

# Stability and scenario trees for multistage stochastic programs

Holger Heitsch and Werner Römisch

Humboldt-University Berlin

Institute of Mathematics

10099 Berlin, Germany

## Abstract

By extending the stability analysis of [20] for multistage stochastic programs we show that their (approximate) solution sets behave stable with respect to the sum of an  $L_r$ -distance and a filtration distance. Based on such stability results we suggest a scenario tree generation method for the (multivariate) stochastic input process. It starts with an initial scenario set and consists of a recursive deletion and branching procedure which is controlled by bounding the approximation error. Some numerical experience for generating scenario trees in electricity portfolio management is reported.

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

**2000 MSC:** 90C15, 90C31

## 1 Introduction

Multistage stochastic programs are often used to model practical decision processes over time and under uncertainty, e.g., in finance, production, energy and logistics. We refer to the pioneering work of Dantzig [5, 6], and to the recent books [39], [44] and the monograph [27] for the state-of-the-art of the theory and solution methods for multistage models and for a variety of applications.

The inputs of multistage stochastic programs are multivariate stochastic processes  $\{\xi_t\}_{t=1}^T$  defined on some probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  and with  $\xi_t$  taking values in some  $\mathbb{R}^d$ . The decision  $x_t$  at  $t$  belonging to  $\mathbb{R}^{m_t}$  is assumed to be *nonanticipative*, i.e., to depend only on  $(\xi_1, \dots, \xi_t)$ . This property is equivalent to the measurability of  $x_t$  with respect to the  $\sigma$ -field  $\mathcal{F}_t(\xi) \subseteq \mathcal{F}$  which is generated by  $(\xi_1, \dots, \xi_t)$ . 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  $\mathcal{F}_1 = \{\emptyset, \Omega\}$ .

The multistage stochastic program is assumed to be of the form

$$\min \left\{ \mathbb{E} \left[ \sum_{t=1}^T \langle b_t(\xi_t), x_t \rangle \right] \left| \begin{array}{l} x_t \in X_t, t = 1, \dots, T, A_{1,0}x_1 = h_1(\xi_1), \\ x_t \text{ is } \mathcal{F}_t(\xi)\text{-measurable, } t = 1, \dots, T, \\ A_{t,0}x_t + A_{t,1}(\xi_t)x_{t-1} = h_t(\xi_t), t = 2, \dots, T \end{array} \right. \right\}, \quad (1)$$

where the sets  $X_t \subseteq \mathbb{R}^{m_t}$  are polyhedral cones, the cost coefficients  $b_t(\xi_t)$  and right-hand sides  $h_t(\xi_t)$  belong to  $\mathbb{R}^{m_t}$  and  $\mathbb{R}^{n_t}$ , respectively, the fixed recourse matrices  $A_{t,0}$  and the technology matrices  $A_{t,1}(\xi_t)$  are  $(n_t, m_t)$ - and  $(n_t, m_{t-1})$ -matrices, respectively. The costs  $b_t(\cdot)$ , technology matrices  $A_{t,1}(\cdot)$  and right-hand sides  $h_t(\cdot)$  are assumed to depend affinely linear on  $\xi_t$ .

While the first and third groups of constraints in (1) have to be satisfied pointwise with probability 1, the second group, the measurability or *information* constraints, are functional and non-pointwise at least if  $T > 2$  and  $\mathcal{F}_2 \subsetneq \mathcal{F}_t \subseteq \mathcal{F}$  for some  $2 < t \leq T$ . The presence of such qualitatively different constraints constitutes the origin of both the theoretical and computational challenges of multistage models. Recent results (see [42, 43]) indicate that multistage stochastic programs have higher computational complexity than two-stage 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  $\xi_1$  of  $\mathbb{R}^d$ . This leads to linear programming models that are very large scale in most cases and can be solved by linear programming techniques, in particular, by decomposition methods that exploit specific structures of the model. We refer to [39, Chapter 3] for a recent survey.

Presently, there exist several approaches to generate scenario trees for multistage stochastic programs (see [8] for a survey). They are based on several different principles. We mention here (i) bound-based constructions [2, 10, 15, 29], (ii) Monte Carlo-based schemes [3, 41, 43] or Quasi Monte Carlo-based methods [32, 33], (iii) (EVPI-based) Sampling within decomposition schemes [4, 7, 21, 26], (iv) the target/moment-matching principle [24, 25], and (v) probability metric based approximations [17, 19, 22, 23, 34].

We add a few more detailed comments on some of the recent work. The approach of (i) relies on constructing discrete probability measures that correspond to lower and upper bounds (under certain assumptions on the model and the stochastic input) and on refinement strategies. The recent paper/monograph [2, 29] belonging to (i) also offer convergence arguments (restricted to linear models containing only stochasticity in right-hand sides in [2] and to convex models whose stochasticity is assumed to follow some linear block-diagonal autoregressive process with compact supports in [29]). The Monte Carlo-based methods in (ii) utilize conditional sampling schemes and lead to a large number of (pseudo) random number generator calls for conditional distributions. Consistency results are shown in [41] and the complexity is discussed in [42]. The Quasi Monte Carlo-based methods in [32, 33] are developed for convex models and for stochastic processes driven by time series models with uniform innovations. While the general theory on epi-convergent discretizations in [32] also applies to conditional sampling procedures, a general procedure for generating scenario trees of such time series driven stochastic processes is developed in [33] by approximating each of the (independent) uniform random variables using Quasi Monte Carlo methods (see [31]). The motivation of using Quasi Monte Carlo schemes originates from their remarkable convergence properties and good performance for the computation of high-dimensional integrals while "generating random samples is difficult" [31, p. 7]. The approach of (v) is based on probability distances that are relevant for the stability of multistage models. While the papers [17, 23, 34] employ Fortet-Mourier or Wasserstein distances, our

recent work [19] is based on the rigorous stability result for linear multistage stochastic programs in [20]. Most of the methods for generating scenario trees require to prescribe (at least partially) the tree structure. Finally, we also mention the importance of evaluating the quality of scenario trees and of a postoptimality analysis [8, 28].

In the present paper we extend the theoretical results obtained in [20] by proving an existence result for solutions of (1) (Theorem 2.1), a Lipschitz stability result for  $\varepsilon$ -approximate solution sets and a (qualitative) stability result for solutions of multistage models. In addition, we review the forward technique of [19] for generating scenario trees. Its idea is to start with an initial finite scenario set with given probabilities which represents a ‘good’ approximation of the underlying stochastic input process  $\xi$ . Such a finite set of scenarios may be obtained by sampling or resampling techniques based on parametric or nonparametric stochastic models of  $\xi$  or by optimal quantization techniques [16]. Starting from the initial scenario set, a tree is constructed recursively by scenario reduction [9, 18] and bundling (Algorithm 3.2). We review an error estimate for Algorithm 3.2 in terms of the  $L_r$ -distance (Theorem 3.1) and a convergence result (Theorem 3.4). Algorithm 3.2 represents a stability-based heuristic for generating scenario trees. It has been implemented and tested on real-life data in several practical applications. Numerical experience was reported in [19] on generating inflow-demand scenario trees based on real-life data provided by the French electricity company EdF. Algorithm 3.2 or a modified version were used in [40] to generate scenario trees in power engineering models and in [30] on generating passenger demand scenario trees in airline revenue management.

