for selecting scenarios.

In the present paper, we study and extend the scenario tree generation technique of [16,17]. Its idea is to start with a good initial approximation of the underlying stochastic input process ξ consisting of a process $\hat{\xi}$ having finitely many scenarios. These scenarios might be obtained by quantization techniques [15], by sampling or resampling techniques based on parametric or nonparametric stochastic models of ξ. Starting from $\hat{\xi}$, a process ξ_{tr} in tree form is constructed by deleting and bundling scenarios recursively. While the recursive method described in [16,17] works backward in time, a forward method was recently proposed in [20]. The main aims of the paper are (1) to derive error estimates for the L_r-distance $\|\hat{\xi} - \xi_{\text{tr}}\|_r$ of both (backward and forward) tree generation techniques, and (2) to provide justifications for both techniques by showing that the distance of optimal values $|v(\xi) - v(\xi_{tr})|$ of the underlying stochastic program (1) is small when additional conditions on $\hat{\xi}$ are imposed. The latter is attained by relying on recent stability results for multistage stochastic programs in [21,22]. The stability results enlighten that ξ has to be approximated in the sense of L_r and in terms of a filtration distance. Consequently, we argue that both distances get for ξ and ξ_{tr} if $\hat{\xi}$ represents a good approximation of ξ (in a sense which is made precise in Sect. 5). In this way, a (stability) theory-based heuristic for generating scenario trees out of a finite number of given scenarios is developed.

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 [29] and for load-price scenario trees in electricity portfolio management [10,18].

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 results of [21,22], which provides the theoretical basis of our tree constructions. Sections 4 and 5 contain the main results of our paper, in particular, the tree generation algorithms and their theoretical justification. In Sect. 6 we derive some estimates of the filtration distance and in Sect. 7 we discuss some numerical experience on backward and forward generation of load-inflow scenario trees based on realistic data.

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 [34,41]. In particular, distances given in terms of Monge–Kantorovich mass transportation problems became relevant. They are of the form

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

where Ξ is a closed subset of $\mathbb{R}^s$ (for some $s \in \mathbb{N}$), 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 [9,42], namely,

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

and

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

for some $r \geq 1$, $\xi_0 \in \Xi$ and a seminorm or a norm $|\cdot|$ in $\mathbb{R}^s$. In both cases, the set $\mathcal{P}_c(\Xi)$ may be chosen as the set $\mathcal{P}_r(\Xi)$ of all probability measures Q on Ξ such that

$\int_{\Xi} |\xi|^r Q(d\xi)$ is finite. The cost (3) leads to *L_r-minimal metrics* ℓ_r [36], which are defined by

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

and sometimes also called Wasserstein metrics of order r [14]. The mass transportation problem (2) with cost (4) defines the Monge–Kantorovich functional $\hat{\mu}_r$ [33,35]. 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 the condition that η is a finite measure on $\Xi \times \Xi$ with $\pi_1\eta - \pi_2\eta = P - Q$. 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 [33]. 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 rth order absolute moments.

For stochastic programs containing information constraints the situation is different. Examples (e.g., [22, Example 2.6]) show that a stability analysis based only on distances of probability distributions may fail. In the recent paper [22] 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 Sect. 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

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

Moreover, since an optimal solution $\eta^* \in \mathcal{P}(\Xi \times \Xi)$ of the mass transportation problem (5) always exists (cf. [33, 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., [33, 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 and probabilities p_i, $i = 1, \ldots, N$, and $\tilde{\xi}^j$ with probabilities q_j, $j = 1, \ldots, M$, respectively. Then we have

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

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 [9, Theorem 2] we conclude

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

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

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 $A_i := \xi^{-1}(\{\xi^i\}) \in \mathcal{F}$, and, thus, $\mathbb{P}(A_i) = p_i$, $i = 1, \ldots, N$. If the random vector ξ_J is defined by

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

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

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

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 any 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

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

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

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

and easily solvable. Their solutions $J = \{l^*\}$ and $J = \{1, \ldots, N\}\backslash\{u^*\}$ 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]: $u_i \in \arg \min_{u\in J^{[i-1]}} \sum_{k\in J^{[i-1]}\backslash\{u\}} p_k \min_{j\notin J^{[i-1]}\backslash\{u\}} |\xi^k - \xi^j|^r,$

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

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

These heuristics were studied in [19] 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 [19] is very encouraging.

3 Stability of multistage models

Here, we record the main results of the recent papers [21,22]. We assume that the stochastic input process ξ belongs to the Banach space $L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ with $s := Td$ and $r \geq 1$. The multistage model (1) is regarded as an optimization problem in the space $L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ with $m = \sum_{t=1}^{T} m_t$ and endowed with the norm

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

where the number r' is defined by

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

The choice of r and the definition of r' are motivated by the knowledge on existing moments of the input process, by having the stochastic program well defined (in particular, such that $\langle b_t(\xi_t), x_t \rangle$ is integrable for every decision x and $t = 1, \ldots, T$) and by satisfying the conditions (A2) and (A3). It is shown in [22] that the decisions belong to L_r (i.e., $r' = r$) if $\xi \in L_r$ enters right-hand sides. If ξ enters the costs, in addition, the condition $\frac{1}{r} + \frac{1}{r'} = 1$ implies $r' = r = 2$ and the finiteness of the objective. If either right-hand sides or costs are random, it is sufficient to require $r \geq 1$ and to select r' such that the objective is again finite. The flexibility of having $r > 1$ may be used to satisfy the conditions (A2) and (A3) (see below) on the feasible set. If the linear stochastic program is fully random (i.e., costs, right-hand sides and technology matrices are random), the conditions $r \geq T$ and $r' = \infty$ allow to derive a stability result for the optimal values, where the norm $\|\xi\|_T$ enters the stability constant (see the proof of [22, Theorem 2.1]).

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

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

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

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

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

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

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

$$S_\alpha(\xi) := \{x \in \mathcal{X}(\xi) : F(\xi, x) \leq v(\xi) + \alpha\} \quad \text{and} \quad S(\xi) := S_0(\xi)$$

denote the *α-approximate solution set* and the *solution set* of the stochastic program (11) with input ξ, respectively.

The following conditions are imposed on (11):

(A1) $\xi \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$, i.e., $\int_\Omega |\xi(\omega)|^r d\mathbb{P}(\omega) < \infty$, for some $r \geq 1$.

(A2) There exists a $\delta > 0$ such that for any $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ with $\|\tilde{\xi} - \xi\|_r \leq \delta$, any $t = 2, \ldots, T$ and any $x_1 \in \mathcal{X}_1(\tilde{\xi}_1)$, $x_\tau \in \mathcal{X}_\tau(x_{\tau-1}; \tilde{\xi}_\tau)$, $\tau = 2, \ldots, t-1$, there exists an $\mathcal{F}_t(\tilde{\xi})$-measurable $x_t \in \mathcal{X}_t(x_{t-1}; \tilde{\xi}_t)$ (*relatively complete recourse locally around* ξ).

(A3) The optimal values $v(\tilde{\xi})$ of (11) with input $\tilde{\xi}$ are finite for all $\tilde{\xi}$ in a neighborhood of ξ and the objective function F is *level-bounded locally uniformly at* ξ, i.e., for some $\alpha > 0$ there exist a constant $\delta > 0$ and a bounded subset B of $L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ such that $S_\alpha(\tilde{\xi})$ is contained in B for all $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ with $\|\tilde{\xi} - \xi\|_r \leq \delta$.

The following stability result for optimal values and solutions of multistage stochastic programs represents a combination of [22, Theorem 2.1] on quantitative continuity properties of optimal values and of [21, Theorem 2.2] on the convergence of approximate solutions. Its main observation is that multistage models behave stable at some stochastic input process if both its probability distribution and its filtration are approximated with respect to the L_r-distance and the filtration distance

$$D_{\mathrm{f}}(\xi, \tilde{\xi}) := \sup_{\varepsilon > 0} \inf_{\substack{x \in S_\varepsilon(\xi) \\ \tilde{x} \in S_\varepsilon(\tilde{\xi})}} \sum_{t=2}^{T-1} \max\{\|x_t - \mathbb{E}[x_t | \mathcal{F}_t(\tilde{\xi})]\|_{r'}, \|\tilde{x}_t - \mathbb{E}[\tilde{x}_t | \mathcal{F}_t(\xi)]\|_{r'}\}, \quad (12)$$

where $\mathcal{F}_t(\xi)$ and $\mathcal{F}_t(\tilde{\xi})$ denote the σ-fields generated by ξ^t and $\tilde{\xi}^t$, and $\mathbb{E}[\cdot|\mathcal{F}_t(\xi)]$ and $\mathbb{E}[\cdot|\mathcal{F}_t(\tilde{\xi})]$, $t = 1, \ldots, T$, the corresponding conditional expectations, respectively. Note that for the supremum in (12) only small ε's are relevant and that the approximate solution sets are bounded for $\varepsilon \in (0, \alpha]$ according to (A3).

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

$$|v(\xi) - v(\tilde{\xi})| \leq L(\|\xi - \tilde{\xi}\|_r + D_{\mathrm{f}}(\xi, \tilde{\xi})) \quad (13)$$

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

An example in [22] 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 both $S(\xi)$ and $S(\tilde{\xi})$ are nonempty, the filtration distance is of the simplified form

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

For example, the solution sets $S(\xi)$ and $S(\tilde{\xi})$ are nonempty if conditions (A1)–(A3) are satisfied and if Ω is finite or $1 < r' < \infty$. This fact is due to the continuity of $F(\tilde{\xi}, \cdot)$ with respect to the weak and norm topologies in $L_{r'}(\Omega, \mathcal{F}, \mathbb{P})$ and to the nonemptiness (because of (A2)), convexity and closedness of $\mathcal{X}(\tilde{\xi})$ for every $\tilde{\xi}$ with $\|\tilde{\xi} - \xi\|_r \leq \delta$. Moreover, if Ω is finite, (A3) implies that certain level sets are bounded in the finite-dimensional space $L_{r'}(\Omega, \mathcal{F}, \mathbb{P})$ and, hence, compact. For general Ω, the condition $1 < r' < \infty$ implies that the Banach space $L_{r'}(\Omega, \mathcal{F}, \mathbb{P})$ is reflexive [7, Corollary IV.8.2], hence, the level sets are weakly compact as they are closed, convex and bounded (due to (A3)). The existence of solutions then follows from Weierstrass' theorem on the existence of minimizers of continuous functions on compact topological spaces. A more thorough discussion of the existence of solutions of (11) is given in [21, Theorem 2.1]. The paper [21] also contains stability results for (approximate) solution sets.

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 (13) 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 Sect. 2).

Another bound for the filtration distance of ξ and $\tilde{\xi}$ in $L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ with $\|\tilde{\xi} - \xi\|_r \leq \delta$ is immediately obtained by the estimate

$$D_{\mathrm{f}}(\xi, \tilde{\xi}) \leq C \sup_{\|x\|_{r'} \leq 1} \sum_{t=2}^{T-1} \|\mathbb{E}[x_t|\mathcal{F}_t(\xi)] - \mathbb{E}[x_t|\mathcal{F}_t(\tilde{\xi})]\|_{r'} =: C\, D_{\mathrm{f}}^*(\xi, \tilde{\xi}), \qquad (15)$$

where $\delta > 0$ and B are the constant and $L_{r'}$-bounded set appearing in (A2) and (A3), respectively, and the constant $C > 0$ is chosen such $\|x\|_{r'} \leq C$ for all $x \in B$. Another estimate for D_{f} may be obtained if (A3) is replaced by the following stronger condition (A3)$'$.

(A3)$'$ The optimal values $v(\tilde{\xi})$ of (11) with input $\tilde{\xi}$ are finite for all $\tilde{\xi}$ in a neighborhood of ξ and for some $\alpha > 0$ there exist constants $\delta > 0$ and $C > 0$ such that $|\tilde{x}(\omega)| \leq C$ for $\mathbb{P}$-almost every $\omega \in \Omega$ and all $\tilde{x} \in S_\alpha(\tilde{\xi})$ with $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ and $\|\tilde{\xi} - \xi\|_r \leq \delta$.

