ORIGINAL PAPER

Scenario tree reduction for multistage stochastic programs

Holger Heitsch · Werner Römisch

Abstract A framework for the reduction of scenario trees as inputs of (linear) multistage stochastic programs is provided such that optimal values and approximate solution sets remain close to each other. The argument is based on upper bounds of the L_r-distance and the filtration distance, and on quantitative stability results for multistage stochastic programs. The important difference from scenario reduction in two-stage models consists in incorporating the filtration distance. An algorithm is presented for selecting and removing nodes of a scenario tree such that a prescribed error tolerance is met. Some numerical experience is reported.

Keywords Stochastic programming · Scenario reduction · Scenario tree · Multistage

1 Introduction

Numerical methods for solving applied stochastic programming models (in finance, production, energy, transportation, etc.) mostly rely on approximating the underlying probability distribution by a finitely discrete probability measure. This approximation technique reduces the original infinite-dimensional optimization problem to a finite-dimensional program. To avoid that these optimization problems are too high-dimensional, a scenario reduction methodology was suggested in Dupačová et al. (2003) and further developed in Heitsch and Römisch (2003, 2007). These scenario reduction methods are based on quantitative stability results for stochastic programs (see the survey, Römisch 2003, and the recent supplement, Römisch and Wets 2007, for two-stage models with random recourse) and on the use of distances of probability distributions relying on Monge–Kantorovich mass transportation problems

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

⚫ Springer

(Rachev and Rüschendorf 1998). Although optimal scenario reduction problems are combinatorial optimization models of k-median type, and, hence, NP-hard, the forward and backward heuristics suggested in Dupačová et al. (2003), Heitsch and Römisch (2003) and refined in Heitsch and Römisch (2007) provide encouraging results and are often used in practical applications. The general idea was recently extended in Henrion et al. (2007, 2008) to chance constrained and mixed-integer two-stage stochastic programming models.

An important class of stochastic programs for practical applications are models with *measurability constraints*, e.g., multistage stochastic programs. Recently, the stability behavior of multistage linear stochastic programs was studied in Heitsch et al. (2006). Its main result states that the distance of optimal values of original and approximate models can be bounded by the L_r-distance (for some $r \geq 1$) and a so-called *filtration distance* of the underlying stochastic processes. The main computational approach for solving multistage models consists in approximating the original stochastic process by a process having finitely many scenarios exhibiting tree structure. Presently, several approaches for the generation of such *scenario trees* are available. Here, we refer to the survey (Dupačová et al. 2000) and to the original papers (Casey and Sen 2005; Heitsch and Römisch 2008; Hochreiter and Pflug 2007; Høyland et al. 2003; Kuhn 2005, 2008; Pennanen 2009). If such a scenario tree is available, it may again be of interest to reduce it by deleting some of its nodes. Due to the stability behavior of multi-stage models, it is argued in Heitsch et al. (2006, Example 2.7) that scenario tree reduction in multistage models should be based on L_r-distances *as well as* on filtration distances.

In this paper, we take up the latter issue and develop a sound theoretical basis for scenario tree reduction in multistage stochastic programming models. To do so, we review stability results for the multistage situation (in Sect. 2) and derive new bounds for both the L_r-distance and the filtration distances between a scenario tree and its reduced version (in Sect. 3). These bounds motivate algorithms for reducing scenario trees. In Sect. 4 we present a specific algorithm based on *recursive single node reduction*. In Sect. 5 we report on numerical experience of the tree reduction algorithm and show that its outcomes strongly depends on the use of both types of distances, namely, the L_r-distance and the filtration distance. In particular, the results indicate that applying the scenario reduction techniques from Heitsch and Römisch (2003, 2007) (i.e., methods that are only based on the L_r-distance) to the multistage situation is *not appropriate*.

2 A review of stability in multistage stochastic programming

Let $\xi = \{\xi_t\}_{t=1}^T$ be a stochastic process defined on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and with ξ_t taking values in $\mathbb{R}^d$. It is assumed that this process enters an optimization model and that the (stochastic) decision x_t at t maps from Ω to $\mathbb{R}^{m_t}$ is *nonanticipative*, i.e., depends only on $\xi^t := (\xi_1, \ldots, \xi_t)$. The latter property is equivalent to the measurability constraint stating that x_t is measurable with respect to the σ-field $\mathscr{F}_t(\xi) \subseteq \mathscr{F}$ generated by ξ^t. We assume that ξ_1 is deterministic, i.e., that $\mathscr{F}_1(\xi) = \{\emptyset, \Omega\}$. Then the stochastic process ξ is accompanied by a filtration $(\mathscr{F}_t(\xi))_{t=1}^T$ of σ-fields satisfying

$$\mathscr{F}_1(\xi) = \{\emptyset, \Omega\} \subseteq \cdots \subseteq \mathscr{F}_t(\xi) \subseteq \mathscr{F}_{t+1}(\xi) \subseteq \cdots \subseteq \mathscr{F}_T(\xi) \subseteq \mathscr{F}.$$

We consider the linear multistage stochastic programming model

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

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 $\mathscr{F}_1(\xi) \subsetneq \mathscr{F}_t(\xi) \subsetneq \mathscr{F}_T(\xi)$ for some $1 < t < T$. The presence of such qualitatively different constraints constitutes the origin of both the theoretical and computational challenges of multistage models.

Next we record results of the recent papers (Heitsch et al. 2006; Heitsch and Römisch 2008). We assume that the stochastic input process ξ belongs to the Banach space $L_r(\Omega, \mathscr{F}, \mathbb{P}; \mathbb{R}^s)$ with $s := Td$ and $r \geq 1$. The multistage model (1) is regarded as an optimization problem in the space $L_{r'}(\Omega, \mathscr{F}, \mathbb{P}; \mathbb{R}^m)$ with $m = \sum_{t=1}^{T} m_t$ and endowed with the norm

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

where the number r' is defined by

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

The choice of r and the definition of r' are motivated by the knowledge on existing moments of the input process and 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 $t = 1, \ldots, T$).