Section 2 presents extensions of the stability result of [20] which provide the basis of our tree constructions. Section 3 reviews some results of [19], in particular, the forward tree construction and error estimates in terms of the  $L_r$ -distances and a convergence result. In Section 4 we discuss some numerical experience on generating load-price scenario trees for an electricity portfolio optimization model based on real-life data of a municipal German power company.

## 2 Stability of multistage models

We assume that the stochastic input process  $\xi = \{\xi_t\}_{t=1}^T$  belongs to the linear space  $\times_{t=1}^T L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d)$  for some  $r \in [1, +\infty]$ . The model (1) is regarded as optimization problem in the space  $\times_{t=1}^T L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{m_t})$  for some  $r' \in [1, \infty]$ , where both linear spaces are Banach spaces when endowed with the norms

$$\|\xi\|_r := \left( \sum_{t=1}^T \mathbb{E}[|\xi_t|^r] \right)^{\frac{1}{r}} \quad \text{for } r \in [1, \infty) \quad \text{and} \quad \|\xi\|_\infty := \max_{t=1, \dots, T} \text{ess sup } |\xi_t|,$$

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

respectively. Here,  $|\cdot|$  denotes some norm on the relevant Euclidean spaces and  $r'$  is defined by

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

The definition of  $r'$  is justified by the proof of [20, Theorem 2.1], which we record as Theorem 2.2. Since  $r'$  depends on  $r$  and our assumptions will depend on both  $r$  and  $r'$ , we will add some comments on the choice of  $r$  and its interplay with the structure of the underlying stochastic programming model. To have the stochastic program well defined, the existence of certain moments of  $\xi$  has to be required. This fact is well known for the two-stage situation (see, e.g., Chapter 2 in [39]). If either right-hand sides or costs in a multistage model (1) are random, it is sufficient to require  $r \geq 1$ . The flexibility in case that the stochastic process  $\xi$  has moments of order  $r > 1$  may be used to choose  $r'$  as small as possible in order to weaken the condition (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), one needs  $r \geq T$  to have the model well defined and no flexibility on  $r'$  remains.

Let us introduce some notation. Let  $F$  denote the objective function defined on  $L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s) \times L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m) \rightarrow \overline{\mathbb{R}}$  by

$$F(\xi, x) := \begin{cases} \mathbb{E} \left[ \sum_{t=1}^T \langle b_t(\xi_t), x_t \rangle \right] & , x \in \mathcal{X}(\xi), \\ +\infty & , \text{ otherwise,} \end{cases}$$

where

$$\mathcal{X}(\xi) := \{x \in L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m) : x_1 \in \mathcal{X}_1(\xi_1), x_t \in \mathcal{X}_t(x_{t-1}; \xi_t), t = 2, \dots, T\}$$

is the set of feasible elements of (1) and

$$\begin{aligned} \mathcal{X}_1(\xi_1) &:= \{x_1 \in X_1 : A_{1,0}x_1 = h_1(\xi_1)\} \\ \mathcal{X}_t(x_{t-1}; \xi_t) &:= \{x_t \in \mathbb{R}^{m_t} : x_t \in X_t, A_{t,0}x_t + A_{t,1}(\xi_t)x_{t-1} = h_t(\xi_t)\} \end{aligned}$$

the  $t$ -th feasibility set for every  $t = 2, \dots, T$ . Denoting by

$$\mathcal{N}_{r'}(\xi) := \times_{t=1}^T L_{r'}(\Omega, \mathcal{F}_t(\xi), \mathbb{P}; \mathbb{R}^{m_t})$$

the nonanticipativity subspace of  $\xi$  allows to rewrite the stochastic program (1) in the form

$$\min\{F(\xi, x) : x \in \mathcal{N}_{r'}(\xi)\}. \quad (3)$$

Let  $v(\xi)$  denote the optimal value of (3) and, for any  $\alpha \geq 0$ , let

$$S_\alpha(\xi) := \{x \in \mathcal{N}_{r'}(\xi) : F(\xi, x) \leq v(\xi) + \alpha\}$$

denote the  $\alpha$ -approximate solution set of the stochastic program (3). Since, for  $\alpha = 0$ , the set  $S_\alpha(\xi)$  coincides with the set solutions to (3), we will also use the notation

$$S(\xi) := S_0(\xi).$$

The following conditions are imposed on (3):

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

(A2) There exists a  $\delta > 0$  such that for any  $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$  with  $\|\tilde{\xi} - \xi\|_r \leq \delta$ , any  $t = 2, \dots, T$  and any  $x_1 \in \mathcal{X}_1(\tilde{\xi}_1)$ ,  $x_\tau \in \mathcal{X}_\tau(x_{\tau-1}; \tilde{\xi}_\tau)$ ,  $\tau = 2, \dots, t-1$ , 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  $\xi$* ).

(A3) The optimal values  $v(\tilde{\xi})$  of (3) with input  $\tilde{\xi}$  are finite for all  $\tilde{\xi}$  in a neighborhood of  $\xi$  and the objective function  $F$  is *level-bounded locally uniformly at  $\xi$* , i.e., for some  $\alpha > 0$  there exists a  $\delta > 0$  and a bounded subset  $B$  of  $L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$  such that  $S_\alpha(\tilde{\xi})$  is contained in  $B$  for all  $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$  with  $\|\tilde{\xi} - \xi\|_r \leq \delta$ .

For any  $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$  with  $\|\tilde{\xi} - \xi\|_r \leq \delta$ , condition (A2) implies the existence of some feasible  $\tilde{x}$  in  $\mathcal{X}(\tilde{\xi})$  and (2) implies the finiteness of the objective  $F(\tilde{\xi}, \cdot)$  at any feasible  $\tilde{x}$ . A sufficient condition for (A2) to hold is the *complete recourse condition* on every recourse matrix  $A_{t,0}$ , i.e.,  $A_{t,0}X_t = \mathbb{R}^{n_t}$ ,  $t = 1, \dots, T$ . The locally uniform level-boundedness of the objective function  $F$  is quite standard in perturbation results for optimization problems (see, e.g., [37, Theorem 1.17]). The finiteness condition on the optimal value  $v(\xi)$  is not implied by the level-boundedness of  $F$  for all relevant pairs  $(r, r')$ . In general, the conditions (A2) and (A3) get weaker for increasing  $r$  and decreasing  $r'$ , respectively.