If (A3)$'$ is satisfied, we immediately have

$$D_{\mathrm{f}}(\xi, \tilde{\xi}) \leq C \sup_{x \in \mathcal{B}_\infty} \sum_{t=2}^{T-1} \|\mathbb{E}[x_t|\mathcal{F}_t(\xi)] - \mathbb{E}[x_t|\mathcal{F}_t(\tilde{\xi})]\|_{r'} =: C\, D_{\mathrm{f},\infty}^*(\xi, \tilde{\xi}) \qquad (16)$$

where $\mathcal{B}_\infty := \{x : \Omega \to \mathbb{R}^m : x \text{ is measurable}, |x(\omega)| \leq 1 \text{ for all } \omega \in \Omega\}$. Such bounds of D_{f} in terms of distances of σ-fields are already discussed in [22, Remark 2.5]. Since D_{f}^* and $D_{\mathrm{f},\infty}^*$ are given as uniform distances of conditional expectation operators, they satisfy the triangle inequality in contrast to D_{f} and do no longer depend on (approximate) solutions of (11).

4 Generating scenario trees

Let $\hat{\xi}$ be a stochastic process on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ having scenarios $\xi^i = (\xi_1^i, \ldots, \xi_T^i) \in \mathbb{R}^{Td}$ with probabilities $p_i > 0, i = 1, \ldots, N$, and common root, i.e., $\xi_1^1 = \cdots = \xi_1^N =: \xi_1^*$. The process $\hat{\xi}$ may be tree-structured or a fan of scenarios. The aim of this section is to describe algorithms for constructing a tree process ξ_{tr} out of $\hat{\xi}$ such that the L_r-distance

$$\|\hat{\xi} - \xi_{\mathrm{tr}}\|_r$$

is bounded by some prescribed tolerance $\varepsilon > 0$. Sections 4.1 and 4.2 contain backward and forward algorithms based on recursive scenario reduction and bundling.

For later use we introduce the L_r-seminorms $\|\cdot\|_{r,t}$ on $L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ (with $s = Td$) given by

$$\|\xi\|_{r,t} := \left(\mathbb{E}[|\xi|_t^r]\right)^{\frac{1}{r}} = \left(\sum_{i=1}^{N} p_i |\xi^i|_t^r\right)^{\frac{1}{r}}, \tag{17}$$

which is needed at step t of the constructions in Sects. 4.1 and 4.2. Here, we denote by $|\cdot|_t$ the seminorm on $\mathbb{R}^s$ defined by $|\xi|_t := |(\xi_1, \ldots, \xi_t, 0, \ldots, 0)|$ for each $\xi = (\xi_1, \ldots, \xi_T) \in \mathbb{R}^s$.

4.1 Backward tree construction

Setting $\bar{\xi}^{T+1} := \hat{\xi}$, 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 $|I_t|$. In this way 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 $J_t := I_{t+1} \backslash I_t$ 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.,

$$\pi_t^i = \pi_{t+1}^i + \sum_{j \in J_{t,i}} \pi_{t+1}^j \quad (i \in I_t), \tag{18}$$

where

$$J_t = \bigcup_{i \in I_t} J_{t,i}, \;\; J_{t,i} := \{j \in J_t : i = i_t(j)\} \;\text{ and }\; i_t(j) \in \arg\min_{i \in I_t} |\xi^i - \xi^j|_t. \tag{19}$$

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

$$\|\bar{\xi}^{t+1} - \bar{\xi}^t\|_{r,t}^r = \sum_{j \in J_t} \pi_{t+1}^j \min_{i \in I_t} |\xi^i - \xi^j|_t^r \tag{20}$$

for $t = 1, \ldots, T$. The final scenario tree ξ_{tr} consists of $|I_T|$ scenarios ξ_{tr}^j with probabilities π_T^j for $j \in I_T$. Each of its components $\xi_{\mathrm{tr},t}^j$ 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

$$\alpha_t(j) := \begin{cases} i_t(\alpha_{t+1}(j)), & \alpha_{t+1}(j) \in J_t, \\ \alpha_{t+1}(j), & \text{otherwise}, \end{cases} \tag{21}$$

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

Theorem 4.1 *Let the stochastic process $\hat{\xi}$ with fixed initial node ξ_1^*, scenarios ξ^i and probabilities p_i, $i = 1, \ldots, N$, be given. Let ξ_{tr} be the stochastic process with scenarios $\xi_{\text{tr}}^i = (\xi_1^*, \xi_{\text{tr},2}^{\alpha_2(i)}, \ldots, \xi_{\text{tr},t}^{\alpha_t(i)}, \ldots, \xi_{\text{tr},T}^i)$ and probabilities π_T^i for $i \in I_T$. Then we have the estimate*

$$\|\hat{\xi} - \xi_{\text{tr}}\|_r \leq \sum_{t=2}^{T} \left(\sum_{j \in J_t} \pi_{t+1}^j \min_{i \in I_t} |\xi^i - \xi^j|_t^r \right)^{\frac{1}{r}}. \tag{22}$$

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 Fig. 1, 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 $\hat{\xi}^1 = \bar{\xi}_{\text{tr}}$. Next we show for $t = 1, \ldots, T-1$ that

$$\|\hat{\xi}^{t+1} - \hat{\xi}^t\|_r = \|\bar{\xi}^{t+1} - \bar{\xi}^t\|_{r,t}. \tag{23}$$

We have

$$\|\hat{\xi}^{t+1} - \hat{\xi}^t\|_r = \sum_{i \in I_T} \pi_T^i |\hat{\xi}^{t+1,i} - \hat{\xi}^{t,i}|^r. \tag{24}$$

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 (24). 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 (21)), we obtain

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

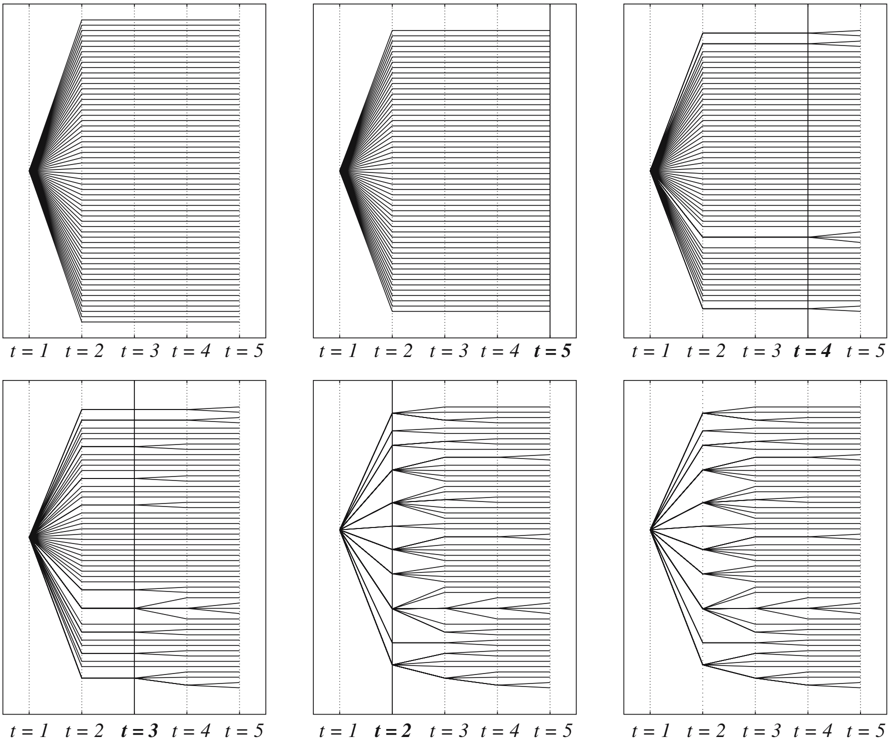

Fig. 1 Illustration of the backward tree construction for an example including $T = 5$ time periods starting with $\hat{\xi}$ containing $N = 58$ scenarios. A possibly existing tree structure of $\hat{\xi}$ is disregarded

With (21) and (20) the latter sum may be rewritten as

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

Hence, the proof of (23) for $t = 1, \ldots, T$ is complete.

Finally, we prove (22) by applying repeatedly the triangle inequality for $\| \cdot \|_r$, using (23) and the identities $\hat{\xi} = \bar{\xi}^{T+1}$, $\hat{\xi}^T = \bar{\xi}^T$ and $\hat{\xi}^1 = \xi_{\text{tr}}$.

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

where the summand for $t = 1$ vanishes. Together with the representation (20) of $\|\cdot\|_{r,t}$, the proof is complete. □

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 $N_t \le N$ 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 (22) 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 $\xi_1^* \in \mathbb{R}^d$, $r \ge 1$, and tolerances ε, ε_t, $t = 2, \ldots, T$, be given such that $\sum_{t=2}^{T} \varepsilon_t \le \varepsilon$.
Step 0: *Set $\bar{\xi}^{T+1} := \hat{\xi}$ 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 $\|\bar{\xi}^{T+1} - \bar{\xi}^T\|_r \le \varepsilon_T$.*
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 $\|\bar{\xi}^{T-t+1} - \bar{\xi}^{T-t}\|_{r,T-t} \le \varepsilon_{T-t}$.*
Step T-1: *Construct the stochastic process $\xi_{\rm tr}$ having $|I_T|$ scenarios $\xi_{\rm tr}^j$, $j \in I_T$, such that $\xi_{{\rm tr},t}^j := \xi_t^{\alpha_t(j)}$, $t = 1, \ldots, T$, where $\alpha_t(\cdot)$ is defined by (21).*

While the first picture in Fig. 1 illustrates the original process $\hat{\xi}$ (disregarding a possibly existing tree structure), 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 $\xi_{\rm tr}$.

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

$$\|\hat{\xi} - \xi_{\rm tr}\|_r \le \sum_{t=2}^{T} \varepsilon_t \le \varepsilon.$$

Proof This is a direct consequence of the estimate (22) in Theorem 4.1, which reads

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

□

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 Sect. 7.1, where the (non-vanishing) tolerances are chosen according to the exponential rule (42).

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

$$C_t^k \cap C_t^{k'} = \emptyset \quad \forall k \neq k' \quad \text{and} \quad \bigcup_{k=1}^{K_t} C_t^k = I. \tag{25}$$

The elements of such a partition $\mathcal{C}_t$ will be called (scenario) clusters. The initialization of the procedure consists in setting $\hat{\xi}^1 = \hat{\xi}$, 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 Sect. 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

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

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

$$\alpha_t(j) = \begin{cases} i_t^k(j), & j \in J_t^k,\; k = 1, \ldots, K_{t-1}, \\ j, & \text{otherwise.} \end{cases} \tag{26}$$

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

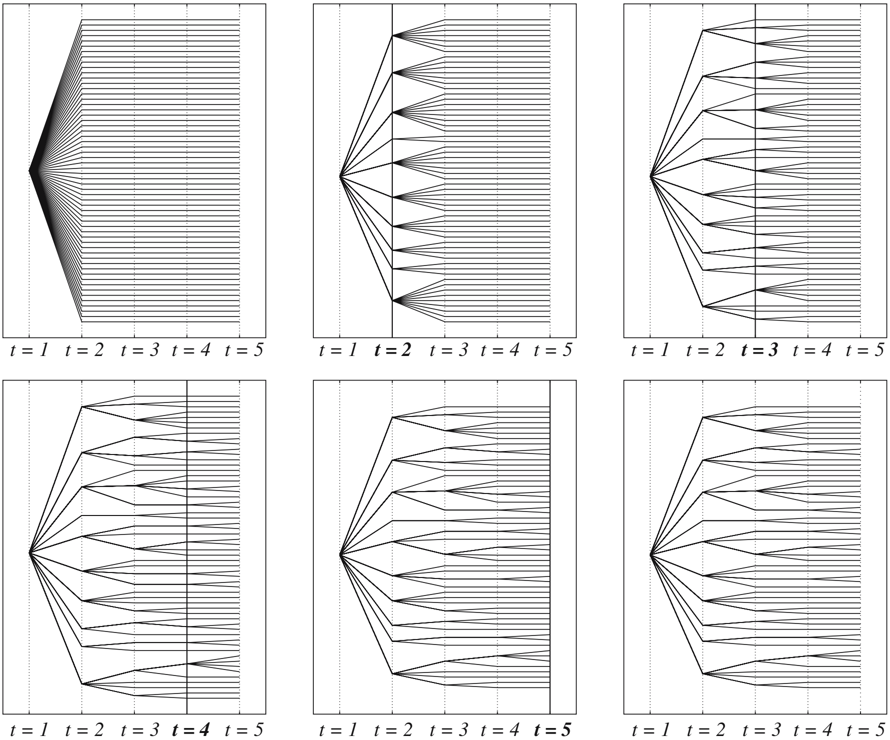

Fig. 2 Illustration of the forward tree construction for an example including $T=5$ time periods starting with $\hat{\xi}$ containing $N=58$ scenarios. A possibly existing tree structure of $\hat{\xi}$ is disregarded

$$\hat{\xi}^{t,i}_{\tau} = \begin{cases} \xi^{\alpha_\tau(i)}_{\tau}, & \tau \leq t, \\ \xi^{i}_{\tau}, & \text{otherwise}, \end{cases} \tag{27}$$

with probabilities p_i for each $i \in I$. The processes $\hat{\xi}^t$ are illustrated in Fig. 2, 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

$$\mathcal{C}_t = \left\{ \alpha_t^{-1}(i) : i \in I_t^k,\ k = 1, \ldots, K_{t-1} \right\}, \tag{28}$$