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

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

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

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

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

$$\min\{F(\xi, x) : x \in \mathscr{X}(\xi)\}. \tag{3}$$

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

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

denote the *α-approximate solution set* and the *solution set* of the stochastic program (3) with input ξ, respectively.

The following conditions are imposed on (3):

(A1) $\xi \in L_r(\Omega, \mathscr{F}, \mathbb{P}; \mathbb{R}^s)$, 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 $\tilde{\xi} \in L_r(\Omega, \mathscr{F}, \mathbb{P}; \mathbb{R}^s)$ with $\|\tilde{\xi} - \xi\|_r \leq \delta$, any $t = 2, \ldots, T$ and any $x_1 \in \mathscr{X}_1(\tilde{\xi}_1)$, $x_\tau \in \mathscr{X}_\tau(x_{\tau-1}; \tilde{\xi}_\tau)$, $\tau = 2, \ldots, t-1$, there exists an $\mathscr{F}_t(\tilde{\xi})$-measurable $x_t \in \mathscr{X}_t(x_{t-1}; \tilde{\xi}_t)$ (*relatively complete recourse locally around ξ*).
(A3) The optimal values $v(\tilde{\xi})$ of (3) 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 $L_\infty(\Omega, \mathscr{F}, \mathbb{P}; \mathbb{R}^m)$ such that $S_\alpha(\tilde{\xi})$ is contained in B for all $\tilde{\xi} \in L_r(\Omega, \mathscr{F}, \mathbb{P}; \mathbb{R}^s)$ with $\|\tilde{\xi} - \xi\|_r \leq \delta$.

The following stability result states that multistage models behave stable at some stochastic input process if both its probability distribution and its filtration are approximated simultaneously in terms of the L_r-distance and of one of the filtration distances

$$D_{\mathrm{f},\infty}(\xi, \tilde{\xi}) := \sup_{\|x\|_\infty \leq 1} \sum_{t=2}^{T-1} \|\mathbb{E}[x_t | \mathscr{F}_t(\xi)] - \mathbb{E}[x_t | \mathscr{F}_t(\tilde{\xi})]\|_{r'}, \tag{4}$$

$$D_{\mathrm{f},\infty}^*(\xi, \tilde{\xi}) := \sup_{\|x\|_\infty \leq 1} \sum_{t=2}^{T} \|\mathbb{E}[x_t | \mathscr{F}_t(\xi)] - \mathbb{E}[x_t | \mathscr{F}_t(\tilde{\xi})]\|_{r'}, \tag{5}$$

where $\mathscr{F}_t(\xi)$ and $\mathscr{F}_t(\tilde{\xi})$ denote the σ-fields generated by ξ^t and $\tilde{\xi}^t$, respectively, and $\mathbb{E}[\cdot | \mathscr{F}_t(\xi)]$ and $\mathbb{E}[\cdot | \mathscr{F}_t(\tilde{\xi})]$ the corresponding conditional expectations.

Fig. 1 Example scenario tree ξ with $|I_t| = 6$, $|I_{t,i}| = 3$ and $|I_T| = 11$
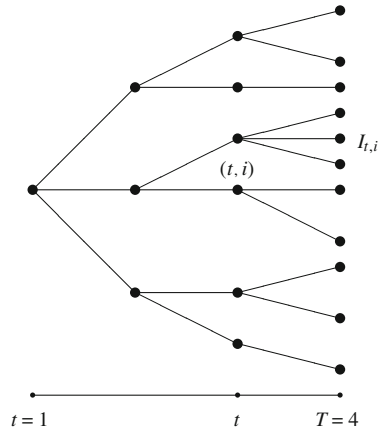
Theorem 1 *Let* (A1), (A2) *and* (A3) *be satisfied and* X_1 *be bounded. Then there exist positive constants* L *and* δ *such that the estimate*

$$|v(\xi) - v(\tilde{\xi})| \le L(\|\xi - \tilde{\xi}\|_r + D_{f,\infty}(\xi, \tilde{\xi})) \tag{6}$$

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

Furthermore, if the solution sets $S(\xi)$ *and* $S(\tilde{\xi})$ *are nonempty, there exist* $\bar{L} > 0$ *and* $\bar{\varepsilon} > 0$ *such that the estimate*

$$dl_\infty(S_\varepsilon(\xi), S_\varepsilon(\tilde{\xi})) \le \frac{\bar{L}}{\varepsilon}(\|\xi - \tilde{\xi}\|_r + D_{f,\infty}^*(\xi, \tilde{\xi})) \tag{7}$$

holds for any $\varepsilon \in (0, \bar{\varepsilon})$. *Here,* dl_∞ *denotes the Pompeiu–Hausdorff distance of bounded subsets of* $L_{r'}$.

The first part of Theorem 1 is essentially Heitsch et al. (2006, Theorem 2.1), where compared to Heitsch et al. (2006), condition (A3) allows to make use of the filtration distances $D_{f,\infty}$ or $D_{f,\infty}^*$ (cf. the discussion in Heitsch and Römisch 2008, Sect. 3). The second part of Theorem 1 is proved in Heitsch and Römisch (2006).

Finally, we mention that Theorem 1 remains valid if the expectation $\mathbb{E}$ in the objective of (1) is replaced by a multi-period polyhedral risk functional satisfying a certain uniform level boundedness property (see Eichhorn and Römisch 2008). Multi-period polyhedral risk functionals and their incorporation into multi-stage stochastic programming models are studied in Eichhorn and Römisch (2005).