To state our first result on the existence of solutions to (3) in full generality, we need two additional conditions:

(A4) There exists a feasible element  $z$  in  $\times_{t=1}^T L_{\hat{r}}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{n_t})$ ,  $\frac{1}{r} + \frac{1}{\hat{r}} = 1$ , of the dual stochastic program to (3), i.e., it holds that

$$A_{t,0}^\top z_t + A_{t+1,1}^\top(\xi_{t+1})z_{t+1} - b_t(\xi_t) \in X_t^*, t = 1, \dots, T-1, A_{T,0}^\top z_T - b_T(\xi_T) \in X_T^*, \quad (4)$$

where  $X_t^*$  denotes the polar to the polyhedral cone  $X_t$ ,  $t = 1, \dots, T$ .

(A5) If  $r' = 1$  we require that, for each  $c \geq 0$ , there exists  $g \in L_1(\Omega, \mathcal{F}, \mathbb{P})$  such that

$$\sum_{t=1}^T \langle b_t(\xi_t(\omega)), x_t \rangle \geq c|x| - g(\omega)$$

for all  $x \in \mathbb{R}^m$  such that  $x_t \in X_t$ ,  $t = 1, \dots, T$ ,  $A_{1,0}x_1 = h_1(\xi_1)$ ,  $A_{t,0}x_t + A_{t,1}(\xi_t(\omega))x_{t-1} = h_t(\xi_t(\omega))$ ,  $t = 2, \dots, T$ , and for  $\mathbb{P}$ -almost all  $\omega \in \Omega$ .

To use Weierstrass' result on the existence of minimizers, we need a topology  $\mathcal{T}$  on  $L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$  such that some approximate solution set  $S_\alpha(\xi)$  is compact with respect to  $\mathcal{T}$ . Since, in general, the norm topology is too strong for infinite-dimensional optimization models in  $L_p$ -spaces, we resort to the weak topologies  $\sigma(L_p, L_q)$  on the spaces  $L_p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ , where  $p \in [1, \infty]$  and  $\frac{1}{p} + \frac{1}{q} = 1$ . They are Hausdorff topological spaces and generated by a basis consisting of the sets

$$\mathcal{O} = \left\{ x \in L_p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m) : \left| \mathbb{E} \left[ \sum_{t=1}^T \langle x_t - x_t^0, y_t^i \rangle \right] \right| < \varepsilon, i = 1, \dots, n \right\}$$

for all  $x^0 \in L_p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ ,  $n \in \mathbb{N}$ ,  $\varepsilon > 0$  and  $y^i \in L_q(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ ,  $i = 1, \dots, n$ . For  $p \in [1, \infty)$ , the weak topology  $\sigma(L_p, L_q)$  is of the form  $\sigma(E, E^*)$  with some Banach space  $E$  and its topological dual  $E^*$ . For  $p = \infty$ , the weak topology  $\sigma(L_\infty, L_1)$  on the

Banach space  $L_\infty(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$  is sometimes called weak\* topology since it is of the form  $\sigma(E^*, E)$ . If  $\Omega$  is finite, the weak topologies coincide with the norm topology. If the space  $L_p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$  is infinite-dimensional, its weak topology  $\sigma(L_p, L_q)$  is even not metrizable. For  $p \in [1, \infty)$ , subsets of  $L_p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$  are (relatively) weakly compact iff they are (relatively) weakly sequentially compact due to the Eberlein-Šmulian theorem. For  $p = \infty$  the latter property is lost in general. However, if a subset  $B$  of  $L_p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$  is compact with respect to the weak topology  $\sigma(L_p, L_q)$ , its restriction to  $B$  is metrizable if  $L_q(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$  is separable. We note that the Banach space  $L_p(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$  with  $p \in [1, \infty)$  is separable if there exists a countable set  $\mathcal{G}$  of subsets of  $\Omega$  such that  $\mathcal{F}$  is the smallest  $\sigma$ -field containing  $\mathcal{G}$  [45]. A  $\sigma$ -field  $\mathcal{F}$  contains such a countable generator if it is generated by a  $\mathbb{R}^m$ -valued random vector. For these and related results we refer to [14, Sections 3 and 4].

Now, we are ready to state our existence result for solutions of (3).

**Theorem 2.1** *Let (A1) – (A5) be satisfied for some pair  $(r, r')$  satisfying (2). Then the solution set  $S(\xi)$  of (3) is nonempty, convex and compact with respect to the weak topology  $\sigma(L_{r'}, L_q)$  ( $\frac{1}{r'} + \frac{1}{q} = 1$ ). Here, the conditions (A4) and (A5) are only needed for  $r' \in \{1, \infty\}$ .*

**Proof:** We define the integrand  $f : \Omega \times \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$

$$f(\omega, x) := \begin{cases} \sum_{t=1}^T \langle b_t(\xi_t(\omega)), x_t \rangle & , x_1 \in \mathcal{X}_1(\xi_1), x_t \in \mathcal{X}_t(x_{t-1}, \xi_t(\omega)), t = 2, \dots, T, \\ +\infty & , \text{otherwise.} \end{cases}$$

Then  $f$  is a proper normal convex integrand (cf. [36] and [37, Chapter 14]).

Let  $(\omega, x) \in \Omega \times \mathbb{R}^m$  be such that  $x_1 \in \mathcal{X}_1(\xi_1)$ ,  $x_t \in \mathcal{X}_t(x_{t-1}, \xi_t(\omega))$ ,  $t = 2, \dots, T$ . Then we conclude from (A4) the existence of  $z \in \times_{t=1}^T L_{\hat{r}}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{n_t})$  such that (4) is satisfied. Hence, for each  $t = 1, \dots, T$ , there exists  $x_t^*(\omega) \in X_t^*$  such that

$$\begin{aligned} b_t(\xi_t(\omega)) &= A_{t,0}^\top z_t(\omega) + A_{t+1,1}^\top(\xi_{t+1}(\omega)) z_{t+1}(\omega) - x_t^*(\omega) \quad (t = 1, \dots, T-1) \\ b_T(\xi_T(\omega)) &= A_{T,0}^\top z_T(\omega) - x_T^*(\omega). \end{aligned}$$

Inserting the latter representation of  $b_t(\xi_t(\omega))$  into the integrand  $f$  (defining  $F(\xi, x) = \mathbb{E}[f(\omega, x)]$ ) leads to

$$\begin{aligned} f(\omega, x) &= \sum_{t=1}^{T-1} \langle A_{t,0}^\top z_t(\omega) + A_{t+1,1}^\top(\xi_{t+1}(\omega)) z_{t+1}(\omega) - x_t^*(\omega), x_t \rangle \\ &\quad + \langle A_{T,0}^\top z_T(\omega) - x_T^*(\omega), x_T \rangle \\ &\geq \sum_{t=1}^{T-1} \langle A_{t,0}^\top z_t(\omega) + A_{t+1,1}^\top(\xi_{t+1}(\omega)) z_{t+1}(\omega), x_t \rangle + \langle A_{T,0}^\top z_T(\omega), x_T \rangle \\ &= \sum_{t=1}^T \langle z_t(\omega), A_{t,0} x_t \rangle + \sum_{t=1}^{T-1} \langle z_{t+1}(\omega), A_{t+1,1}(\xi_{t+1}(\omega)) x_t \rangle \\ &= \sum_{t=1}^T \langle z_t(\omega), h_t(\xi_t(\omega)) \rangle. \end{aligned}$$