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

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

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

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 $\xi_{\rm tr} := \hat{\xi}^T$ are given by the structure of the final partition $\mathcal{C}_T$, i.e., they are of the form

$$\xi_{\rm tr}^k = \left(\xi_1^*, \xi_2^{\alpha_2(i)}, \dots, \xi_t^{\alpha_t(i)}, \dots, \xi_T^{\alpha_T(i)}\right) \quad \text{and} \quad \pi_T^i \quad \text{if} \;\; i \in C_T^k$$

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

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

$$\|\hat{\xi} - \xi_{\rm tr}\|_r \le \sum_{t=2}^{T} \left(\sum_{k=1}^{K_{t-1}} \sum_{j \in J_t^k} p_j \min_{i \in I_t^k} |\xi_t^i - \xi_t^j|^r \right)^{\frac{1}{r}}. \tag{29}$$

Proof We recall that $\hat{\xi}^1 = \hat{\xi}$ and $\hat{\xi}^T = \xi_{\rm tr}$ and obtain

$$\|\hat{\xi} - \xi_{\rm tr}\|_r \le \sum_{t=2}^{T} \|\hat{\xi}^t - \hat{\xi}^{t-1}\|_r,$$

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

$$\|\xi - \tilde{\xi}_{\rm tr}\|_r \le \sum_{t=2}^{T} \|\hat{\xi}^t - \hat{\xi}^{t-1}\|_{r,t}. \tag{30}$$

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 (25) and the definitions (26) of the mappings α_t and i_t^k. Inserting the latter result into (30) 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, $i = 1, \ldots, N$, fixed root $\xi_1^* \in \mathbb{R}^d$ and probability distribution P, $r \geq 1$, and tolerances ε, ε_t, $t = 2, \ldots, T$, be given such that $\sum_{t=2}^T \varepsilon_t \leq \varepsilon$.
Step 1: *Set $\hat{\xi}^1 := \hat{\xi}$ 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 (26) and a stochastic process $\hat{\xi}^t$ having N scenarios $\hat{\xi}^{t,i}$ with probabilities p_i according to (27) 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 ξ_{tr} having K_T scenarios ξ_{tr}^k such that $\xi_{\text{tr},t}^k := \xi_t^{\alpha_t(i)}$ if $i \in C_T^k$, $k = 1, \ldots, K_T$, $t = 1, \ldots, T$.*

While the first picture in Fig. 2 illustrates the original process $\hat{\xi}$ (disregarding a possibly existing tree structure), 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 ξ_{tr}.

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

$$\|\hat{\xi} - \xi_{\text{tr}}\|_r \leq \sum_{t=2}^{T} \varepsilon_t \leq \varepsilon.$$

Proof This is a direct consequence of (30). □

When using Algorithm 4.5, the selection of $r > 1$ should be done according to the same reasons as mentioned at the end of Sect. 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 (43) in Sect. 7.2.

5 Convergence

Let ξ be the original stochastic process on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\hat{\xi}$ another stochastic process on $(\Omega, \mathcal{F}, \mathbb{P})$ having scenarios $\xi^i = (\xi_1^i, \ldots, \xi_T^i) \in \mathbb{R}^{Td}$

with probabilities $p_i > 0, i = 1, \dots, N$, and common root, i.e., $\xi_1^1 = \cdots = \xi_1^N =: \xi_1^*$, as in Sect. 4. Let $\xi_{\rm tr}$ be a process obtained by one of the tree generation algorithms in Sect. 4 starting from $\hat{\xi}$ with tolerance $\varepsilon_r > 0$, i.e., it holds

$$\|\hat{\xi} - \xi_{\rm tr}\|_r \leq \varepsilon_r. \tag{31}$$

Our aim is to establish conditions on $\hat{\xi}$ implying that $\xi_{\rm tr}$ is a good approximation of ξ in terms of the L_r-distance and in terms of $D_{\rm f}$, $D_{\rm f}^*$ or $D_{{\rm f},\infty}^*$. In this section we assume that one of the following conditions (C1) and (C2) on $\hat{\xi}$ is satisfied:

(C1) $\|\xi - \hat{\xi}\|_r$ and $D_{\rm f}^*(\xi, \hat{\xi})$ or $D_{{\rm f},\infty}^*(\xi, \hat{\xi})$ are small.

(C2) $\|\xi - \hat{\xi}\|_r$ is small and $\hat{\xi}$ is adapted to the filtration of ξ, i.e., $\mathcal{F}_t(\hat{\xi}) \subseteq \mathcal{F}_t(\xi)$ for all $t = 1, \dots, T$.

It will turn out in both cases (see Propositions 5.2 and 5.4) that the distance of optimal values $|v(\xi) - v(\xi_{\rm tr})|$ gets small if ε_r is small and $\hat{\xi}$ is a sufficiently good approximation in the sense stated in (C1) or (C2). The two conditions are illustrated by two special cases, namely, by sampling from finitely discrete probability measures (Example 5.3) and discretization schemes for general probability distributions (Example 5.5).

Proposition 5.1 *Let $\xi_{\rm tr}$ be obtained by one of the tree generation algorithms in Sect. 4 starting from $\hat{\xi}$. Then $\mathcal{F}_t(\xi_{\rm tr}) \subseteq \mathcal{F}_t(\hat{\xi})$ holds for all $t = 1, \dots, T$.*

Proof According to the construction of $\xi_{\rm tr}$ we have $\mathcal{F}_1(\xi_{\rm tr}) = \mathcal{F}_1(\hat{\xi})$. Now, let $t \in \{2, \dots, T\}$ and $A_{t,i} := (\hat{\xi}_1, \dots, \hat{\xi}_t)^{-1}(\{(\xi_1^*, \xi_2^i, \dots, \xi_t^i)\})$, $i = 1, \dots, N$. Then $\mathcal{F}_t(\hat{\xi})$ is the smallest σ-field containing $\{A_{t,i}\}_{i=1}^N$. With the notation in Sect. 4 we obtain for $t \in \{2, \dots, T\}$ and $k \in I_t$ that

$$\begin{aligned}(\xi_{{\rm tr},1}, \dots, \xi_{{\rm tr},t})^{-1}(\xi_{{\rm tr},1}^k, \dots, \xi_{{\rm tr},t}^k) &= \{\omega : \xi_{{\rm tr},\tau}(\omega) = \xi_\tau^{\alpha_\tau(i)}, \tau = 2, \dots, t, i \in \bar{I}_{t,k}\} \\ &= \bigcap_{\tau=2}^{t} \bigcup_{i \in \bar{I}_{t,k}} A_{\tau,\alpha_\tau(i)} \in \mathcal{F}_t(\hat{\xi}).\end{aligned}$$

Hence, we have $\mathcal{F}_t(\xi_{\rm tr}) \subseteq \mathcal{F}_t(\hat{\xi})$. □

Proposition 5.2 *Let $1 \leq r' < \infty$, (A1), (A2) and (A3) be satisfied and X_1 be bounded. Let $L > 0$, $C > 0$ and $\delta > 0$ be the constants appearing in Theorem 3.1 and (15), and let $\|\xi - \hat{\xi}\|_r \leq \delta - \varepsilon_r$. Then we have*

$$|v(\xi) - v(\xi_{\rm tr})| \leq L(\varepsilon_r + \|\xi - \hat{\xi}\|_r + C\, D_{\rm f}^*(\xi, \hat{\xi}) + C\, D_{\rm f}^*(\hat{\xi}, \xi_{\rm tr})) \tag{32}$$

If $(\varepsilon_r^{(n)})$ is a sequence tending to 0 such that the corresponding tolerances $\varepsilon_t^{(n)}$ in Algorithms 4.2 and 4.5 are nonincreasing for all $t = 2, \dots, T$, the corresponding sequence $(\xi_{\rm tr}^{(n)})$ has the property $D_{\rm f}^*(\hat{\xi}, \xi_{\rm tr}^{(n)}) \underset{n\to\infty}{\longrightarrow} 0$.

Proof The assumption on $\hat{\xi}$ implies $\|\xi - \xi_{\mathrm{tr}}\|_r \leq \delta$. Hence, Theorem 3.1, (16) and the triangle inequality for $\|\cdot\|_r$ and D_{f}^* provide the estimate

$$\begin{aligned}|v(\xi) - v(\xi_{\mathrm{tr}})| &\leq L(\|\xi - \xi_{\mathrm{tr}}\|_r + C\, D_{\mathrm{f}}^*(\xi, \xi_{\mathrm{tr}}))\\ &\leq L(\|\xi - \hat{\xi}\|_r + C\, D_{\mathrm{f}}^*(\xi, \hat{\xi}) + \varepsilon_r + C\, D_{\mathrm{f}}^*(\hat{\xi}, \xi_{\mathrm{tr}}^{(n)})).\end{aligned}$$

If the sequences $(\varepsilon_t^{(n)})$ are nonincreasing for all $t = 1, \ldots, T$, the σ-fields $\mathcal{F}_t(\xi_{\mathrm{tr}}^{(n)})$ are nondecreasing with respect to $n \in \mathbb{N}$ and for every $t = 1, \ldots, T$. If the sequence $(\varepsilon_r^{(n)})$ converges to 0, $(\xi_{\mathrm{tr}}^{(n)})$ converges to $\hat{\xi}$ in L_r. Hence, $\hat{\xi}_t$ is measurable with respect to

$$\hat{\mathcal{F}}_t := \sigma\left(\bigcup_{n\in\mathbb{N}} \mathcal{F}_t(\xi_{\mathrm{tr}}^{(n)})\right) \quad (t = 1, \ldots, T).$$

Hence, for any $x \in L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ with $\|x\|_{r'} \leq 1$, we obtain

$$\|\mathbb{E}[\mathbb{E}[x_t|\mathcal{F}_t(\hat{\xi})]|\mathcal{F}_t(\xi_{\mathrm{tr}}^{(n)})] - \mathbb{E}[x_t|\mathcal{F}_t(\hat{\xi})]\|_{r'} = \|\mathbb{E}[x_t|\mathcal{F}_t(\xi_{\mathrm{tr}}^{(n)})] - \mathbb{E}[x_t|\mathcal{F}_t(\hat{\xi})]\|_{r'} \underset{n\to\infty}{\longrightarrow} 0$$

for every $t = 1, \ldots, T$ due to classical convergence results for conditional expectations (see, e.g., [11]). Moreover, we have

$$\begin{aligned}D_{\mathrm{f}}^*(\hat{\xi}, \xi_{\mathrm{tr}}^{(n)}) &= \sup_{\|x\|_{r'}\leq 1} \sum_{t=2}^{T-1} \|\mathbb{E}[x_t|\mathcal{F}_t(\xi_{\mathrm{tr}}^{(n)})] - \mathbb{E}[x_t|\mathcal{F}_t(\hat{\xi})]\|_{r'}\\ &= \sup_{\|x\|_{r'}\leq 1} \sum_{t=2}^{T-1} \|\mathbb{E}[\mathbb{E}[x_t|\mathcal{F}_t(\hat{\xi})]|\mathcal{F}_t(\xi_{\mathrm{tr}}^{(n)})] - \mathbb{E}[x_t|\mathcal{F}_t(\hat{\xi})]\|_{r'}\\ &= \sup_{\|\hat{x}\|_{r'}\leq 1} \sum_{t=2}^{T-1} \|\mathbb{E}[\hat{x}_t|\mathcal{F}_t(\xi_{\mathrm{tr}}^{(n)})] - \hat{x}_t\|_{r'}\end{aligned}$$

where $\hat{x}$ varies in the finite-dimensional space $\times_{t=1}^T L_{r'}(\Omega, \mathcal{F}_t(\hat{\xi}), \mathbb{P}; \mathbb{R}^{m_t})$. Hence, the unit ball in this space is compact. Since $\sum_{t=2}^{T-1} \|\mathbb{E}[\hat{x}_t|\mathcal{F}_t(\xi_{\mathrm{tr}}^{(n)})] - \hat{x}_t\|_{r'}$ converges to 0 for every $\hat{x}$ in the unit ball as $n \to \infty$, we conclude by standard compactness arguments that

$$D_{\mathrm{f}}^*(\hat{\xi}, \xi_{\mathrm{tr}}^{(n)}) = \sup_{\|\hat{x}\|_{r'}\leq 1} \sum_{t=2}^{T-1} \|\mathbb{E}[\hat{x}_t|\mathcal{F}_t(\xi_{\mathrm{tr}}^{(n)})] - \hat{x}_t\|_{r'} \underset{n\to\infty}{\longrightarrow} 0.$$

□

Since the set $\mathcal{B}_\infty$ (see (16)) is contained in $\{x \in L_{r'} : \|\hat{x}\|_{r'} \leq 1\}$, Proposition 5.2 remains valid if (A3) and D_{f}^* are replaced by (A3)' and $D_{\mathrm{f},\infty}^*$, respectively. We note that the case $r' = \infty$ has to be excluded in Proposition 5.2 as the convergence result for conditional expectations with increasing σ-fields is violated for $r' = \infty$ [11].