3 Bounding the L_r-minimal and filtration distance

Let $\xi = \{\xi_t\}_{t=1}^T$ be a stochastic process on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ having a finite number of scenarios ξ^i with probabilities p_i, $i = 1, \ldots, N$, in form of a scenario tree. Let I_t denote the index set of realizations of ξ_t. If we set $A_{ti} := (\xi_1, \ldots, \xi_t)^{-1}(\{(\xi_1^i, \ldots, \xi_t^i)\})$ for every $i \in I_t$, the system $\{A_{ti}\}_{i \in I_t}$ is a partition of

Fig. 2 Example scenario tree ξ^{red} with $|I_t^{\mathrm{red}}| = 5$, $|J_{t,j}| = 2$ and $|I_T^{\mathrm{red}}| = 8$
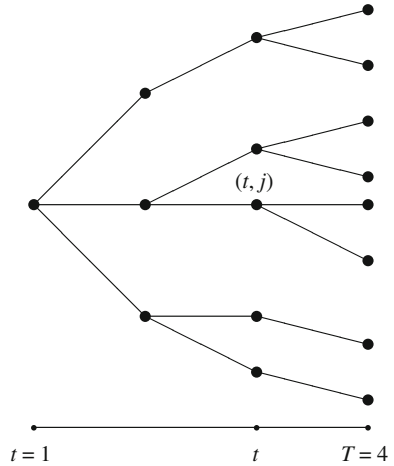
Ω and generates the σ-field $\mathscr{F}_t(\xi)$. We set $p_t^i := \mathbb{P}(A_{ti})$, $i \in I_t$, $t = 1, \ldots, T$, and have, in particular, that $I_T = \{1, \ldots, N\}$ and $p_T^i = p_i$ for every $i \in I_T$. Furthermore, we have

$$\xi_t = \sum_{i \in I_t} \xi_t^i \, \mathbb{1}_{A_{ti}} \quad (t = 1, \ldots, T),$$

where $\mathbb{1}_A$ denotes the characteristic function of a subset A of Ω. The tree structure of ξ implies that I_1 is a singleton and that

$$I_1 \subseteq I_2 \subseteq \cdots \subseteq I_t \subseteq I_{t+1} \subseteq \cdots \subseteq I_T = \{1, \ldots, N\}$$

holds. Moreover, if $I_{t,i} \subseteq I_{t+1}$ denotes the index set of successors to ξ_t^i at $t + 1$, the relations

$$p_t^i = \sum_{j \in I_{t,i}} p_{t+1}^j \quad (i \in I_t)$$

are valid for every $t = 1, \ldots, T - 1$. Any node of the tree corresponds to a pair $(t, i) \in \{1, \ldots, T\} \times I_t$ (Fig. 1).

Now, let ξ^{red} be a stochastic process on $(\Omega, \mathscr{F}, \mathbb{P})$ that we regard as reduced scenario tree obtained from ξ. If I_t^{red} denotes the index set of realizations of ξ_t^{red}, the latter means

$$I_1^{\mathrm{red}} = I_1 \quad \text{and} \quad I_t^{\mathrm{red}} \subseteq I_t \quad (t = 2, \ldots, T),$$

where at least for some $t \in \{2, \ldots, T\}$ we have $I_t^{\mathrm{red}} \subset I_t$. Let $\xi^{j,\mathrm{red}}$, $j \in I_T^{\mathrm{red}}$, denote the scenarios of ξ^{red}. Let us further denote by $J_t := I_t \setminus I_t^{\mathrm{red}}$ the index set of all withdrawn realizations at time t and by E_{tj} the set $E_{tj} := (\xi_1^{\mathrm{red}}, \ldots, \xi_t^{\mathrm{red}})^{-1}(\{(\xi_1^j, \ldots, \xi_t^j)\})$ for every $j \in I_t^{\mathrm{red}}$. The system $\{E_{tj}\}_{j \in I_t^{\mathrm{red}}}$ forms a partition of Ω and generates the

σ-field $\mathscr{F}_t(\xi^{\text{red}})$. Moreover, for every $j \in I_t^{\text{red}}$, let $J_{t,j} \subset I_t$ denote the index set such that $\xi_t^i = \xi_t^j = \xi_t^{j,\,\text{red}}$ holds for all $i \in J_{t,j}$, i.e., the index set of all scenarios in I_t which have been identified with ξ_t^j during the reduction process (Fig. 2). The index sets $\{J_{t,j}\}_{j \in I_t^{\text{red}}}$ form a partition of I_t and it holds

$$E_{tj} = \bigcup_{i \in J_{t,j}} A_{ti},$$

$$\pi_t^j := \mathbb{P}(E_{tj}) = \sum_{i \in J_{t,j}} p_t^i \quad (j \in I_t^{\text{red}}).$$

If the assumptions of Theorem 1 are satisfied for some $1 \leq r < +\infty$, the distances of optimal values and ε-approximate solution sets get small if both distances

$$\|\xi - \xi^{\text{red}}\|_r \quad \text{and} \quad D_{f,\infty}^*(\xi, \xi^{\text{red}}) \tag{8}$$

are small. Hence, if a tolerance $\varepsilon > 0$ is given, it is reasonable to require

$$w_1 \|\xi - \xi^{\text{red}}\|_r + w_2 D_{f,\infty}^*(\xi, \xi^{\text{red}}) \leq \varepsilon, \tag{9}$$

where w_i, $i = 1, 2$, denote positive weighting factors such that $w_1 \|\xi - \xi^{\text{red}}\|_r$ and $w_2 D_{f,\infty}^*(\xi, \xi^{\text{red}})$ belong to $(0, 1]$.

The condition (9) appears as a natural condition for reducing the scenario tree ξ. A canonical choice for the factors w_1 and w_2 is obtained by selecting $i^* \in I_T$ such that the corresponding scenario ξ^{i^*} represents the best approximation of ξ with respect to the L_r-distance. More precisely, if ξ^* denotes the corresponding deterministic scenario process, i.e., $\xi^*(\omega) = \xi^{i^*}$, for all $\omega \in \Omega$, we have

$$\|\xi - \xi^*\|_r \leq \|\xi - \tilde{\xi}^*\|_r,$$

for all deterministic processes $\tilde{\xi}^*$ consisting of only one given scenario, i.e., for all processes with $\tilde{\xi}^*(\omega) = \xi^i$, for all $\omega \in \Omega$, where $i \in I_T$. Then the weighting factors are defined by

$$w_1 = \frac{1}{\|\xi - \xi^*\|_r} \quad \text{and} \quad w_2 = \frac{1}{D_{f,\infty}^*(\xi, \xi^*)}. \tag{10}$$