Hence, we have

$$f(\omega, x) \geq g(\omega), \quad \text{where} \quad g := \sum_{t=1}^T \langle z_t, h_t(\xi_t) \rangle \in L_1(\Omega, \mathcal{F}, \mathbb{P}).$$

This implies for the conjugate normal convex integrand  $f^* : \Omega \times \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$  given by

$$f^*(\omega, y) := \sup_{x \in \mathbb{R}^m} \{ \langle y, x \rangle - f(\omega, x) \}$$

that the estimate  $f^*(\omega, 0) \leq -g(\omega)$  holds. Hence, the assumption of [36, Corollary 3D] is satisfied and we conclude that the integral functional  $F(\xi, \cdot) = \mathbb{E}[f(\omega, \cdot)]$  is lower semicontinuous on  $L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$  with respect to the weak topology  $\sigma(L_{r'}, L_q)$ .

The nonanticipativity subspace  $\mathcal{N}_{r'}(\xi)$  is closed with respect to the weak topology  $\sigma(L_{r'}, L_q)$  for all  $r' \in [1, \infty]$ . For  $r' \in [1, \infty)$  this fact is a consequence of the norm closedness and convexity of  $\mathcal{N}_{r'}(\xi)$ . For  $r' = \infty$ , let  $(x_\alpha)_{\alpha \in I}$  be a net in  $\mathcal{N}_\infty(\xi)$  with some partially ordered set  $(I, \leq)$  that converges to some  $x^* \in L_\infty(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ . Any neighborhood  $U(x^*)$  of  $x^*$  with respect to the weak topology  $\sigma(L_\infty, L_1)$  is of the form

$$U(x^*) = \left\{ x \in L_\infty(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m) : \left| \mathbb{E} \left[ \sum_{t=1}^T \langle x_t - x_t^*, y_t^i \rangle \right] \right| < \varepsilon_i, i = 1, \dots, n \right\},$$

where  $n \in \mathbb{N}$ ,  $y^i \in L_1(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ ,  $\varepsilon_i > 0$ ,  $i = 1, \dots, n$ . Since the net  $(x_\alpha)_{\alpha \in I}$  converges to  $x^*$ , there exists  $\alpha_0 \in I$  such that  $x_\alpha \in U(x^*)$  whenever  $\alpha_0 \leq \alpha$ . If the elements  $y^i$  belong to  $\times_{t=1}^T L_1(\Omega, \mathcal{F}_t, \mathbb{P}; \mathbb{R}^{m_t})$  for each  $i = 1, \dots, n$ , we obtain

$$\begin{aligned} \left| \mathbb{E} \left[ \sum_{t=1}^T \langle x_{\alpha,t} - x_t^*, y_t^i \rangle \right] \right| &= \left| \mathbb{E} \left[ \sum_{t=1}^T \mathbb{E}[\langle x_{\alpha,t} - x_t^*, y_t^i \rangle | \mathcal{F}_t] \right] \right| \\ &= \left| \mathbb{E} \left[ \sum_{t=1}^T \langle x_{\alpha,t} - \mathbb{E}[x_t^* | \mathcal{F}_t], y_t^i \rangle \right] \right| < \varepsilon_i \end{aligned}$$

due to the fact that  $\mathbb{E}[x_{\alpha,t} | \mathcal{F}_t] = x_{\alpha,t}$  for each  $t = 1, \dots, T$  and  $\alpha \in I$ . Hence, we have in this case

$$U(x^*) = U(\mathbb{E}[x_1^* | \mathcal{F}_1], \dots, \mathbb{E}[x_T^* | \mathcal{F}_T]).$$

Since the net  $(x_\alpha)_{\alpha \in I}$  converges to  $x^*$  and the weak topology is Hausdorff, we conclude  $x_t^* = \mathbb{E}[x_t^* | \mathcal{F}_t]$ ,  $t = 1, \dots, T$ , and, thus,  $x^* \in \mathcal{N}_\infty(\xi)$ .

It remains to show that, for some  $\alpha > 0$ , the  $\alpha$ -approximate solution set  $S_\alpha(\xi)$  is compact with respect to the weak topology  $\sigma(L_{r'}, L_q)$ . For  $r' \in (1, \infty)$  the Banach space  $L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$  is reflexive. Furthermore, any  $\alpha$ -approximate solution set  $S_\alpha(\xi)$  is closed and convex. For some  $\alpha > 0$  the level set is also bounded due to (A3) and, hence, compact with respect to  $\sigma(L_{r'}, L_q)$ . For  $r' = 1$  the compactness of any  $\alpha$ -level set with respect to  $\sigma(L_1, L_\infty)$  follows from [36, Theorem 3K] due to condition (A5). For  $r' = \infty$ , some  $\alpha$ -level set is bounded due to (A3) and, hence, relatively compact with respect to  $\sigma(L_\infty, L_1)$  due to Alaoglu's theorem [14, Theorem 3.21]. Since the objective function  $F(\xi, \cdot)$  is lower semicontinuous and  $\mathcal{N}_\infty(\xi)$  weakly closed with respect to  $\sigma(L_\infty, L_1)$ , the  $\alpha$ -level set is even compact with respect to  $\sigma(L_\infty, L_1)$ .

Altogether,  $S(\xi)$  is nonempty due to Weierstrass' theorem and compact with respect to  $\sigma(L_{r'}, L_q)$ . The convexity of  $S(\xi)$  is an immediate consequence of the convexity of the objective  $F(\xi, \cdot)$  of the stochastic program (3).