Example 5.3 (Sampling) Let P be a probability distribution on $\Xi = \{\xi^1, \ldots, \xi^N\} \subseteq \mathbb{R}^{Td}$, where the scenarios ξ^i have a common root, i.e., $\xi_1^1 = \cdots = \xi_1^N =: \xi_1^*$ and are possibly tree-structured, i.e., there are index sets I_t and $\bar{I}_{t,i}$, $i \in I_t$, contained in $\{1, \ldots, N\}$ such that $\xi_t^j = \xi_t^i$ for each $j \in \bar{I}_{t,i}$, $i \in I_t$ and that any two elements of $\{\xi_t^i\}_{i \in I_t}$ do not coincide for all $t = 1, \ldots, T$. Furthermore, we set $p_i := P(\{\xi^i\})$, $i = 1, \ldots, N$, $p := (p_1, \ldots, p_N) \in \mathcal{X} := \{x \in \mathbb{R}_+^N : \sum_{i=1}^N x_i = 1\}$.

Let $(\xi_n)_{n \in \mathbb{N}}$ be a sequence of independent and identically distributed Ξ-valued random variables on some probability space $(\Omega^*, \mathcal{F}^*, \mathbb{P}^*)$ such that $P = \mathbb{P}^* \xi_1^{-1}$. We consider the (random) empirical measures

$$P_n(\omega^*) := \frac{1}{n} \sum_{j=1}^{n} \delta_{\xi_j(\omega^*)} \quad (n \in \mathbb{N}, \omega^* \in \Omega^*),$$

where δ_z denotes the probability measure on Ξ placing unit mass at $z \in \Xi$. Then the sequence $(P_n(\omega^*))$ converges to P in the sense of weak convergence for $\mathbb{P}^*$-almost every $\omega^* \in \Omega^*$ (see, e.g., [6, Chap. 11.4]). Since P and $P_n(\omega^*)$ belong to the set $\mathcal{P}_r(\Xi)$ (see Sect. 2) for any $r \geq 1$, $\ell_r(P, P_n(\omega^*))$ tends to 0 as $n \to \infty$ for $\mathbb{P}^*$-almost every $\omega^* \in \Omega^*$. Let $\bar{\omega}^*$ be such that $(P_n(\bar{\omega}^*))$ converges to P with respect to ℓ_r. The measure $P_n(\bar{\omega}^*)$ is of the form

$$P_n(\bar{\omega}^*) = \sum_{i=1}^{N} p_i^{(n)} \delta_{\xi^i},$$

where $p_i^{(n)} \geq 0$ and $p_i^{(n)} \underset{n \to \infty}{\longrightarrow} P(\{\xi^i\})$ due to the weak convergence of $(P_n(\bar{\omega}^*))$ to P. We set $p^{(n)} := (p_1^{(n)}, \ldots, p_N^{(n)}) \in \mathcal{X}$ and $\hat{P} := P_n(\bar{\omega}^*)$ for some $n \in \mathbb{N}$ that will be specified later on.

Let $r \geq 1$ and r' according to (10) be fixed. In order to define a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and Ξ-valued random variables ξ and $\hat{\xi}$ having probability distributions P and $\hat{P}$, respectively, we consider the parametric linear transportation problem

$$\min \left\{ \sum_{i,j=1}^{N} \eta_{ij} |\xi^i - \xi^j|^r : \eta_{ij} \geq 0, \sum_{i} \eta_{ij} = q_j, \sum_{j} \eta_{ij} = p_i, i, j = 1, \ldots, N \right\}, \tag{33}$$

where $q = (q_1, \ldots, q_N) \in \mathcal{X}$ plays the role of a parameter. Let $\Sigma(q) \subset \mathbb{R}^{N \times N}$ denote the solution set of the minimization problem (33). Its optimal value coincides with $\ell_r^r(P, Q)$, where $Q = \sum_{j=1}^N q_j \delta_{\xi^j}$ (see also (5) and (7)). Clearly, the diagonal matrix $\mathrm{diag}(p_1, \ldots, p_N)$ belongs to $\Sigma(p)$. The set-valued mapping Σ from $\mathcal{X}$ to $\mathbb{R}^{N \times N}$ is polyhedral, i.e., its graph is the union of finitely many polyhedral convex sets, and, hence, is locally upper Lipschitzian at each element of $\mathcal{X}$ [37, Proposition 1]. Hence, there exists constants $L > 0$ and $\gamma > 0$ such that

$$\inf_{\mathcal{E} \in \Sigma(q)} \|\mathrm{diag}(p_1, \ldots, p_N) - \mathcal{E}\|_* \leq L \|p - q\| \tag{34}$$

for all $q \in \mathcal{X}$ with $\|p - q\| \leq \gamma$. Here, the norm $\|\cdot\|_*$ is some matrix norm on $\mathbb{R}^{N\times N}$ and the norm on the right-hand side is defined on $\mathbb{R}^N$. For each $n \in \mathbb{N}$, let $\mathcal{E}^{*,(n)} = (\eta_{ij}^{*,(n)})_{i,j=1,\ldots,N}$ be selected such that

$$\|\text{diag}(p_1, \ldots, p_N) - \mathcal{E}^{*,(n)}\|_* = \inf_{\mathcal{E}\in\Sigma(p^{(n)})} \|\text{diag}(p_1, \ldots, p_N) - \mathcal{E}\|_*.$$

Due to (34) we have that

$$\|\text{diag}(p_1, \ldots, p_N) - \mathcal{E}^{*,(n)}\|_* \leq L\|p - p^{(n)}\|$$

for sufficiently large n. Hence, the matrix norm $\|\text{diag}(p_1, \ldots, p_N) - \mathcal{E}^{*,(n)}\|_*$ gets small for large n and, thus, $\eta_{ii}^{*,(n)}$ converges to p_i for each $i = 1, \ldots, N$ and $\eta_{ij}^{*,(n)}$ converges to 0 if $i \neq j$.

Now, let $\Omega = \{\omega_{ij} : i, j = 1, \ldots, N\}$, $\mathcal{F}$ be the power set of Ω and $\mathbb{P}$ defined by $\mathbb{P}(\omega_{ij}) = \eta_{ij}^{*,(n)}$, $i, j = 1, \ldots, N$. Furthermore, we define $\xi, \hat{\xi} : \Omega \to \Xi$ by $\xi(\omega_{ij}) := \xi^i$ and $\hat{\xi}(\omega_{ij}) := \xi^j$ for all $i, j = 1, \ldots, N$. Then we have

$$\ell_r^r(P, \hat{P}) = \sum_{i,j=1}^{N} \eta_{ij}^{*,(n)} |\xi^i - \xi^j|^r = \|\xi - \hat{\xi}\|_r^r. \tag{35}$$

Let $\varepsilon > 0$ be some tolerance. Our preceding arguments imply that $\|\xi - \hat{\xi}\|_r \leq \varepsilon$ for some sufficiently large $n \in \mathbb{N}$. Next we show that the filtration distance $D^*_{\text{f},\infty}(\xi, \hat{\xi})$ gets small for large n. To this end, we first consider the σ-fields $\mathcal{F}_t(\xi)$ and $\mathcal{F}_t(\hat{\xi})$. They are generated by the partitions $\{E_{tk}\}_{k\in I_t}$ and $\{\hat{E}_{tk}\}_{k\in I_t}$, respectively, of Ω, where

$$E_{tk} := (\xi_1, \ldots, \xi_t)^{-1}(\{\xi^k\}) = \left\{\omega_{ij} : i \in \bar{I}_{t,k}, j = 1, \ldots, N\right\}$$
$$\hat{E}_{tk} := (\hat{\xi}_1, \ldots, \hat{\xi}_t)^{-1}(\{\xi^k\}) = \left\{\omega_{ij} : j \in \bar{I}_{t,k}, i = 1, \ldots, N\right\}$$

for every $k \in I_t$. The probabilities of the sets in each partition are

$$\mathbb{P}(E_{tk}) = \sum_{j=1}^{N} \sum_{i\in\bar{I}_{t,k}} \eta_{ij}^{*,(n)} = \sum_{i\in\bar{I}_{t,k}} p_i = \pi_t^k$$
$$\mathbb{P}(\hat{E}_{tk}) = \sum_{i=1}^{N} \sum_{j\in\bar{I}_{t,k}} \eta_{ij}^{*,(n)} = \sum_{j\in\bar{I}_{t,k}} p_j^{(n)} = \pi_t^{k,(n)}.$$

Thus, we obtain

$$\mathbb{E}[x_t|\mathcal{F}_t(\xi)] = \sum_{k\in I_t} \mathbb{E}[x_t|E_{tk}]\chi_{E_{tk}} = \sum_{k\in I_t} \frac{1}{\pi_t^k} \sum_{j=1}^N \sum_{i\in \bar{I}_{t,k}} x_t(\omega_{ij})\eta_{ij}^{*,(n)} \chi_{E_{tk}}$$

$$\mathbb{E}[x_t|\mathcal{F}_t(\hat{\xi})] = \sum_{k\in I_t} \mathbb{E}[x_t|\hat{E}_{tk}]\chi_{\hat{E}_{tk}} = \sum_{k\in I_t} \frac{1}{\pi_t^{k,(n)}} \sum_{i=1}^N \sum_{j\in \bar{I}_{t,k}} x_t(\omega_{ij})\eta_{ij}^{*,(n)} \chi_{\hat{E}_{tk}}$$

for any random variable $x_t : \Omega \to \mathbb{R}^{m_t}$, where $\chi_A : \Omega \to \mathbb{R}$ denotes the characteristic function of $A \in \mathcal{F}$. Hence, we have

$$\begin{aligned}\|\mathbb{E}[x_t|\mathcal{F}_t(\xi)] - \mathbb{E}[x_t|\mathcal{F}_t(\hat{\xi})]\|_{r'}^{r'} &= \sum_{i,j=1}^N \eta_{ij}^{*,(n)} |\mathbb{E}[x_t|\mathcal{F}_t(\xi)](\omega_{ij}) - \mathbb{E}[x_t|\mathcal{F}_t(\hat{\xi})](\omega_{ij})|^{r'} \\ &= \sum_{i,j=1}^N \eta_{ij}^{*,(n)} |\mathbb{E}[x_t|E_{ti}] - \mathbb{E}[x_t|\hat{E}_{tj}]|^{r'},\end{aligned}$$

where the factor of $\eta_{ij}^{*,(n)}$ in each summand is of the form

$$|\mathbb{E}[x_t|E_{ti}] - \mathbb{E}[x_t|\hat{E}_{tj}]| = \left| \frac{1}{\pi_t^i} \sum_{j=1}^N \sum_{k\in \bar{I}_{t,i}} x_t(\omega_{kj})\eta_{kj}^{*,(n)} - \frac{1}{\pi_t^{j,(n)}} \sum_{i=1}^N \sum_{k\in \bar{I}_{t,j}} x_t(\omega_{ik})\eta_{ik}^{*,(n)} \right|.$$