Next we derive bounds for both distances in (9). They are of the form

$$\|\xi - \xi^{\text{red}}\|_r = \left(\sum_{t=2}^{T} \mathbb{E}\left[|\xi_t - \xi_t^{\text{red}}|^r \right] \right)^{\frac{1}{r}}, \tag{11}$$

$$D_{f,\infty}^*(\xi, \xi^{\text{red}}) = \sum_{t=2}^{T} \sup_{\substack{x_t \in L_\infty(\mathscr{F}_t(\xi)) \\ \|x_t\|_\infty \leq 1}} \|x_t - \mathbb{E}[x_t | \mathscr{F}_t(\xi^{\text{red}})]\|_{r'}, \tag{12}$$

where $L_\infty(\mathscr{F}_t(\xi)) = L_\infty(\Omega, \mathscr{F}_t(\xi), \mathbb{P}; \mathbb{R}^{m_t})$. The latter formula is a consequence of the identity

$$
\begin{aligned}
D_{f,\infty}^*(\xi, \xi^{\text{red}}) &= \sup_{\|x\|_\infty \leq 1} \sum_{t=2}^T \|\mathbb{E}[x_t | \mathscr{F}_t(\xi)] - \mathbb{E}[x_t | \mathscr{F}_t(\xi^{\text{red}})]\|_{r'} \\
&= \sup_{\|x\|_\infty \leq 1} \sum_{t=2}^T \|\mathbb{E}[x_t | \mathscr{F}_t(\xi)] - \mathbb{E}[\mathbb{E}[x_t | \mathscr{F}_t(\xi)] | \mathscr{F}_t(\xi^{\text{red}})]\|_{r'} \\
&= \sup_{\substack{\|x\|_\infty \leq 1 \\ x_t \in L_\infty(\mathscr{F}_t(\xi))}} \sum_{t=2}^T \|x_t - \mathbb{E}[x_t | \mathscr{F}_t(\xi^{\text{red}})]\|_{r'},
\end{aligned}
$$

which is due to the inclusion $\mathscr{F}_t(\xi^{\text{red}}) \subseteq \mathscr{F}_t(\xi)$, and the fact that the condition $\|x\|_\infty \leq 1$ is equivalent to $\|\mathbb{E}[x_t | \mathscr{F}_t(\xi)]\|_\infty \leq 1$ for every $t = 1, \ldots, T$.

To derive explicit expressions for (12), we use the measurability of x_t with respect to $\mathscr{F}_t(\xi^{\text{red}})$ and denote the scenarios x_t by x_t^i, $i \in I_t$, for every $t = 2, \ldots, T$. We obtain for the conditional expected values

$$
\mathbb{E}[x_t | E_{tj}] = \frac{\int_{E_{tj}} x_t \mathbb{P}(d\omega)}{\mathbb{P}(E_{tj})} = \frac{1}{\pi_t^j} \sum_{i \in J_{t,j}} p_t^i x_t^i
$$

for every $j \in I_t^{\text{red}}$ and $t = 2, \ldots, T$. For $1 \leq r' < \infty$ we get from (12)

$$
D_{f,\infty}^*(\xi, \xi^{\text{red}}) = \sum_{t=2}^T \sup_{\|x_t\|_\infty \leq 1} \left(\mathbb{E}\left[\left| \sum_{i \in I_t} x_t^i \mathbb{1}_{A_{ti}} - \sum_{j \in I_t^{\text{red}}} \mathbb{E}[x_t | E_{tj}] \mathbb{1}_{E_{tj}} \right|^{r'} \right] \right)^{\frac{1}{r'}}
$$

and continue

$$
\begin{aligned}
D_{f,\infty}^*(\xi, \xi^{\text{red}}) &= \sum_{t=2}^T \sup_{\|x_t\|_\infty \leq 1} \left(\mathbb{E}\left[\left| \sum_{j \in I_t^{\text{red}}} \sum_{i \in J_{t,j}} x_t^i \mathbb{1}_{A_{ti}} - \sum_{j \in I_t^{\text{red}}} \mathbb{E}[x_t | E_{tj}] \sum_{i \in J_{t,j}} \mathbb{1}_{A_{ti}} \right|^{r'} \right] \right)^{\frac{1}{r'}} \\
&= \sum_{t=2}^T \sup_{\|x_t\|_\infty \leq 1} \left(\sum_{j \in I_t^{\text{red}}} \sum_{i \in J_{t,j}} p_t^i \left| x_t^i - \frac{1}{\pi_t^j} \sum_{k \in J_{t,j}} p_t^k x_t^k \right|^{r'} \right)^{\frac{1}{r'}}.
\end{aligned} \tag{13}
$$

3.1 The L_r-distance

Now we start to discuss the L_r-distance $\|\xi - \xi^{\text{red}}\|_r$ between the two processes ξ and ξ^{red}. According to our notations we directly obtain from (11)

$$\|\xi - \xi^{\text{red}}\|_r^r = \sum_{t=2}^{T} \int_{\Omega} |\xi_t - \xi_t^{\text{red}}|^r = \sum_{t=2}^{T} \sum_{j \in I_t^{\text{red}}} \sum_{i \in J_{t,j}} \int_{A_{ti}} |\xi_t - \xi_t^{\text{red}}|^r$$

$$= \sum_{t=2}^{T} \sum_{j \in I_t^{\text{red}}} \sum_{i \in J_{t,j}} p_t^i |\xi_t^i - \xi_t^j|^r. \tag{14}$$

This means that the L_r-distance between a given process and a reduced one depends on the probabilities of all withdrawn scenario components and on their distances to some of the remaining scenario components.

3.2 The filtration distance

Next we derive an estimate for $D_{\text{f},\infty}^*(\xi, \xi^{\text{red}})$ given by (13) in case of $r' < \infty$ and $\|x_t\|_\infty := \max_{i \in I_t} |x_t^i|$ for every $x_t \in L_\infty(\mathscr{F}_t(\xi))$.