Finally, we note that assumptions (A4) and (A5) are not needed for proving that  $S(\xi)$  is nonempty and compact with respect to the topology  $\sigma(L_{r'}, L_q)$  in case  $r' \in (1, \infty)$ . This fact is an immediate consequence of minimizing a linear continuous functional on a closed, convex, bounded subset of a reflexive Banach space.  $\square$

To state our next result we introduce the functional  $D_f(\xi, \tilde{\xi})$  depending on the filtrations of  $\xi$  and of its perturbation  $\tilde{\xi}$ , respectively. It is defined by

$$D_f(\xi, \tilde{\xi}) := \sup_{\varepsilon \in (0, \alpha]} \inf_{\substack{x \in S_\varepsilon(\xi) \\ \tilde{x} \in S_\varepsilon(\tilde{\xi})}} \sum_{t=2}^{T-1} \max\{\|x_t - \mathbb{E}[x_t | \mathcal{F}_t(\tilde{\xi})]\|_{r'}, \|\tilde{x}_t - \mathbb{E}[\tilde{x}_t | \mathcal{F}_t(\xi)]\|_{r'}\}, \quad (5)$$

In the following, we call the functional  $D_f$  *filtration distance*, although it fails to satisfy the triangle inequality in general. If solutions of (3) for the inputs  $\xi$  and  $\tilde{\xi}$  exist (see Theorem 2.1), the filtration distance is of the simplified form

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

We note that the conditional expectations  $\mathbb{E}[x_t | \mathcal{F}_t(\tilde{\xi})]$  and  $\mathbb{E}[\tilde{x}_t | \mathcal{F}_t(\xi)]$  may be written equivalently in the form  $\mathbb{E}[x_t | \tilde{\xi}_1, \dots, \tilde{\xi}_t]$  and  $\mathbb{E}[\tilde{x}_t | \xi_1, \dots, \xi_t]$ , respectively.

The following stability result for optimal values of program (3) is essentially [20, Theorem 2.1].

**Theorem 2.2** *Let (A1), (A2) and (A3) be satisfied and the sets  $\mathcal{X}_1(\tilde{\xi}_1)$  be nonempty and uniformly bounded in  $\mathbb{R}^{m_1}$  if  $|\tilde{\xi}_1 - \xi_1| \leq \delta$  (where  $\delta > 0$  is taken from (A3)).*

*Then there exists positive constants  $L$  and  $\delta$  such that the estimate*

$$|v(\xi) - v(\tilde{\xi})| \leq L(\|\xi - \tilde{\xi}\|_r + D_f(\xi, \tilde{\xi})) \quad (6)$$

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

The proof of [20, Theorem 2.1] extends easily to constraints for  $x_1$  that depend on  $\xi_1$  (via the right-hand side of the equality constraint  $A_{1,0}x_1 = h(\xi_1)$ ). We note that the constant  $L$  depends on  $\|\xi\|_r$  in all cases.

To prove a stability result for (approximate) solutions of (3), we need a stronger version of the filtration distance  $D_f$ , namely,

$$D_f^*(\xi, \tilde{\xi}) = \sup_{\|x\|_{r'} \leq 1} \sum_{t=2}^T \|\mathbb{E}[x_t | \mathcal{F}_t(\xi)] - \mathbb{E}[x_t | \mathcal{F}_t(\tilde{\xi})]\|_{r'}. \quad (7)$$

Notice that the sum is extended by the additional summand for  $t = T$  and that the former infimum is replaced by a supremum with respect to a sufficiently large bounded set (the unit ball in  $L_{r'}$ ). Clearly, the conditions (A1)–(A3) imply the estimate

$$D_f(\xi, \tilde{\xi}) \leq \sup_{x \in B} \sum_{t=2}^{T-1} \|\mathbb{E}[x_t | \mathcal{F}_t(\xi)] - \mathbb{E}[x_t | \mathcal{F}_t(\tilde{\xi})]\|_{r'} \leq C D_f^*(\xi, \tilde{\xi}) \quad (8)$$

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

Sometimes, the unit ball in  $L_{r'}$  in the definition of  $D_f^*$  is too large. It may be replaced by the smaller set  $\mathcal{B}_\infty := \{x : \Omega \rightarrow \mathbb{R}^m : x \text{ is measurable, } |x(\omega)| \leq 1 \text{ for all } \omega \in \Omega\}$  if the following stronger condition (A3)' is satisfied.

**(A3)'** The optimal values  $v(\tilde{\xi})$  of (3) with input  $\tilde{\xi}$  are finite for all  $\tilde{\xi}$  in a neighborhood of  $\xi$  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 define

$$D_f^*(\xi, \tilde{\xi}) := \sup_{x \in \mathcal{B}_\infty} \sum_{t=2}^T \|\mathbb{E}[x_t | \mathcal{F}_t(\xi)] - \mathbb{E}[x_t | \mathcal{F}_t(\tilde{\xi})]\|_{r'} \quad (9)$$

and have  $D_f(\xi, \tilde{\xi}) \leq C D_f^*(\xi, \tilde{\xi})$ . We note that  $D_f^*$  always satisfies the triangle inequality.

In the next result we derive a (local) Lipschitz property of the feasible set-valued mapping  $\mathcal{X}(\cdot)$  from  $L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$  into  $L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$  in terms of a 'truncated' Pompeiu-Hausdorff type distance

$$\hat{d}_\rho(B, \tilde{B}) = \inf \left\{ \eta \geq 0 : B \cap \rho\mathbb{B} \subset \tilde{B} + \eta\mathbb{B}, \tilde{B} \cap \rho\mathbb{B} \subset B + \eta\mathbb{B} \right\}$$

of closed subsets  $B$  and  $\tilde{B}$  of the space  $L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$  with  $\mathbb{B}$  denoting its unit ball. The Pompeiu-Hausdorff distance may be defined by

$$d_\infty(B, \tilde{B}) = \lim_{\rho \rightarrow \infty} \hat{d}_\rho(B, \tilde{B})$$

(see [37, Corollary 4.38]).

**Proposition 2.3** *Let (A1), (A2) and (A3) be satisfied with  $r' \in [1, \infty)$  and the sets  $\mathcal{X}_1(\tilde{\xi}_1)$  be nonempty and uniformly bounded in  $\mathbb{R}^{m_1}$  if  $|\tilde{\xi}_1 - \xi_1| \leq \delta$  (with  $\delta > 0$  from (A3)). Then there exist positive constants  $L$  and  $\delta$  such that the estimate*

$$\hat{d}_\rho(\mathcal{X}(\xi), \mathcal{X}(\tilde{\xi})) \leq L(\|\xi - \tilde{\xi}\|_r + \rho D_f^*(\xi, \tilde{\xi}))$$

holds for any  $\rho > 0$  and any  $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$  with  $\|\xi - \tilde{\xi}\|_r \leq \delta$ . If (A3)' is satisfied instead of (A3), the estimate is valid with  $\hat{d}_\rho$  denoting the 'truncated' Pompeiu-Hausdorff distance in  $L_\infty(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$  and  $D_f^*$  defined by (9).

**Proof:** Let  $\rho > 0$ ,  $\delta > 0$  be selected as in (A2) and (A3),  $x \in \mathcal{X}(\xi) \cap \rho\mathbb{B}$  and  $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$  be such that  $\|\tilde{\xi} - \xi\|_r < \delta$ .