Now, we consider the set $\mathcal{B}_\infty$ (see (16)) in $\mathbb{R}^{mN^2}$. Since it is compact, there exist $\ell \in \mathbb{N}$ and random variables $x_l : \Omega \to \mathbb{R}^m$, $l = 1, \ldots \ell$, such that $\mathcal{B}_\infty$ is contained in the union of the balls $B_l := \{x \in \mathcal{B}_\infty : \max_{\omega\in\Omega} \max_{t=1,\ldots,T} |x_t(\omega) - x_{lt}(\omega)| \le \frac{\varepsilon}{4T}\}$, $l = 1, \ldots, \ell$. Next we select $n \in \mathbb{N}$ such that the conditions

$$\left| \frac{1}{\pi_t^i} \sum_{j=1}^N \sum_{k\in \bar{I}_{t,i}} x_{lt}(\omega_{kj})\eta_{kj}^{*,(n)} - \frac{1}{\pi_t^i} \sum_{k\in \bar{I}_{t,i}} x_{lt}(\omega_{kk})p_k \right| \le \frac{\varepsilon}{8T}$$

$$\left| \frac{1}{\pi_t^j} \sum_{k\in \bar{I}_{t,j}} x_{lt}(\omega_{kk})p_k - \frac{1}{\pi_t^{j,(n)}} \sum_{i=1}^N \sum_{k\in \bar{I}_{t,j}} x_{lt}(\omega_{ik})\eta_{ik}^{*,(n)} \right| \le \frac{\varepsilon}{8T}$$

$$\sum_{i,j=1}^N \eta_{ij}^{*,(n)} \left(\frac{\varepsilon}{4T} + \left| \frac{1}{\pi_t^i} \sum_{k\in \bar{I}_{t,i}} x_{lt}(\omega_{kk})p_k - \frac{1}{\pi_t^j} \sum_{k\in \bar{I}_{t,j}} x_{lt}(\omega_{kk})p_k \right| \right)^{r'} \le \left(\frac{\varepsilon}{2T}\right)^{r'}$$

are satisfied for all $l = 1, \ldots, \ell$, $t = 2, \ldots, T-1$, and $i, j = 1, \ldots, N$. This implies

$$\|\mathbb{E}[x_{lt}|\mathcal{F}_t(\xi)] - \mathbb{E}[x_{lt}|\mathcal{F}_t(\hat{\xi})]\|_{r'} \le \frac{\varepsilon}{2T}$$

for all $l = 1, \ldots, \ell$, $t = 2, \ldots, T-1$. Thus, we have for any $x \in \mathcal{B}_\infty$

$$\begin{aligned}\|\mathbb{E}[x_t|\mathcal{F}_t(\xi)] - \mathbb{E}[x_t|\mathcal{F}_t(\hat{\xi})]\|_{r'} &\leq \|\mathbb{E}[x_t|\mathcal{F}_t(\xi)] - \mathbb{E}[x_{lt}|\mathcal{F}_t(\xi)]\|_{r'} + \|\mathbb{E}[x_{lt}|\mathcal{F}_t(\xi)] \\ &\quad - \mathbb{E}[x_{lt}|\mathcal{F}_t(\hat{\xi})]\|_{r'} + \|\mathbb{E}[x_{lt}|\mathcal{F}_t(\hat{\xi})] - \mathbb{E}[x_t|\mathcal{F}_t(\hat{\xi})]\|_{r'} \\ &\leq 2\|x_t - x_{lt}\|_{r'} + \|\mathbb{E}[x_{lt}|\mathcal{F}_t(\xi)] - \mathbb{E}[x_{lt}|\mathcal{F}_t(\hat{\xi})]\|_{r'} \\ &\leq \frac{\varepsilon}{2T} + \frac{\varepsilon}{2T} = \frac{\varepsilon}{T},\end{aligned}$$

where $l \in \{1, \ldots, \ell\}$ is chosen such that $x \in B_l$, i.e.,

$$\|x_t - x_{lt}\|_{r'} \leq \max_{\omega \in \Omega} |x_t(\omega) - x_{lt}(\omega)| \leq \frac{\varepsilon}{4T}.$$

Hence, we obtain $D^*_{\mathrm{f},\infty}(\xi, \hat{\xi}) = \sup_{x \in \mathcal{B}_\infty} \sum_{t=2}^{T-1} \|\mathbb{E}[x_t|\mathcal{F}_t(\xi)] - \mathbb{E}[x_t|\mathcal{F}_t(\hat{\xi})]\|_{r'} \leq \varepsilon$.

Proposition 5.4 *Let* $1 \leq r' < \infty$, (A1), (A2) *and* (A3) *be satisfied and* X_1 *be bounded. Assume that* $S(\xi)$ *is nonempty. Let* $(\hat{\xi}^{(n)})$ *be a sequence of discrete processes that converge to* ξ *in* $L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ *and has the property that the* σ*-fields* $\mathcal{F}_t(\hat{\xi}^{(n)})$ *are nondecreasing and contained in* $\mathcal{F}_t(\xi)$, $t = 1, \ldots, T$. *Furthermore, let* $\xi_{\mathrm{tr}}^{(n)}$ *be processes obtained by one of the Algorithms 4.2 and 4.5 starting from* $\hat{\xi}^{(n)}$ *with tolerance* $\varepsilon_r^{(n)}$ *tending to zero as* $n \to \infty$ *and such that the corresponding tolerances* $\varepsilon_t^{(n)}$ *in the algorithms are nonincreasing for all* $t = 2, \ldots, T$. *Then we have*

$$\lim_{n\to\infty} D_{\mathrm{f}}(\xi, \xi_{\mathrm{tr}}^{(n)}) = 0 \quad \textit{and, hence,} \quad \lim_{n\to\infty} |v(\xi) - v(\xi_{\mathrm{tr}}^{(n)})| = 0.$$

Moreover, if $1 < r' < \infty$ *and* $\xi_{\mathrm{tr}} := \xi_{\mathrm{tr}}^{(n_0)}$ *for some* $n_0 \in \mathbb{N}$, *it holds*

$$\liminf_{n\to\infty} \inf_{x^{(n)} \in S(\hat{\xi}^{(n)})} \sum_{t=2}^{T-1} \|x_t^{(n)} - \mathbb{E}[x_t^{(n)}|\mathcal{F}_t(\xi_{\mathrm{tr}})]\|_{r'} \geq D_{\mathrm{f}}(\xi, \xi_{\mathrm{tr}}).$$

Proof Let $L > 0$ and $\delta > 0$ denote the constants in Theorem 3.1. For sufficiently large $n \in \mathbb{N}$, $\xi_{\mathrm{tr}}^{(n)}$ satisfies $\|\xi - \xi_{\mathrm{tr}}^{(n)}\| \leq \delta$. Hence, we conclude from Theorem 3.1 that

$$\begin{aligned}|v(\xi) - v(\xi_{\mathrm{tr}}^{(n)})| &\leq L(\|\xi - \xi_{\mathrm{tr}}^{(n)}\|_r + D_{\mathrm{f}}(\xi, \xi_{\mathrm{tr}}^{(n)})) \\ &\leq L(\varepsilon_r^{(n)} + \|\xi - \hat{\xi}^{(n)}\|_r + D_{\mathrm{f}}(\xi, \xi_{\mathrm{tr}}^{(n)}))\end{aligned}$$

Proposition 5.1 implies $\mathcal{F}_t(\xi_{\mathrm{tr}}^{(n)}) \subseteq \mathcal{F}_t(\hat{\xi}^{(n)}) \subseteq \mathcal{F}_t(\xi)$. In particular, we conclude that $x_t = \mathbb{E}[x_t|\mathcal{F}_t(\xi)]$ holds for every $t = 1, \ldots, T$ and $x \in S_\varepsilon(\xi_{\mathrm{tr}}^{(n)})$. Thus, we obtain for some $x^* \in S(\xi)$

$$D_{\mathrm{f}}(\xi, \xi_{\mathrm{tr}}^{(n)}) = \sup_{\varepsilon>0} \inf_{x \in S_\varepsilon(\xi)} \sum_{t=2}^{T-1} \|x_t - \mathbb{E}[x_t|\mathcal{F}_t(\xi_{\mathrm{tr}}^{(n)})]\|_{r'} \leq \sum_{t=2}^{T-1} \|x_t^* - \mathbb{E}[x_t^*|\mathcal{F}_t(\xi_{\mathrm{tr}}^{(n)})]\|_{r'}.$$

As in the proof of Proposition 5.2 we consider the smallest σ-fields $\hat{\mathcal{F}}_t$ ($t = 1, \dots, T$) containing $\mathcal{F}_t(\xi_{\mathrm{tr}}^{(n)})$ for each $n \in \mathbb{N}$. Again the convergence of $(\xi_{\mathrm{tr}}^{(n)})$ to ξ in L_r implies $\xi_t \in L_r(\Omega, \hat{\mathcal{F}}_t, \mathbb{P}; \mathbb{R}^d)$ for every $t = 2, \dots, T$. Since the σ-fields $\mathcal{F}_t(\hat{\xi}^{(n)})$ are nondecreasing and the tolerances $\varepsilon_t^{(n)}$ in both algorithms are nonincreasing, the σ-fields $\mathcal{F}_t(\xi_{\mathrm{tr}}^{(n)})$ are nondecreasing with respect to $n \in \mathbb{N}$ for every $t = 2, \dots, T$. Moreover, the sequence $(\xi_{\mathrm{tr}}^{(n)})$ also converges to ξ in L_r. Hence, ξ_t and, thus, x_t^* is measurable with respect to $\hat{\mathcal{F}}_t$. Classical convergence results for conditional expectations (e.g., [11]) then imply

$$\|\mathbb{E}[x_t^*|\mathcal{F}_t(\xi_{\mathrm{tr}}^{(n)})] - \mathbb{E}[x_t^*|\hat{\mathcal{F}}_t]\|_{r'} = \|\mathbb{E}[x_t^*|\mathcal{F}_t(\xi_{\mathrm{tr}}^{(n)})] - x_t^*\|_{r'} \underset{n\to\infty}{\longrightarrow} 0.$$

for all $t = 1, \dots, T$. Hence, we obtain $D_{\mathrm{f}}(\xi, \xi_{\mathrm{tr}}^{(n)}) \underset{n\to\infty}{\longrightarrow} 0$ and, thus, according to Theorem 3.1 that $|v(\xi) - v(\xi_{\mathrm{tr}}^{(n)})| \underset{n\to\infty}{\longrightarrow} 0$.

Now, let $\xi_{\mathrm{tr}} = \xi_{\mathrm{tr}}^{(n_0)}$ for some $n_0 \in \mathbb{N}$. We note that $S(\hat{\xi}^{(n)})$ is nonempty for large n (due to (A2), (A3)). Let $\bar{x}^n \in S(\hat{\xi}^{(n)})$ for $n \geq n_1 \geq n_0$ be selected such that

$$\sum_{t=2}^{T-1} \|\bar{x}_t^{(n)} - \mathbb{E}[\bar{x}_t^{(n)}|\mathcal{F}_t(\xi_{\mathrm{tr}})]\|_{r'} \leq \inf_{x^{(n)} \in S(\hat{\xi}^{(n)})} \sum_{t=2}^{T-1} \|x_t^{(n)} - \mathbb{E}[x_t^{(n)}|\mathcal{F}_t(\xi_{\mathrm{tr}})]\|_{r'} + \frac{1}{n}.$$

Then any subsequence of $(\bar{x}^{(n)})_{n \geq n_1}$ contains a further subsequence converging weakly in $L_{r'}$ to some element of $S(\xi)$ (due to [21, Theorem 2.5 and Remark 2.6]). Since $\mathbb{E}[\cdot|\mathcal{F}_t(\xi_{\mathrm{tr}})]$ and $\|\cdot\|_{r'}$ are continuous and lower semicontinuous with respect to the weak convergence in $L_{r'}$, respectively, we obtain after setting

$$\mathrm{li} := \liminf_{n\to\infty} \inf_{x^{(n)} \in S(\hat{\xi}^{(n)})} \sum_{t=2}^{T-1} \|x_t^{(n)} - \mathbb{E}[x_t^{(n)}|\mathcal{F}_t(\xi_{\mathrm{tr}})]\|_{r'}$$

the desired estimate

$$\mathrm{li} = \liminf_{n\to\infty} \sum_{t=2}^{T-1} \|\bar{x}_t^{(n)} - \mathbb{E}[\bar{x}_t^{(n)}|\mathcal{F}_t(\xi_{\mathrm{tr}})]\|_{r'} \geq \sum_{t=2}^{T-1} \|x_t^* - \mathbb{E}[x_t^*|\mathcal{F}_t(\xi_{\mathrm{tr}})]\|_{r'} \geq D_{\mathrm{f}}(\xi, \xi_{\mathrm{tr}}),$$

where x^* is some element in $S(\xi)$. □

The first part of the preceding result is related to [1, Theorem III.30]. If the assumptions of Proposition 5.4 are satisfied and if (i) $\hat{\xi}$ is *adapted to the filtration* of the original process ξ and (ii) represents a good *approximation to* ξ in the sense of $L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$, the tree generation using Algorithms 4.2 and 4.5, respectively, leads to a small distance $|v(\xi) - v(\xi_{\mathrm{tr}})|$ of optimal values. Moreover, the quantity

$$\Delta(\hat{\xi}, \xi_{\mathrm{tr}}) := \inf_{x \in S(\hat{\xi})} \sum_{t=2}^{T-1} \|x_t - \mathbb{E}[x_t|\mathcal{F}_t(\xi_{\mathrm{tr}})]\|_{r'} \qquad (36)$$

approximately represents an upper bound of $D_{\mathrm{f}}(\xi, \xi_{\mathrm{tr}})$. Next we show that conditions (i) and (ii) are satisfied on *some* probability space in case of *discretization* schemes for P.