Proposition 1 *Consider the $\ell_{r'}$-norms*

$$|y_t| := \left(\sum_{s=1}^{m_t} |y_{t,s}|^{r'} \right)^{\frac{1}{r'}}$$

in $\mathbb{R}^{m_t}$ for every $t = 2, \ldots, T$. Then we have

$$D_{\text{f},\infty}^*(\xi, \xi^{\text{red}}) \leq \max_{t=2,\ldots,T} 2m_t^{\frac{1}{r'}} \sum_{t=2}^{T} \left(\sum_{j \in I_t^{\text{red}}} \max \left\{ f_{t,r'}^j \left(\sum_{i \in J} p_t^i \right) : J \subset J_{t,j} \right\} \right)^{\frac{1}{r'}}, \tag{15}$$

where the function $f_{t,r'}^j$ is defined by $f_{t,r'}^j(p) := \dfrac{p(\pi_t^j - p)^{r'} + (\pi_t^j - p)p^{r'}}{(\pi_t^j)^{r'}}$ for every $p \in [0, \pi_t^j], j \in I_t^{\text{red}}, t = 2, \ldots, T.$

Proof From (13) we obtain

$$D_{\text{f},\infty}^*(\xi, \xi^{\text{red}}) = \sum_{t=2}^{T} \sup_{\|x_t\|_\infty \leq 1} \left(\sum_{j \in I_t^{\text{red}}} \sum_{i \in J_{t,j}} p_t^i \left| x_t^i - \sum_{k \in J_{t,j}} \frac{p_t^k}{\pi_t^j} x_t^k \right|^{r'} \right)^{\frac{1}{r'}}.$$

Let $x_{t,s}^i$, $s = 1, \ldots, m_t$, denote the components of $x_t^i \in \mathbb{R}^{m_t}$ for every $i \in I_t$ and $t = 2, \ldots, T$. We may continue

$$
D_{f,\infty}^*(\xi, \xi^{\text{red}}) = \sum_{t=2}^{T} \sup_{\|x_t\|_\infty \le 1} \left(\sum_{j \in I_t^{\text{red}}} \sum_{i \in J_{t,j}} p_t^i \sum_{s=1}^{m_t} \left| x_{t,s}^i - \sum_{k \in J_{t,j}} \frac{p_t^k}{\pi_t^j} x_{t,s}^k \right|^{r'} \right)^{\frac{1}{r'}}
$$

$$
= \sum_{t=2}^{T} \sup_{\|x_t\|_\infty \le 1} \left(\sum_{s=1}^{m_t} \sum_{j \in I_t^{\text{red}}} \pi_t^j \sum_{i \in J_{t,j}} \frac{p_t^i}{\pi_t^j} \left| x_{t,s}^i - \sum_{k \in J_{t,j}} \frac{p_t^k}{\pi_t^j} x_{t,s}^k \right|^{r'} \right)^{\frac{1}{r'}}.
$$

Hence, to estimate the filtration distance we have to solve maximum problems of the form

$$
\max \left\{ \sum_{i \in J} \lambda_i \left| y_i - \sum_{k \in J} \lambda_k y_k \right|^{r'} : y \in \mathbb{R}^{|J|}, \max_{i \in J} |y_i| \le 1 \right\}, \tag{16}
$$

where J is a given finite index set with cardinality $|J|$ and $\lambda_i > 0$, $i \in J$, are given with $\sum_{i \in J} \lambda_i = 1$. Let $y^{(j)}$, $j = 1, \ldots, 2^{|J|}$, denote the vertices of the polytope $Y := \{ y \in \mathbb{R}^{|J|} : \max_{i \in J} |y_i| \le 1 \}$. Any element $y \in Y$ can be represented as convex combination of the vertices, i.e.,

$$
y = \sum_{j=1}^{2^{|J|}} \alpha_j y^{(j)}, \quad \text{where } \alpha_j \ge 0 \text{ and } \sum_{j=1}^{2^{|J|}} \alpha_j = 1.
$$

Since the objective function $g(y) := \sum_{i \in J} \lambda_i \left| y_i - \sum_{k \in J} \lambda_k y_k \right|^{r'}$ in (16) is convex, one obtains

$$
g(y) = g\left(\sum_{j=1}^{2^{|J|}} \alpha_j y^{(j)} \right) \le \sum_{j=1}^{2^{|J|}} \alpha_j g(y^{(j)}) \le \max_{j=1,\ldots,2^{|J|}} g(y^{(j)}).
$$

Hence, the maximum in (16) is attained at some $y^* \in Y$ with $y_i^* \in \{+1, -1\}$ for all $i \in J$. Let $J^+ \subseteq J$ and $J^- \subseteq J$ denote the index sets, where y^* is positive and negative, respectively. Furthermore, let

$$
\lambda^+ = \sum_{i \in J^+} \lambda_i \quad \text{and} \quad \lambda^- = \sum_{i \in J^-} \lambda_i.
$$

Then we have $\lambda^+ + \lambda^- = 1$, and, we obtain

$$
\sum_{i \in J} \lambda_i \left| y_i^* - \sum_{k \in J} \lambda_k y_k^* \right|^{r'} = \sum_{i \in J^+} \lambda_i (1 - \lambda^+ + \lambda^-)^{r'} + \sum_{i \in J^-} \lambda_i (1 + \lambda^+ - \lambda^-)^{r'}
$$
$$
= 2^{r'} \lambda^+ (1 - \lambda^+)^{r'} + 2^{r'} (1 - \lambda^+)(\lambda^+)^{r'}.
$$

If we solve the problem (16) for all $s \in \{1, \dots, m_t\}$, $j \in I_t^{\mathrm{red}}$ with $J := J_{t,j}$ and $\lambda_i := \frac{p_t^i}{\pi_t^j}$ we get as final estimate for the filtration distance

$$
D_{\mathrm{f},\infty}^*(\xi, \xi^{\mathrm{red}}) \leq \max_{t=2,\dots,T} 2m_t^{\frac{1}{r'}} \sum_{t=2}^{T} \left(\sum_{j \in I_t^{\mathrm{red}}} M_{t,r'}^j \right)^{\frac{1}{r'}}, \qquad \text{where}
$$

$$
M_{t,r'}^j := \max \left\{ \frac{p(\pi_t^j - p)^{r'} + (\pi_t^j - p)p^{r'}}{(\pi_t^j)^{r'}} : J \subset J_{t,j}, \ p = \sum_{i \in J} p_t^i \right\}.
$$

$\square$

Proposition 1 says that the filtration distance between the given process ξ and the reduced one ξ^{red} only depends on the particular partition structure of the scenarios of ξ and on the choice of representative scenarios of ξ^{red}. The typical cases $r' = 1$ and $r' = 2$ are now discussed in more detail.