With the same arguments as in the proof of [20, Theorem 2.1], there exists  $\tilde{x} \in \mathcal{X}(\tilde{\xi})$  such that the estimate

$$|\mathbb{E}[x_t | \mathcal{F}_t(\tilde{\xi})] - \tilde{x}_t| \leq \hat{L}_t \left( \sum_{\tau=1}^t \mathbb{E}[|\xi_\tau - \tilde{\xi}_\tau| | \mathcal{F}_\tau(\tilde{\xi})] + \sum_{\tau=2}^{t-1} \mathbb{E}[|x_\tau - \mathbb{E}[x_\tau | \mathcal{F}_\tau(\tilde{\xi})]| | \mathcal{F}_{\tau+1}(\tilde{\xi})] \right) \quad (10)$$

holds  $\mathbb{P}$ -almost surely with some positive constant  $\hat{L}_t$  for  $t = 1, \dots, T$ . Note that  $r' < \infty$  means that only costs and/or right-hand sides in (3) are random and that the

first sum on the right-hand side of (10) disappears if only costs are random. From the definition of  $r'$  we know that  $r \neq r'$  may occur only in the latter case.

Hence, together with the estimate

$$|x_t - \tilde{x}_t| \leq |x_t - \mathbb{E}[x_t | \mathcal{F}_t(\tilde{\xi})]| + |\mathbb{E}[x_t | \mathcal{F}_t(\tilde{\xi})] - \tilde{x}_t|$$

$\mathbb{P}$ -almost surely and for all  $t = 1, \dots, T$ , (10) implies for all pairs  $(r, r')$  with  $r' \in [1, \infty)$  that

$$\mathbb{E}[|x_t - \tilde{x}_t|^{r'}] \leq L_t \left( \sum_{\tau=1}^t \mathbb{E}[|\xi_\tau - \tilde{\xi}_\tau|^r] + \sum_{\tau=2}^t \mathbb{E}[|x_\tau - \mathbb{E}[x_\tau | \mathcal{F}_\tau(\tilde{\xi})]|^{r'}] \right)$$

holds with certain constants  $L_t$ ,  $t = 1, \dots, T$ . We conclude

$$\|x - \tilde{x}\|_{r'} \leq L(\|\xi - \tilde{\xi}\|_r + \rho D_f^*(\xi, \tilde{\xi})).$$

with some constant  $L > 0$ . The second estimate follows by interchanging the role of the pairs  $(x, \tilde{\xi})$  and  $(\tilde{x}, \xi)$ . If (A3)' is satisfied instead of (A3), the changes are obvious.  $\square$

Now, we are ready to establish a Lipschitz property of approximate solution sets.

**Theorem 2.4** *Let (A1), (A2) and (A3) be satisfied with  $r' \in [1, \infty)$  and the sets  $\mathcal{X}_1(\tilde{\xi}_1)$  be nonempty and uniformly bounded in  $\mathbb{R}^{m_1}$  if  $|\tilde{\xi}_1 - \xi_1| \leq \delta$ . Assume that the solution sets  $S(\xi)$  and  $S(\tilde{\xi})$  are nonempty for some  $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$  with  $\|\xi - \tilde{\xi}\|_r \leq \delta$  (with  $\delta > 0$  from (A3)). Then there exist  $\bar{L} > 0$  and  $\bar{\varepsilon} > 0$  such that*

$$d_\infty(S_\varepsilon(\xi), S_\varepsilon(\tilde{\xi})) \leq \frac{\bar{L}}{\varepsilon} (\|\xi - \tilde{\xi}\|_r + D_f^*(\xi, \tilde{\xi})) \quad (11)$$

holds for any  $\varepsilon \in (0, \bar{\varepsilon})$ .

**Proof:** Let  $\rho_0 \geq 1$  be chosen such that the  $L_{r'}$ -bounded set  $B$  in (A3) is contained in  $\rho_0 \mathbb{B}$  (with  $\mathbb{B}$  denoting the unit ball in  $L_{r'}$ ) and  $\min\{v(\xi), v(\tilde{\xi})\} \geq -\rho_0$ . Let  $\rho > \rho_0$ ,  $\bar{\varepsilon} = \min\{\alpha, \rho - \rho_0\}$  and  $0 < \varepsilon < \bar{\varepsilon}$ . Let  $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$  with  $\|\xi - \tilde{\xi}\|_r \leq \delta$ . Then the assumptions of Theorem 7.69 in [37] are satisfied for the functions  $F(\xi, \cdot)$  and  $F(\tilde{\xi}, \cdot)$ . We note that most of the results in [37] are stated in finite-dimensional spaces. However, the proof of [37, Theorem 7.69] carries over to linear normed spaces (see also [1, Theorem 4.3]). We obtain from the proof the inclusion

$$S_\varepsilon(\xi) = S_\varepsilon(\xi) \cap \rho \mathbb{B} \subseteq S_\varepsilon(\tilde{\xi}) + \frac{2\eta}{\varepsilon + 2\eta} 2\rho \mathbb{B} \subseteq S_\varepsilon(\tilde{\xi}) + \frac{4\rho}{\varepsilon} \eta \mathbb{B}, \quad (12)$$

for all  $\eta > \hat{d}_{\rho+\varepsilon}^+(F(\xi, \cdot), F(\tilde{\xi}, \cdot))$ , where the auxiliary epi-distance  $\hat{d}_\rho^+(F(\xi, \cdot), F(\tilde{\xi}, \cdot))$  is defined as the infimum of all  $\eta \geq 0$  such that for all  $x, \tilde{x} \in \rho \mathbb{B}$ ,

$$\min_{\tilde{y} \in \mathbb{B}(x, \eta)} F(\tilde{\xi}, \tilde{y}) \leq \max\{F(\xi, x), -\rho\} + \eta \quad (13)$$

$$\min_{y \in \mathbb{B}(\tilde{x}, \eta)} F(\xi, y) \leq \max\{F(\tilde{\xi}, \tilde{x}), -\rho\} + \eta. \quad (14)$$

The estimate (12) implies

$$S_\varepsilon(\xi) \subseteq S_\varepsilon(\tilde{\xi}) + \frac{4\rho}{\varepsilon} \hat{d}_{\rho+\varepsilon}^+(F(\xi, \cdot), F(\tilde{\xi}, \cdot)) \mathbb{B}.$$

Since the same argument works with  $\xi$  and  $\tilde{\xi}$  interchanged, we obtain

$$d_\infty(S_\varepsilon(\xi), S_\varepsilon(\tilde{\xi})) \leq \frac{4\rho}{\varepsilon} \hat{d}_{\rho+\varepsilon}^+(F(\xi, \cdot), F(\tilde{\xi}, \cdot))$$

and it remains to estimate  $\hat{d}_{\rho+\varepsilon}^+(F(\xi, \cdot), F(\tilde{\xi}, \cdot))$ . Let  $\eta > \hat{d}_{\rho+\varepsilon}^+(F(\xi, \cdot), F(\tilde{\xi}, \cdot))$  and  $x \in \mathcal{X}(\xi)$ . Proposition 2.3 implies the existence of  $\tilde{x} \in \mathcal{X}(\tilde{\xi})$  such that

$$\|x - \tilde{x}\|_{r'} \leq L(\|\xi - \tilde{\xi}\|_r + \|x\|_{r'} D_f^*(\xi, \tilde{\xi})).$$