Example 5.5 (Discretization) Let P be a Borel probability measure on $\Xi = \times_{t=1}^{T} \Xi_t$, where $\Xi_t \in \mathcal{B}(\mathbb{R}^d)$, $t = 1, \ldots, T$. We consider the probability space $(\Xi, \mathcal{B}(\Xi), P)$ and the identity mapping id of Ξ as original stochastic process. We assume that Ξ is bounded. To construct an approximation $\hat{\xi}$ of id, we first consider a sequence $(\mathcal{D}_t^{(n)})$ of Borel partitions of Ξ_t satisfying the following properties for any $t = 1, \ldots, T$:

(P1) $D_t \cap \tilde{D}_t = \emptyset$ for all $D_t, \tilde{D}_t \in \mathcal{D}_t^{(n)}$, $D_t \neq \tilde{D}_t$, $n \in \mathbb{N}$.
(P2) $\cup_{D_t \in \mathcal{D}_t^{(n)}} D_t = \Xi_t$ for all $n \in \mathbb{N}$.
(P3) $\delta_{t,n} := \sup_{D_t \in \mathcal{D}_t^{(n)}} \sup_{\xi_t, \tilde{\xi}_t \in D_t} |\xi_t - \tilde{\xi}_t| \underset{n \to \infty}{\longrightarrow} 0.$

In this way, we obtain a sequence $(\mathcal{P}^{(n)}) = (\{\times_{t=1}^{T} D_t : D_t \in \mathcal{D}_t^{(n)}, t = 1, \ldots, T\})$ of partitions of Ξ. Furthermore, we select elements

$$\hat{\xi}_t^{D_t,n} \in D_t \in \mathcal{D}_t^{(n)}$$

for all $D_t \in \mathcal{D}_t^{(n)}$, $n \in \mathbb{N}$ and $t = 1, \ldots, T$, and define $\hat{\xi}^{(n)} : \Xi \to \Xi$ by setting

$$\begin{aligned}
\hat{\xi}^{(n)}(\xi) &:= \sum_{\times_{t=1}^{T} D_t \in \mathcal{P}^{(n)}} (\hat{\xi}_1^{D_1,n}, \ldots, \hat{\xi}_T^{D_T,n}) \chi_{\times_{t=1}^{T} D_t}(\xi) \quad (\xi = (\xi_1, \ldots, \xi_T) \in \Xi) \\
&= \sum_{D_1 \in \mathcal{D}_1^{(n)}} \cdots \sum_{D_T \in \mathcal{D}_T^{(n)}} (\hat{\xi}_1^{D_1,n}, \ldots, \hat{\xi}_T^{D_T,n}) \prod_{t=1}^{T} \chi_{D_t}(\xi_t)
\end{aligned}$$

for some $n \in \mathbb{N}$. Next we verify that $\hat{\xi}^{(n)}$ is adapted to the filtration of id, i.e., to $\mathcal{F}_t(\mathrm{id}) = \sigma(\{B_1 \times \cdots \times B_t \times \Xi_{t+1} \times \cdots \times \Xi_T : B_\tau \in \mathcal{B}(\Xi_\tau), \tau = 1, \ldots, t\})$, $t = 1, \ldots, T$. By construction of $\hat{\xi}^{(n)}$ we have that

$$\mathcal{F}_t(\hat{\xi}^{(n)}) = \sigma(\{D_1 \times \cdots \times D_t \times \Xi_{t+1} \times \cdots \times \Xi_T : D_\tau \in \mathcal{D}_\tau^{(n)}, \tau = 1, \ldots, t\})$$

and, hence, $\mathcal{F}_t(\hat{\xi}^{(n)})$ is contained in $\mathcal{F}_t(\mathrm{id})$. It remains to consider the L_r-distance

$$\begin{aligned}
\|\mathrm{id} - \hat{\xi}^{(n)}\|_r^r &= \int_{\Xi} |\xi - \hat{\xi}^{(n)}(\xi)|^r P(d\xi) \\
&= \sum_{D_1 \times \cdots \times D_T \in \mathcal{P}^{(n)}} \int_{D_1 \times \cdots \times D_T} |(\xi_1, \ldots, \xi_T) - (\hat{\xi}_1^{D_1,n}, \ldots, \hat{\xi}_T^{D_T,n})|^r P(d\xi) \\
&\leq \sum_{D_1 \times \cdots \times D_T \in \mathcal{P}^{(n)}} \max_{t=1,\ldots,T} \delta_{n.t}^r P(\times_{t=1}^{T} D_t) = \max_{t=1\ldots,T} \delta_{n.t},
\end{aligned}$$

where the norm $|\cdot|$ on $\mathbb{R}^{Td}$ is defined by $|\xi| = \max_{t=1,\dots,T} |\xi_t|$. Hence, $\hat{\xi}^{(n)}$ is close to id if n is sufficiently large.

In general, it is difficult to determine the probabilities $P(\times_{t=1}^{T} D_t)$ for all $D_t \in \mathcal{D}_t^{(n)}$ and some $n \in \mathbb{N}$. However, if additional structure is available on ξ, the discretization scheme may be adapted such that the probabilities are computationally accessible. For example, let the stochastic process ξ be driven by a finite number of mutually independent $\mathbb{R}^{d_t}$-valued random variables z_t with probability distributions P_t, $t = 2, \dots, T$, i.e.,

$$\xi_t = g_t(\xi_1, \dots, \xi_{t-1}, z_t),$$

where the g_t, $t = 2, \dots, T$, denote certain measurable functions from $\mathbb{R}^{td} \times \mathbb{R}^{d_t}$ to $\mathbb{R}^d$ (see, e.g., [28,31,39,40]). Then there exists a measurable function G such that $\xi = G(z_2, \dots, z_T)$. If $\mathcal{D}_t^{(n)}$ is now a partition of the support of z_t in $\mathbb{R}^{d_t}$, $t = 2, \dots, T$, then $\hat{\xi}^{(n)}$ may be defined by

$$\begin{aligned}\hat{\xi}_t^{(n)} &= g_t(\hat{\xi}_1^{(n)}, \dots, \hat{\xi}_{t-1}^{(n)}, z_t^{(n)})\\ z_t^{(n)} &= \sum_{D_t \in \mathcal{D}_t^{(n)}} \hat{z}_t^{D_t,n} \chi_{D_t}\end{aligned}$$

where $\hat{z}_t^{D_t,n} \in D_t$, $t = 2, \dots, T$. The probability distribution of $\hat{\xi}^{(n)}$ is then known if $P_t(D_t)$ is known for all $D_t \in \mathcal{D}_t^{(n)}$, $t = 2, \dots, T$. This covers situations, where ξ is a Gaussian process or is given by certain time series models.

If, in Example 5.5, the partitions $\mathcal{D}_t^{(n+1)}$ represent a refinement of $\mathcal{D}_t^{(n)}$ for each $n \in \mathbb{N}$ and $t = 1, \dots, T$, the approximations $\hat{\xi}^{(n)}$ have the additional property

$$v(\hat{\xi}^{(n)}) \underset{n\to\infty}{\longrightarrow} v(\mathrm{id})$$

due to Proposition 5.4 (if the relevant assumptions are satisfied). This is not surprising when recalling the convergence results for discretizations in [30,31].

If $\hat{\xi} := \hat{\xi}^{(n)}$ is obtained by a discretization scheme based on some partition $\mathcal{D}_t^{(n)}$ for each $t = 2, \dots, T$, the action of the backward and forward Algorithms 4.2 and 4.5 may be interpreted as finding appropriate partitions and elements belonging to the subsets of each partition (for $t = 1, \dots, T$) such that ξ_{tr} is a good discrete approximation of the original stochastic process.

6 Bounds on the filtration distance

Let $\hat{\xi}$ be a (discrete) approximation of the original stochastic process ξ and ξ_{tr} be a process obtained by means of one of the tree construction approaches in Sects. 4.1 and 4.2, respectively. Let all processes be defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. So far we obtained estimates of the form (31). In this section, we derive estimates for $\Delta(\hat{\xi}, \xi_{\mathrm{tr}})$ (see (36)). Recall that the latter quantity may serve as an approximate upper bound of $D_{\mathrm{f}}(\xi, \xi_{\mathrm{tr}})$ if $\hat{\xi}$ is adapted to the filtration of the original process ξ and $\|\xi - \hat{\xi}\|_r$ is small.

We assume that conditions (A2) and (A3) of Sect. 3 are satisfied and that the stochastic processes $\hat{\xi}$ and ξ_{tr} belong to the ball $\{\tilde{\xi} \in L_r : \|\tilde{\xi} - \xi\|_r \le \delta\}$ (see (A2)). Hence, the solution set $S(\hat{\xi})$ is nonempty and we have

$$\Delta(\hat{\xi}, \xi_{\mathrm{tr}}) = \inf_{x \in S(\hat{\xi})} \begin{cases} \sum_{t=2}^{T-1} (\mathbb{E}[|x_t - \mathbb{E}[x_t|\mathcal{F}_t(\xi_{\mathrm{tr}})]|^{r'}])^{\frac{1}{r'}}, & 1 \le r' < \infty, \\ \sum_{t=2}^{T-1} \|x_t - \mathbb{E}[x_t|\mathcal{F}_t(\xi_{\mathrm{tr}})]\|_\infty, & r' = \infty. \end{cases} \tag{37}$$

Let ξ^i, $i = 1, \ldots, N$, denote the scenarios of $\hat{\xi}$ with probabilities p_i, $i = 1, \ldots, N$, and let $A_i := \hat{\xi}^{-1}(\{\xi^i\})$, $i = 1, \ldots, N$. Then $\{A_i\}_{i=1}^N$ is a partition of Ω, we have $\mathbb{P}(A_i) = p_i$, $i = 1, \ldots, N$ and $\hat{\xi}$ is of the form

$$\hat{\xi}_t = \sum_{i=1}^{N} \xi_t^i \chi_{A_i} \quad (t = 1, \ldots, T),$$

where χ_A denotes the characteristic function of a subset A of Ω. Furthermore, let I_t denote the index set of realizations of $\xi_{\mathrm{tr},t}$, $t = 1, \ldots, T$, as in the Sects. 4.1 and 4.2 and let $\mathcal{E}_t$ denote the families of nonempty elements of $\mathcal{F}_t(\xi_{\mathrm{tr}})$ that form partitions of Ω and generate the corresponding σ-fields. We set $E_{ti} := \{\omega \in \Omega : (\xi_{\mathrm{tr},1}(\omega), \ldots, \xi_{\mathrm{tr},t}(\omega)) = (\xi^i_{\mathrm{tr},1}, \ldots, \xi^i_{\mathrm{tr},t})\}$ for all $i \in I_t$ and $t = 1, \ldots, T$. Hence, we have $E_{ti} = \cup_{j \in \bar{I}_{t,i}} A_j$ and $\pi_t^i := \mathbb{P}(E_{ti}) = \sum_{j \in \bar{I}_{t,i}} p_j$ for $i \in I_t$, $t = 1, \ldots, T$. With x^i denoting the scenarios of $x \in S(\hat{\xi})$ we obtain

$$\mathbb{E}[x_t|E_{ts}] = \sum_{i \in I_t} \sum_{j \in \bar{I}_{t,i}} x_t^j \mathbb{E}[\chi_{A_j}|E_{ts}] = \sum_{j \in \bar{I}_{t,s}} x_t^j \mathbb{E}[\chi_{A_j}|E_{ts}] = \sum_{j \in \bar{I}_{t,s}} x_t^j \frac{p_j}{\pi_t^s} \tag{38}$$

for every $s \in I_t$. For $1 \le r' < \infty$ we have from (37)

$$\Delta(\hat{\xi}, \xi_{\mathrm{tr}}) = \inf_{x \in S(\hat{\xi})} \sum_{t=2}^{T-1} \left(\mathbb{E}\left[\left| \sum_{i=1}^{N} x_t^i \chi_{A_i} - \sum_{i \in I_t} \mathbb{E}[x_t|E_{ti}] \chi_{E_{ti}} \right|^{r'} \right] \right)^{\frac{1}{r'}}$$

and continue

$$\Delta(\hat{\xi}, \xi_{\mathrm{tr}}) = \inf_{x \in S(\hat{\xi})} \sum_{t=2}^{T-1} \left(\mathbb{E}\left[\left| \sum_{i=1}^{N} x_t^i \chi_{A_i} - \sum_{i \in I_t} \mathbb{E}[x_t|E_{ti}] \sum_{j \in \bar{I}_{t,i}} \chi_{A_j} \right|^{r'} \right] \right)^{\frac{1}{r'}}$$

$$= \inf_{x \in S(\hat{\xi})} \sum_{t=2}^{T-1} \left(\sum_{i \in I_t} \sum_{j \in \bar{I}_{t,i}} p_j \left| x_t^j - \mathbb{E}[x_t | E_{ti}] \right|^{r'} \right)^{\frac{1}{r'}}. \tag{39}$$

Analogously, we have for $r' = \infty$

$$\Delta(\hat{\xi}, \xi_{\text{tr}}) = \inf_{x \in S(\hat{\xi})} \sum_{t=2}^{T-1} \max_{i \in I_t} \max_{j \in \bar{I}_{t,i}} \left| x_t^j - \mathbb{E}[x_t | E_{ti}] \right|.$$

Starting from these representations of $\Delta(\hat{\xi}, \xi_{\text{tr}})$ the following estimates are valid.