The filtration distance for $r' = 1$ and $r' = 2$

In case $r' = 1$ the estimate of the filtration distance is of the form

$$
D_{\mathrm{f},\infty}^*(\xi, \xi^{\mathrm{red}}) \leq \max_{t=2,\dots,T} 2m_t \sum_{t=2}^{T} \sum_{j \in I_t^{\mathrm{red}}} M_{t,1}^j, \qquad \text{where}
$$

$$
M_{t,1}^j := \max \left\{ \frac{2p(\pi_t^j - p)}{\pi_t^j} \ \middle| \ J \subset J_{t,j}, \ p = \sum_{i \in J} p_t^i \right\} \leq \frac{\pi_t^j}{2} \qquad (17)
$$

This allows the following interpretation. For any scenario cluster $J_{t,j}$ defined by the realization $j \in I_t^{\mathrm{red}}$, at some stage t, the contribution $M_{t,1}^j$ to the total filtration distance depends on the total probability π_t^j of the set $J_{t,j}$ as well as on the partitioning of the probability weights. Note that $M_{t,1}^j$ is always bounded from above by π_t^j.

Example: Let $J_{t,j} = \{k_1, \dots, k_n\}$ and $p_t^{k_1} + \cdots + p_t^{k_n} = \pi_t^j$.

(a) In case $p_t^{k_i} = p_t^{k_j}$ for all $i, j = 1, \dots, n$ we obtain $M_{t,1}^j = \frac{n^2-1}{n^2} \frac{\pi_t^j}{2}$ and $M_{t,1}^j = \frac{\pi_t^j}{2}$ if n is odd or even, respectively.

(b) In case of one dominant probability in the sense that $p_t^{k_{j_0}} = \lambda \frac{\pi_t^j}{2}$ with $\lambda \in [1, 2]$ we obtain $M_{t,1}^j = \lambda(2 - \lambda)\frac{\pi_t^j}{2}$. For example, if $\lambda = 1.95$ we have $M_{t,1}^j = 0.0975 \frac{\pi_t^j}{2}$.

Fig. 3 Objective function $f_{t,r'}^j(p)$ to determine $M_{t,r'}^j$ in cases $r' = 1$ and $r' = 2$
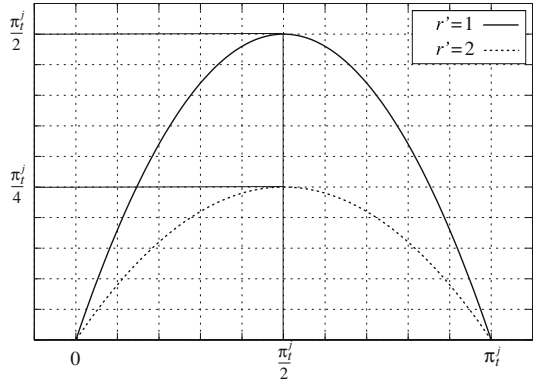
Likewise, in case $r' = 2$ we obtain from Proposition 1

$$D_{f,\infty}^*(\xi, \xi^{\text{red}}) \leq \max_{t=2,\ldots,T} m_t^{\frac{1}{2}} \sum_{t=2}^T \left(\sum_{j \in I_t^{\text{red}}} M_{t,2}^j \right)^{\frac{1}{2}}, \qquad \text{where}$$

$$M_{t,2}^j := \max \left\{ \left. \frac{p(\pi_t^j - p)}{\pi_t^j} \right| J \subset J_{t,j}, \ p = \sum_{i \in J} p_t^i \right\} \leq \frac{\pi_t^j}{4}. \quad (18)$$

Figure 3 shows a plot of the objective in problem (17) and (18), respectively, for determining $M_{t,1}^j$ and $M_{t,2}^j$.

4 Scenario tree reduction in multistage models

In this section we are going to describe a simple scenario reduction algorithm which is based on recursive *single node reduction*. For a given tree structure of ξ, the criterion (9), the representation (14), and the estimate (15) suggest the following *reduction strategy* of ξ given some tolerance $\varepsilon > 0$.

Algorithm (*Single node reduction*)

[Initialization]
   The reduction procedure is initialized by starting from the initial process, i.e., by setting

$$\begin{aligned}
I_t^{\text{red}} &:= I_t \quad \text{for all} \quad t = 1, \ldots, T, \\
I_{t,i}^{\text{red}} &:= I_{t,i} \quad \text{for all} \quad t = 1, \ldots, T-1, \ i \in I_t, \\
J_{t,j} &:= \{j\} \quad \text{for all} \quad t = 1, \ldots, T, \ j \in I_t, \\
q_t^j &:= p_t^j \quad \text{for all} \quad t = 1, \ldots, T, \ j \in I_t, \\
\varepsilon_{\text{appr}} &:= 0,
\end{aligned}$$

where $\varepsilon_{\text{appr}}$ denotes the approximation error of the reduction process.

[Node selection]

The node selection aims at determining an acceptable pair of nodes. A pair (t, j) and (τ, i) is called *acceptable* whenever $t = \tau$, $i \neq j$, the unique predecessors of both nodes coincide, and, simultaneously, the approximation error is small enough. To this end, we search for some $(t, j) \in \{2, \ldots, T\} \times I_t^{\mathrm{red}}$ such that there exists $i \in I_t^{\mathrm{red}}$ with $i \neq j$, $i, j \in I_{t-1,k}^{\mathrm{red}}$ for some $k \in I_{t-1}^{\mathrm{red}}$ and such that the pair (t, j) and (t, i) of nodes satisfies the estimate

$$\varepsilon_{\mathrm{step}} \leq \varepsilon - \varepsilon_{\mathrm{appr}}, \quad \text{where}$$

$$\varepsilon_{\mathrm{step}} := w_1 \left(q_t^i\right)^{\frac{1}{r}} |\xi_t^i - \xi_t^j| + w_2 \frac{\left(q_t^i (q_t^j)^{r'} + (q_t^i)^{r'} q_t^j\right)^{\frac{1}{r'}}}{q_t^i + q_t^j}. \tag{19}$$

If such a pair cannot be found, go to the termination step, otherwise continue with the following reduction step.