In order to check condition (13), we have to distinguish three cases, namely, that randomness appears in costs and right-hand sides, only in costs, and only in right-hand sides. Next we consider the first case, i.e.,  $r = r' = 2$ , and obtain as in the proof of [20, Theorem 2.1] the estimate

$$\begin{aligned} F(\tilde{\xi}, \tilde{x}) &\leq F(\xi, x) + |F(\tilde{\xi}, \tilde{x}) - F(\tilde{\xi}, x)| + |F(\tilde{\xi}, x) - F(\xi, x)| \\ &\leq F(\xi, x) + \left| \mathbb{E} \left[ \sum_{t=1}^T \langle b_t(\tilde{\xi}_t), \tilde{x}_t - \mathbb{E}[x_t | \mathcal{F}_t(\tilde{\xi})] \rangle \right] \right| + \left| \mathbb{E} \left[ \sum_{t=1}^T \langle b_t(\tilde{\xi}_t) - b_t(\xi_t), x_t \rangle \right] \right| \\ &\leq F(\xi, x) + \hat{K} \left( \left( \sum_{t=1}^T (1 + \mathbb{E}[|\tilde{\xi}_t|^2]) \right)^{\frac{1}{2}} \left( \sum_{t=1}^T \mathbb{E}[|\tilde{x}_t - \mathbb{E}[x_t | \mathcal{F}_t(\tilde{\xi})]|^2] \right)^{\frac{1}{2}} \right. \\ &\quad \left. + \|\tilde{\xi} - \xi\|_2 \|x\|_2 \right) \\ &\leq F(\xi, x) + \hat{L} \left( \rho \|\tilde{\xi} - \xi\|_2 + \sum_{t=2}^{T-1} \|x_t - \mathbb{E}[x_t | \mathcal{F}_t(\tilde{\xi})]\|_2 \right) \\ &\leq F(\xi, x) + L\rho(\|\tilde{\xi} - \xi\|_2 + D_f^*(\xi, \tilde{\xi})) \end{aligned}$$

with certain constants  $\hat{K}$ ,  $\hat{L}$  and  $L$  (depending on  $\|\xi\|_2$ ), where the Cauchy-Schwarz inequality, (A3) and the estimate (10) are used. Hence, condition (13) is satisfied if

$$\eta = L\rho(\|\tilde{\xi} - \xi\|_2 + D_f^*(\xi, \tilde{\xi}))$$

holds with certain constant  $L > 0$ . The same estimate holds in the remaining two cases and when checking condition (14) (possibly with different constants). Taking the maximal constant  $L > 0$  we conclude

$$\hat{d}_{\rho+\varepsilon}^+(F(\xi, \cdot), F(\tilde{\xi}, \cdot)) \leq L\rho(\|\tilde{\xi} - \xi\|_r + D_f^*(\xi, \tilde{\xi}))$$

and, hence,

$$d_\infty(S_\varepsilon(\xi), S_\varepsilon(\tilde{\xi})) \leq \frac{4L\rho^2}{\varepsilon} (\|\tilde{\xi} - \xi\|_r + D_f^*(\xi, \tilde{\xi})).$$

Setting  $\bar{L} = 4L\rho^2$  completes the proof.  $\square$

For solution sets the situation is less comfortable. Stability of solutions can only be derived with respect to the weak topology  $\sigma(L_{r'}, L_r)$ .

**Theorem 2.5** *Let (A1), (A2) and (A3) be satisfied with  $r' \in (1, \infty)$  and the sets  $\mathcal{X}_1(\tilde{\xi}_1)$  be nonempty and uniformly bounded in  $\mathbb{R}^{m_1}$  if  $|\tilde{\xi}_1 - \xi_1| \leq \delta$  (with  $\delta > 0$  from (A3)). If  $(\xi^{(n)})$  is a sequence in  $L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$  converging to  $\xi$  in  $L_r$  and with respect to  $D_f^*$  and if  $(x^{(n)})$  is a sequence of solutions of the approximate problems, i.e.,  $x^{(n)} \in S(\xi^{(n)})$ , then there exists a subsequence  $(x^{(n_k)})$  of  $(x^{(n)})$  that converges with respect to the weak topology  $\sigma(L_{r'}, L_r)$  to some element of  $S(\xi)$ . If  $S(\xi)$  is a singleton, the sequence  $(x^{(n)})$  converges with respect to the weak topology  $\sigma(L_{r'}, L_r)$  to the unique solution of (3).*

**Proof:** Let  $(\xi^{(n)})$  and  $(x^{(n)})$  be selected as above. Since there exists  $n_0 \in \mathbb{N}$  such that  $\|\xi^{(n)} - \xi\|_r \leq \delta$  and  $x^{(n)} \in S_\alpha(\xi^{(n)})$  for any  $n \geq n_0$ , where  $\alpha > 0$  and  $\delta > 0$  are chosen as in (A3), the sequence  $(x^{(n)})$  is contained in a bounded set of the reflexive Banach space  $L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ . Hence, there exists a subsequence  $(x^{(n_k)})$  of  $(x^{(n)})$  that converges with respect to the weak topology  $\sigma(L_{r'}, L_r)$  to some element  $x^*$  in  $L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ . Theorem 2.2 implies

$$v(\xi^{(n_k)}) = F(\xi^{(n_k)}, x^{(n_k)}) = \mathbb{E} \left[ \sum_{t=1}^T \langle b_t(\xi_t^{(n_k)}), x_t^{(n_k)} \rangle \right] \rightarrow v(\xi).$$

Due to the norm convergence of  $(\xi^{(n_k)})$  and the weak convergence of  $(x^{(n_k)})$ , we also obtain

$$\mathbb{E} \left[ \sum_{t=1}^T \langle b_t(\xi_t^{(n_k)}), x_t^{(n_k)} \rangle \right] \rightarrow \mathbb{E} \left[ \sum_{t=1}^T \langle b_t(\xi_t), x_t^* \rangle \right].$$

Hence, it remains to show that  $x^*$  is feasible for (3), i.e.,  $x^* \in \mathcal{X}(\xi)$  and  $x^* \in \mathcal{N}_{r'}(\xi)$ .