Proposition 6.1 *Let* (A2) *and* (A3) *be satisfied and assume that* $\hat{\xi}$ *and* ξ_{tr} *belong to the ball* $\{\tilde{\xi} \in L_r : \|\tilde{\xi} - \xi\|_r \leq \delta\}$. *Let the stochastic process* $\hat{\xi}$ *have scenarios* ξ^i *with probabilities* p_i, $i = 1, \ldots, N$, *and* ξ_{tr} *be a scenario tree with index set* I_t *of realizations and scenario clusters* $\bar{I}_{t,i}$ *at* t. *Then we have*

$$\Delta(\hat{\xi}, \xi_{\text{tr}}) \leq \begin{cases} \sum_{t=2}^{T-1} \left(\sum_{i \in I_t} \sum_{j \in \bar{I}_{t,i}} \frac{p_j}{\pi_t^i} \left| \sum_{k \in \bar{I}_{t,i}} p_k (x_t^j - x_t^k) \right|^{r'} \right)^{\frac{1}{r'}}, & 1 \leq r' < \infty \\ \sum_{t=2}^{T-1} \max_{i \in I_t} \max_{j,k \in \bar{I}_{t,i}} |x_t^k - x_t^j|, & r' = \infty \end{cases} \tag{40}$$

$$\Delta(\hat{\xi}, \xi_{\text{tr}}) \leq K \begin{cases} \sum_{t=2}^{T-1} \left(\sum_{i \in I_t} \sum_{j \in I_{t,i}} p_j |x_t^j - x_t^i|^{r'} \right)^{\frac{1}{r'}}, & 1 \leq r' < \infty \\ \sum_{t=2}^{T-1} \max_{i \in I_t} \max_{j \in I_{t,i}} |x_t^j - x_t^i|, & r' = \infty \end{cases} \tag{41}$$

for any solution $x \in S(\hat{\xi})$ *and some constant* $K > 0$.

Proof Let $x \in S(\hat{\xi})$. Notice that $S(\hat{\xi})$ is nonempty according to (A2) and (A3). The proof is carried out for the case $1 \leq r' < \infty$. In case $r' = \infty$ the estimates follow by immediate modifications. To derive (40), we start from (39) and insert the representation (38) of the conditional expected values. This leads to

$$\Delta(\hat{\xi}, \xi_{\text{tr}}) \leq \sum_{t=2}^{T-1} \left(\sum_{i \in I_t} \sum_{j \in \bar{I}_{t,i}} \frac{p_j}{\pi_t^i} \left| \pi_t^i x_t^j - \sum_{k \in \bar{I}_{t,i}} p_k x_t^k \right|^{r'} \right)^{\frac{1}{r'}}.$$

and, thus, to (40). For the second estimate (41) we consider for any $i \in I_t$ the index $\alpha_t(i)$ defined by (21) and (26), respectively. Starting again from the representation

(39) of $\Delta(\hat{\xi}, \xi_{\mathrm{tr}})$ we get

$$\begin{aligned}
\Delta(\hat{\xi}, \xi_{\mathrm{tr}}) &\leq \sum_{t=2}^{T-1} \left(\sum_{i \in I_t} \sum_{j \in \bar{I}_{t,i}} p_j \left(|x_t^j - x_t^{\alpha_t(i)}| + \sum_{k \in \bar{I}_{t,i}} \frac{p_k}{\pi_t^i} |x_t^{\alpha_t(i)} - x_t^k| \right)^{r'} \right)^{\frac{1}{r'}} \\
&\leq \bar{K} \sum_{t=2}^{T-1} \left(\sum_{i \in I_t} \sum_{j \in \bar{I}_{t,i}} p_j \left(|x_t^j - x_t^{\alpha_t(i)}|^{r'} + \sum_{k \in \bar{I}_{t,i}} \left(\frac{p_k}{\pi_t^i} \right)^{r'} |x_t^{\alpha_t(i)} - x_t^k|^{r'} \right) \right)^{\frac{1}{r'}} \\
&\leq \bar{K} \sum_{t=2}^{T-1} \left(\sum_{i \in I_t} \left(\sum_{j \in \bar{I}_{t,i}} p_j |x_t^j - x_t^{\alpha_t(i)}|^{r'} + \sum_{k \in \bar{I}_{t,i}} p_k |x_t^{\alpha_t(i)} - x_t^k|^{r'} \right) \right)^{\frac{1}{r'}} \\
&= 2\bar{K} \sum_{t=2}^{T-1} \left(\sum_{i \in I_t} \sum_{j \in \bar{I}_{t,i}} p_j |x_t^j - x_t^{\alpha_t(i)}|^{r'} \right)^{\frac{1}{r'}} \\
&= 2\bar{K} \sum_{t=2}^{T-1} \left(\sum_{i \in I_t} \sum_{j \in I_{t,i}} p_j |x_t^j - x_t^i|^{r'} \right)^{\frac{1}{r'}},
\end{aligned}$$

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 $I_{t,i}$. □

Proposition 6.1 may be used to obtain a posteriori estimates of the filtration distance. Unfortunately, the solution process $x = \{x_t\}_{t=1}^T \in S(\hat{\xi})$ of the stochastic programming model (11) is only available at certain extra cost.

Estimates for $D_{\mathrm{f}}^*(\hat{\xi}, \xi_{\mathrm{tr}})$ (appearing in Proposition 5.2) or $D_{\mathrm{f},\infty}^*(\hat{\xi}, \xi_{\mathrm{tr}})$ may be obtained from (40) by taking the supremum of the right-hand side of (40) with respect to some ball containing $S(\hat{\xi})$. This leads to some matrix norm that might be evaluated or further estimated. Such matrix norms will be dealt with in a follow-up to this paper.

7 Numerical experience

The tree generation Algorithms 4.2 and 4.5 have been tested on data provided by the French electricity company Electricité de France (EDF). The data consists of a finite number of scenarios representing realizations of a bivariate stochastic process whose components are electrical load and water inflow to a hydro unit of a power generation system for a time horizon of 2 years. Equal probabilities are assigned to each scenario. Both stochastic processes appear as right-hand sides of linear constraints in stochastic electricity portfolio optimization models. 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 Discretization of the 2-year time horizon

Random variable	Discretization	Number time steps
Electrical load	3 per day	2,184
Water inflow	Weekly	104

Table 2 Dimension of the initial scenario fan

	Number
Scenarios	456
Time periods	2,184
Initial nodes	995,449

Tables 1 and 2 show the discretization of the data for the time horizon of 2 years and provide the number of scenarios, the total number of time periods and the corresponding number of nodes of the initial scenario set. 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.

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 set and the distribution of one of its scenarios endowed with unit mass. Since the stochastic optimization model of EDF was not accessible to us, the computation of bounds for the filtration distance and a comparison of optimal values was impossible. The test runs were performed on a PC with a 3 GHz Intel Pentium CPU and 1 GByte main memory.

Table 3 Results for backward tree construction without branching restriction

ε_{rel}	Scenarios	Nodes	Stages	Time (s)
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 4 Results for backward tree construction with weekly branching restriction

ε_{rel}	Scenarios	Nodes	Stages	Time (s)
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

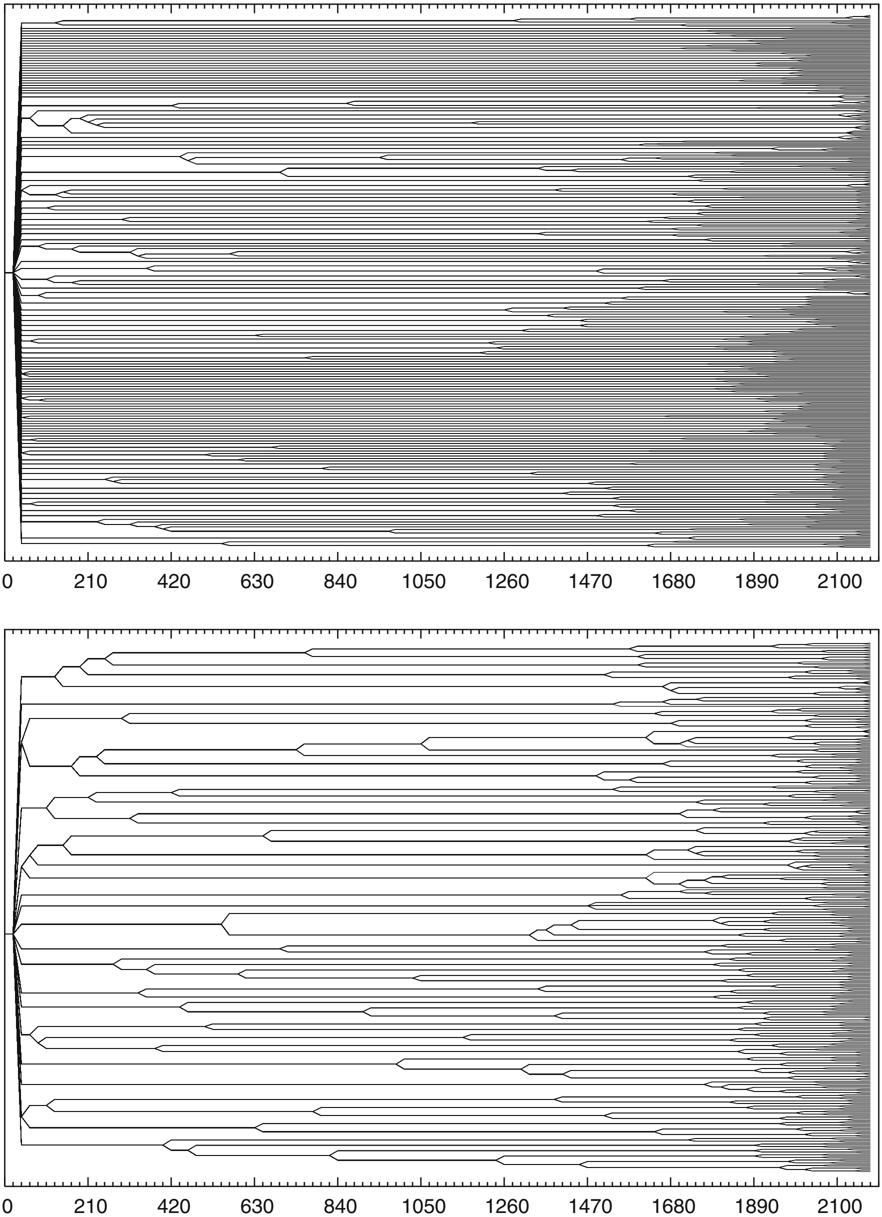

Fig. 3 Scenario trees obtained with $\varepsilon_{rel} = 0.2/0.5$ and weekly branching structure

7.1 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

Table 5 Results for forward tree construction without branching restriction

ε_{rel}	Scenarios	Nodes	Stages	Time (s)
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 6 Results for forward tree construction with weekly branching restriction

ε_{rel}	Scenarios	Nodes	Stages	Time (s)
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

$$\varepsilon_T = \varepsilon \cdot (1-q), \quad q \in (0,1) \quad \text{and} \quad \varepsilon_t = q \cdot \varepsilon_{t+1}, \quad t = T-1, \ldots, 2. \tag{42}$$

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 set 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 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 Fig. 3 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.