[Reduction]

In this step we perform the node reduction according to the selection of acceptable nodes before. We adjust the relevant index sets and probabilities of the scenario tree by enlarging the set $J_{t,j}$, reducing the set I_t, updating the successor information and changing the probabilities. The new index sets and probabilities are given by

$$J_{t,j} := J_{t,j} \cup J_{t,i},$$
$$I_t^{\mathrm{red}} := I_t \setminus \{i\},$$
$$I_{t,j}^{\mathrm{red}} := I_{t,j} \cup I_{t,i},$$
$$q_t^j := q_t^j + q_t^i.$$

Finally, the approximation error is updated by

$$\varepsilon_{\mathrm{appr}} := \varepsilon_{\mathrm{appr}} + \varepsilon_{\mathrm{step}},$$

and the iteration is continued by a new node selection step.

[Termination]

Whenever the termination step is reached all relevant index sets giving the structure of the reduced scenario tree process are stored as I_t^{red}, $I_{t,j}^{\mathrm{red}}$ and $J_{t,j}$ (cf. Sect. 3). It remains to define the (node) probabilities by

$$\pi_t^j := q_t^j \quad \text{for all} \quad t = 1, \ldots, T, \ j \in I_t^{\mathrm{red}}.$$

The process ξ^{red} is well-defined now by the given index sets and probabilities.

We conclude this section with some comments on the above algorithm. In fact, we obtain for the probabilities of the scenario tree process ξ^{red} that

$$\pi_t^j = \sum_{i \in J_{t,j}} p_t^i \quad \text{for all} \quad t = 1, \ldots, T, \ j \in I_t^{\mathrm{red}}.$$

Moreover, the approximation error between the initial scenario tree process and the reduced one can be bounded by $\varepsilon_{\text{appr}}$. More precisely, it holds

$$w_1 \|\xi - \xi^{\text{red}}\|_r \;+\; w_2 D^*_{\text{f},\infty}(\xi, \xi^{\text{red}}) \leq \varepsilon_{\text{appr}},$$

which is a direct consequence of (19) and the triangle inequality for both the L_r-distance and the filtration distance $D^*_{\text{f},\infty}$. Note that the reduction algorithm also can be easily performed with respect to only one distance by setting the weighting factors $w_1 = 0$ and $w_2 = 0$, respectively. If we define $w_1 = 0$ in condition (19) the term controlling the L_r-distance disappears. On the other hand, when using $w_2 = 0$ the filtration term disappears and, hence, the reduction is only performed with respect to the L_r-distance.

Table 1 Structure of the scenario tree processes ξ^{red} obtained by the single node reduction algorithm starting from the input tree containing 221 nodes and terminating with trees containing 150 and 100 nodes, respectively

Nodes	Reduction	Nodes (per Stage)						Scenarios
	w.r.t.	1	2	3	4	5	6	
150	L_r only	1	8	15	23	40	63	63
150	L_r and $D^*_{\text{f},\infty}$	1	8	16	23	41	61	61
150	$D^*_{\text{f},\infty}$ only	1	8	16	24	41	60	60
100	L_r only	1	6	12	17	25	39	39
100	L_r and $D^*_{\text{f},\infty}$	1	8	15	19	25	32	32
100	$D^*_{\text{f},\infty}$ only	1	8	16	20	26	29	29

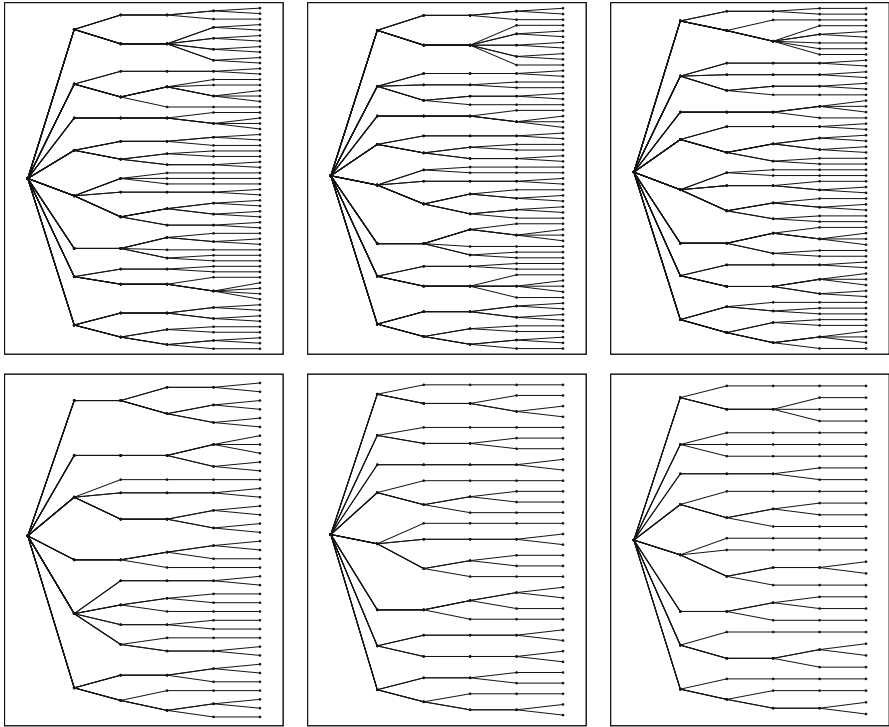**Fig. 4** Structure of the trivariate initial scenario tree ξ serving as input for the single node reduction algorithm
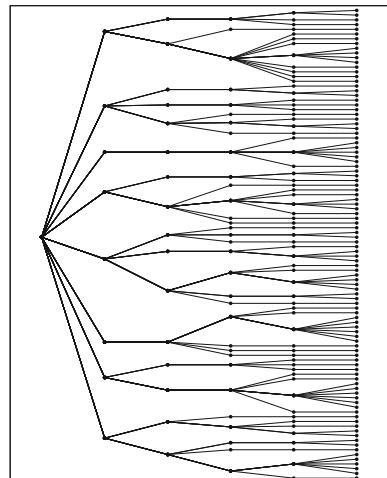
Fig. 5 Illustration of the reduced (trivariate) scenario trees ξ^{red}. The trees are obtained by the single node reduction algorithm until 150 (*above*) and 100 (*below*) remaining nodes are reached. The reduction is carried out with respect to the L_r-distance only (*left*), both the L_r- and the $D_{f,\infty}^*$-distance (*middle*), and the $D_{f,\infty}^*$-distance only (*right*)

5 Numerical experience

Finally, we report on some preliminary numerical experience for scenario tree reduction in multistage stochastic programs. For testing the single node reduction algorithm of the previous section, we consider a stochastic optimization model for electricity portfolios of a German municipal power company. The portfolio consists of the own (thermal) electricity production, the spot market contracts, supply contracts and electricity futures. For details of the optimization model we refer to Eichhorn and Römisch (2005). It takes into account the stochastic nature of the input parameters for every hour of the underlying time horizon, namely, the electricity demand, the heat demand, the EEX spot prices, and base and peak future prices (for each month). Here, we focus on the input scenario tree process and assume that it is obtained by the scenario tree generation method of Heitsch and Römisch (2008). Since the future prices are considered as fair prices and can be derived from the spot prices, the input scenarios correspond to a trivariate time discrete stochastic input process whose components are electricity demand, heat demand, and (EEX) spot prices.

For our purposes a generated scenario tree process ξ is singled out and reduced by the algorithm in Sect. 4 until a prescribed number of nodes is reached. To study, in particular, the impact of the filtration distance, scenario trees ξ^{red} are computed by the single node reduction algorithm, where the reduction is done with respect to the L_r-distance and the $D_{\text{f},\infty}^*$-distance separately as well as with respect to the sum of both distances as advised by the stability analysis of Sect. 2.

Due to modeling reasons the input scenario tree exhibits a monthly branching structure. For our numerical test we considered a time horizon of 6 months which correspond, hence, to six stages of the stochastic program (Table 1). In order to cope with this monthly structure, each component of the scenarios (corresponding to electricity demand, heat demand or spot prices) is represented by 6 vectors, where each vector contains the inputs of one month in hourly discretization. The tree structure of the input process is illustrated in Fig. 4.

Figure 5 illustrates the results of the scenario tree reduction by applying the single node reduction algorithm until 150 and 100 nodes remain, respectively. They show that the filtration distance influences the structure of the reduced scenario trees noticeably. The incorporation of the filtration distance leads to a smaller number of remaining scenarios in both cases. The opposite effect appears when using the L_r-distance only.

6 Conclusions

Summarizing our theoretical arguments and preliminary numerical experience indicates that the incorporation of the filtration distance into the reduction of scenario trees is indispensable. This implies, in particular, that deleting scenarios in input trees for multi-stage models according to the methodology presented in Dupačová et al. (2003) and Heitsch and Römisch (2003, 2007) is *not* appropriate as the information (filtration) structure is not taken into account. The numerical results in Sect. 4 are obtained by a simple straightforward strategy of reducing single nodes recursively. But, the estimates (14) and (15) for the L_r- and filtration distance offer further potential for algorithmic extensions.

References

Casey M, Sen S (2005) The scenario generation algorithm for multistage stochastic linear programming. Math Oper Res 30:615–631

Dupačová J, Consigli G, Wallace SW (2000) Scenarios for multistage stochastic programs. Ann Oper Res 100:25–53

Dupačová J, Gröwe-Kuska N, Römisch W (2003) Scenario reduction in stochastic programming: an approach using probability metrics. Math Program 95:493–511

Eichhorn A, Römisch W (2005) Polyhedral risk measures in stochastic programming. SIAM J Optim 16:69–95

Eichhorn A, Römisch W (2006) Mean-risk optimization models for electricity portfolio management. In: Proceedings of PMAPS 2006 (Probabilistic Methods Applied to Power Systems), Stockholm, Sweden

Eichhorn A, Römisch W (2008) Stability of multistage stochastic programs incorporating polyhedral risk measures. Optimization 57:295–318

Heitsch H, Römisch W (2003) Scenario reduction algorithms in stochastic programming. Comp Optim Appl 24:187–206

Heitsch H, Römisch W (2006) Stability and scenario trees for multistage stochastic programs. In: Infanger G (ed) Stochastic Programming—The State of the Art (submitted)

Heitsch H, Römisch W (2007) A note on scenario reduction for two-stage stochastic programs. Oper Res Let 35:731–736

Heitsch H, Römisch W (2008) Scenario tree modeling for multistage stochastic programs. Math Program (to appear)

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

Henrion R, Küchler C, Römisch W (2007) Scenario reduction in stochastic programming with respect to discrepancy distances. Comp Optim Appl (to appear)

Henrion R, Küchler C, Römisch W (2008) Discrepancy distances and scenario reduction in two-stage stochastic integer programming. J Ind Manage Optim 4:363–384

Hochreiter R, Pflug GCh (2007) Financial scenario generation for stochastic multi-stage decision processes as facility location problems. Ann Oper Res 152:257–272

Høyland K, Kaut M, Wallace SW (2003) A heuristic for moment-matching scenario generation. Comp Optim Appl 24:169–185

Kuhn D (2005) Generalized bounds for convex multistage stochastic programs. Lecture Notes in Economics and Mathematical Systems, vol 548. Springer, Berlin

Kuhn D (2008) Aggregation and discretization in multistage stochastic programming. Math Program 113:61–94

Pennanen T (2009) Epi-convergent discretizations of multistage stochastic programs via integration quadratures. Math Program 116:461–479

Rachev ST, Rüschendorf L (1998) Mass transportation problems, vol I, II. Springer, Berlin

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

Römisch W, Wets RJ-B (2007) Stability of ε-approximate solutions to convex stochastic programs. SIAM J Optim 18:961–979