7.2 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

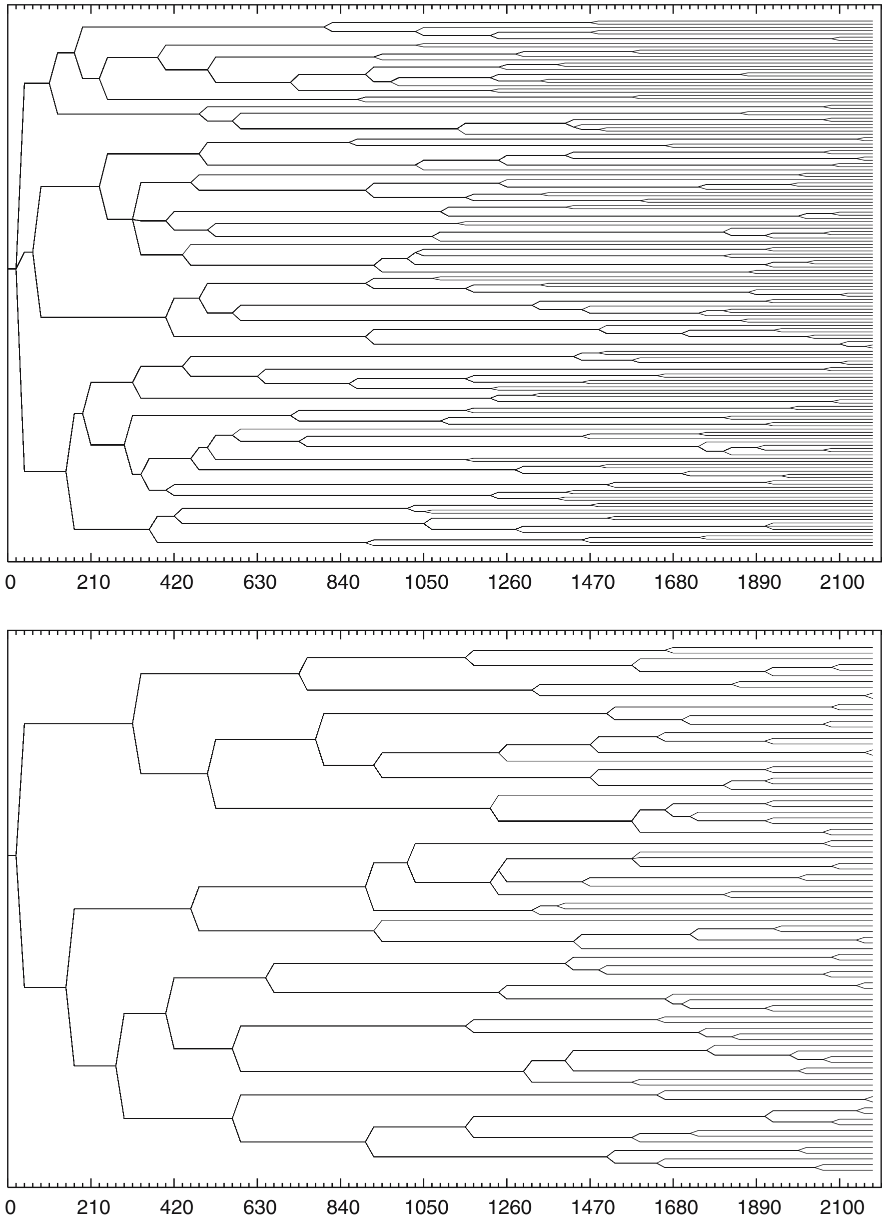

Fig. 4 Scenario trees obtained with $\varepsilon_{rel} = 0.4/0.5$ and weekly branching structure

$$\varepsilon_t = \frac{\varepsilon}{T}\left[1 + \overline{q}\left(\frac{1}{2} - \frac{t}{T}\right)\right], \quad t = 2, \ldots, T, \tag{43}$$

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 Algorithm 4.2 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 (22) in Sect. 4.1 is derived by employing the triangle inequality extensively and, hence, is more pessimistic than (29).

Figure 4 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.

8 Conclusions

In many applications of multistage stochastic programming the available statistical data allows to generate a (large) number of scenarios including their probabilities. This constitutes an initial approximation $\hat{\xi}$, which is considered as a good representation of the underlying stochastic process ξ. In this paper, we develop algorithms for generating scenario trees ξ_{tr} starting from $\hat{\xi}$ and provide conditions on $\hat{\xi}$ under which the distance $|v(\xi) - v(\xi_{\mathrm{tr}})|$ of optimal values of the stochastic programming model gets small. The theoretical results rely on a stability result (Theorem 3.1) for multistage stochastic programs. The conditions on $\hat{\xi}$ are justified for two relevant cases in applications, namely, if $\hat{\xi}$ is obtained by sampling from a discrete probability distribution or by discretization schemes for general probability distributions. In other cases, the algorithms are considered as heuristics for scenario tree generation. Some computational experience is provided for generating load-inflow scenario trees in electric power management. Further experience is reported in [10,12,29] for multistage stochastic programs in electricity and airline revenue management.

Acknowledgments This work was supported by the DFG Research Center MATHEON "Mathematics for key technologies" in Berlin, the BMBF under the grant 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. Further thanks are due to two anonymous referees for their constructive criticism and many helpful suggestions.

References

1. Barty, K.: Contributions à la discrétisation des contraintes de mesurabilité pour les problèmes d'optimisation stochastique. Thèse de Doctorat, École Nationale des Ponts et Chaussées, (2004)

2. Casey, M., Sen, S.: The scenario generation algorithm for multistage stochastic linear programming. Math. Oper. Res. **30**, 615–631 (2005)
3. Chiralaksanakul, A.; Morton, D.P.: Assessing policy quality in multi-stage stochastic programming. Stochastic Programming E-Print Series 12–2004 (<www.speps.org>).
4. Corvera Poiré, X.: Model Generation and Sampling Algorithms for Dynamic Stochastic Programming. PhD Thesis, Department of Mathematics, University of Essex (1995)
5. Dempster, M.A.H.: Sequential importance sampling algorithms for dynamic stochastic programming. Zap. Nauchn. Semin. POMI **312**, 94–129 (2004)
6. Dudley, R.M.: Real Analysis and Probability. Chapman & Hall, New York (1989)
7. Dunford, N., Schwartz, J.T.: Linear Operators, Part I: General Theory. Wiley Classics Library, New York (1988)
8. Dupačová, J., Consigli, G., Wallace, S.W.: Scenarios for multistage stochastic programs. Ann. Oper. Res. **100**, 25–53 (2000)
9. Dupačová, J., Gröwe-Kuska, N., Römisch, W.: Scenario reduction in stochastic programming: an approach using probability metrics. Math. Program. **95**, 493–511 (2003)
10. Eichhorn, A.; Römisch, W.; Wegner, I.: Mean-risk optimization of electricity portfolios using multiperiod polyhedral risk measures, IEEE St. Petersburg Power Tech (2005)
11. Fetter, H.: On the continuity of conditional expectations. J. Math. Anal. Appl. **61**, 227–231 (1977)
12. Fleten, S.-E.; Kristoffersen, T.K.: Short-term hydropower production planning by stochastic programming. Comput. Oper. Res. (to appear)
13. Frauendorfer, K.: Barycentric scenario trees in convex multistage stochastic programming. Math. Programm. Ser. B **75**, 277–293 (1996)
14. Givens, C.R., Shortt, R.M.: A class of Wasserstein metrics for probability distributions. Michigan Math. J. **31**, 231–240 (1984)
15. Graf, S., Luschgy, H.: Foundations of Quantization for Probability Distributions. Lecture Notes in Mathematics, vol. 1730. Springer, Berlin (2000)
16. 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, pp. 69–78. VDI-Verlag, Düsseldorf (2001)
17. Gröwe-Kuska, N., Heitsch, H., Römisch, W.: Scenario reduction and scenario tree construction for power management problems. In: Borghetti, A., Nucci, C.A., Paolone, M. (eds.) IEEE Bologna Power Tech Proceedings (2003)
18. Gröwe-Kuska, N., Römisch, W.: Stochastic unit commitment in hydro-thermal power production planning. In: Wallace, S.W., Ziemba, W.T. (eds.) Chapter 30 in Applications of Stochastic Programming, MPS-SIAM Series in Optimization (2005)
19. Heitsch, H., Römisch, W.: Scenario reduction algorithms in stochastic programming. Comput. Optim. Appl. **24**, 187–206 (2003)
20. Heitsch, H., Römisch, W.: Generation of multivariate scenario trees to model stochasticity in power management. IEEE St. Petersburg Power Tech (2005)
21. Heitsch, H., Römisch, W.: Stability and scenario trees for multistage stochastic programs, Preprint 324, DFG Research Center Matheon. In: Dantzig, G., Infanger, G. (eds.) Mathematics for key technologies, 2006 and submitted to Stochastic Programming—The State of the Art (2006)
22. Heitsch, H., Römisch, W., Strugarek, C.: Stability of multistage stochastic programs. SIAM J. Optim. **17**, 511–525 (2006)
23. Higle, J.L., Rayco, B., Sen, S.: Stochastic scenario decomposition for multistage stochastic programs. Ann. Oper. Res. (submitted)
24. Hochreiter, R., Pflug, G. Ch.: Financial scenario generation for stochastic multi-stage decision processes as facility location problems. Ann. Oper. Res. **152**, 257–272 (2007)
25. Høyland, K., Wallace, S.W.: Generating scenario trees for multi-stage decision problems. Manage. Sci. **47**, 295–307 (2001)
26. Høyland, K., Kaut, M., Wallace, S.W.: A heuristic for moment-matching scenario generation. Comput. Optim. Appl. **24**, 169–185 (2003)
27. Kaut, M., Wallace, S.W.: Evaluation of scenario-generation methods for stochastic programming. Stochastic Programming E-Print Series 14–2003 (<www.speps.org>).
28. Kuhn, D.: Generalized Bounds for Convex Multistage Stochastic Programs. Lecture Notes in Economics and Mathematical Systems, vol. 548. Springer, Berlin (2005)

29. Möller, A., Römisch, W., Weber, K.: Airline network revenue management by multistage stochastic programming. Comput. Manage. Sci. (to appear)
30. Pennanen, T.: Epi-convergent discretizations of multistage stochastic programs. Math. Oper. Res. **30**, 245–256 (2005)
31. Pennanen, T.: Epi-convergent discretizations of multistage stochastic programs via integration quadratures. Stochastic Programming E-Print Series 19–2004 (<www.speps.org>). Math. Programm. (to appear)
32. Pflug, G.Ch.: Scenario tree generation for multiperiod financial optimization by optimal discretization. Math. Program. **89**, 251–271 (2001)
33. Rachev, S.T.: Probability Metrics and the Stability of Stochastic Models. Wiley, New York (1991)
34. Rachev, S.T., Römisch, W.: Quantitative stability in stochastic programming: The method of probability metrics. Math. Oper. Res. **27**, 792–818 (2002)
35. Rachev, S.T., Rüschendorf, L.: Mass Transportation Problems, vols. I and II. Springer, Berlin (1998)
36. Rachev, S.T., Schief, A.: On L_p-minimal metrics. Prob. Math. Stat. **13**, 311–320 (1992)
37. Robinson, S.M.: Some continuity properties of polyhedral multifunctions. Math. Program. Study **14**, 206–214 (1981)
38. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, Berlin (1998)
39. Römisch, W.: Kennwertmethoden für stochastische Volterrasche Integralgleichungen. Wiss. Zeitschr. Humb.-Univ. Berlin, Math.-Nat. R. **XXVIII**, 523–533 (1979)
40. Römisch, W.: On the approximate solution of random operator equations. Wiss. Zeitschr. Humb.-Univ. Berlin, Math.-Nat. R. **XXX**, 455–462 (1981)
41. Römisch, W: Stability of stochastic programming problems. In: Ruszczyński, A., Shapiro, A. (eds.) Stochastic Programming, Handbooks in Operations Research and Management Science, vol 10., pp. 483–554. Elsevier, Amsterdam (2003)
42. Römisch, W., Schultz, R.: Stability analysis for stochastic programs. Ann. Oper. Res. **30**, 241–266 (1991)
43. Ruszczyński, A., Shapiro, A. (eds.): Stochastic Programming, Handbooks in Operations Research and Management Science, vol. 10. Elsevier, Amsterdam (2003)
44. Shapiro, A.: Inference of statistical bounds for multistage stochastic programming problems. Math. Meth. Oper. Res. **58**, 57–68 (2003)
45. Shapiro, A.: On complexity of multistage stochastic programs. Oper. Res. Lett. **34**, 1–8 (2006)