In the present situation, the set  $\mathcal{X}(\xi)$  is of the form

$$\mathcal{X}(\xi) = \{x \in L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m) : x \in X, Ax = h(\xi)\}, \quad (15)$$

where  $X := \times_{t=1}^T X_t$ ,  $h(\xi) := (h_1(\xi_1), \dots, h_T(\xi_T))$  and

$$A := \begin{pmatrix} A_{1,0} & 0 & 0 & \cdots & 0 & 0 & 0 \\ A_{2,1} & A_{2,0} & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & A_{T,1} & A_{T,0} \end{pmatrix}.$$

The graph of  $\mathcal{X}$ , i.e.,  $\text{graph } \mathcal{X} = \{(x, \xi) \in L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m) \times L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s) | x \in \mathcal{X}(\xi)\}$  is closed and convex. Since  $(\xi^{(n_k)})$  norm converges in  $L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$  to  $\xi$  and  $(x^{(n_k)})$  weakly converges to  $x^*$ , the sequence  $((x^{(n_k)}, \xi^{(n_k)}))$  of pairs in  $\text{graph } \mathcal{X}$  converges weakly to  $(x^*, \xi)$ . Due to the closedness and convexity of  $\text{graph } \mathcal{X}$ , Mazur's theorem [14, Theorem 3.19] implies that  $\text{graph } \mathcal{X}$  is weakly closed and, thus,  $(x^*, \xi) \in \text{graph } \mathcal{X}$  or  $x^* \in \mathcal{X}(\xi)$ .

Finally, we have to show that  $x^*$  belongs to  $\mathcal{N}_{r'}(\xi)$ . For any  $y \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$  we obtain the estimate

$$\left| \mathbb{E} \left[ \sum_{t=1}^T \langle y_t, x_t^* - \mathbb{E}[x_t^* | \mathcal{F}_t(\xi)] \rangle \right] \right| \leq \left| \sum_{t=1}^T \mathbb{E}[\langle y_t, x_t^* - x_t^{(n_k)} \rangle] \right|$$

$$\begin{aligned}
& + \left| \sum_{t=1}^T \mathbb{E}[\langle y_t, x_t^{(n_k)} - \mathbb{E}[x_t^{(n_k)} | \mathcal{F}_t(\xi)] \rangle] \right| \\
& + \left| \sum_{t=1}^T \mathbb{E}[\langle y_t, \mathbb{E}[x_t^{(n_k)}] - x_t^* | \mathcal{F}_t(\xi) \rangle] \right| \\
& \leq 2 \left| \sum_{t=1}^T \mathbb{E}[\langle y_t, x_t^* - x_t^{(n_k)} \rangle] \right| \\
& \quad + \max_{t=1, \dots, T} \|y_t\|_r \sum_{t=2}^T \|x_t^{(n_k)} - \mathbb{E}[x_t^{(n_k)} | \mathcal{F}_t(\xi)]\|_{r'}.
\end{aligned}$$

The first term on the right-hand side converges to 0 for  $k$  tending to  $\infty$  as the sequence  $(x^{(n_k)})$  converges weakly to  $x^*$ . The second term converges to 0 due to the estimate (8) since  $(D_f^*(\xi, \xi^{(n_k)}))$  also converges to 0. We conclude that

$$\mathbb{E} \left[ \sum_{t=1}^T \langle y_t, x_t^* - \mathbb{E}[x_t^* | \mathcal{F}_t(\xi)] \rangle \right] = 0$$

holds for any  $y \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$  and, hence, that  $x_t^* = \mathbb{E}[x_t^* | \mathcal{F}_t(\xi)]$  for each  $t = 1, \dots, T$ . This means  $x^* \in \mathcal{N}_{r'}(\xi)$ .  $\square$

**Remark 2.6** *Theorem 2.5 remains true if the filtration distance  $D_f^*$  is replaced by the weaker distance*

$$\hat{D}_f(\xi, \tilde{\xi}) = \sup_{\tilde{x} \in S(\tilde{\xi})} \sum_{t=2}^T \|\tilde{x}_t - \mathbb{E}[\tilde{x}_t | \mathcal{F}_t(\xi)]\|_{r'}.$$

Furthermore, if the solutions  $x^{(n)} \in S(\xi^{(n)})$  are adapted to the filtration  $\mathcal{F}_t(\xi)$ ,  $t = 1, \dots, T$ , of the original process  $\xi$  (as in [19, Proposition 5.5]), the convergence of  $(\xi^{(n)})$  to  $\xi$  in  $L_r$  is sufficient for the weak convergence of some subsequence of  $(x^{(n)})$  to some element of  $S(\xi)$  (in the sense of  $\sigma(L_{r'}, L_r)$ ).

**Remark 2.7** *The stability analysis of (linear) two-stage stochastic programs (see, e.g., [35, Section 3.1], [38]) mostly studied the continuity behavior of first-stage (approximate) solution sets. Hence, for the specific case  $T = 2$ , our stability results in Theorems 2.4 and 2.5 extend earlier work because they concern first- and second-stage solutions. The new important assumption is (A3), i.e., the level-boundedness of the objective (locally uniformly at  $\xi$ ) with respect to both first- and second-stage variables.*

**Remark 2.8** *In many applications of stochastic programming it is of interest to develop risk-averse models (e.g., in electricity risk management and in finance). For example, this can be achieved if the expectation in the objective of (1) is replaced by a (convex) risk functional (measure). Typically, risk functionals are inherently nonlinear. If, however, a multiperiod polyhedral risk functional [11] replaces the expectation in (1), the resulting risk-averse stochastic program may be reformulated as a linear multistage stochastic program of the form (1) by introducing new state variables and (linear) constraints (see [11, Section 4]). Moreover, it is shown in [12] that the stability*

behavior of the reformulation does not change (when compared with the original problem with expectation objective) if the multiperiod polyhedral (convex) risk functional has bounded  $L_1$ -level sets. The latter property is shared by the Conditional or Average Value-at-Risk and several of its multiperiod extensions [12, Section 4].

### 3 Generating scenario trees

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

$$\|\xi - \xi_{\text{tr}}\|_r \quad \text{and} \quad D_{\text{f}}^*(\xi, \xi_{\text{tr}}) \quad (16)$$

are small and, hence, the optimal values  $v(\xi)$  and  $v(\xi_{\text{tr}})$ , and the approximate solution sets  $S_\varepsilon(\xi)$  and  $S_\varepsilon(\xi_{\text{tr}})$  are close to each other according to Theorems 2.2 and 2.4, respectively.

The idea is to start with a good initial approximation  $\hat{\xi}$  of  $\xi$  having a finite number of scenarios  $\xi^i = (\xi_1^i, \dots, \xi_T^i) \in \mathbb{R}^{Td}$  with probabilities  $p_i > 0$ ,  $i = 1, \dots, N$ , and common root, i.e.,  $\xi_1^1 = \dots = \xi_1^N =: \xi_1^*$ . These scenarios might be obtained by quantization techniques [16] or by sampling or resampling techniques based on parametric or nonparametric stochastic models of  $\xi$ .

In the following we assume that

$$\|\xi - \hat{\xi}\|_r + D_{\text{f}}^*(\xi, \hat{\xi}) \leq \varepsilon \quad (17)$$

holds for some given (initial) tolerance  $\varepsilon > 0$ . For example, condition (17) may be satisfied for  $D_{\text{f}}^*$  given by (9) and for any tolerance  $\varepsilon > 0$  if  $\hat{\xi}$  is obtained by sampling from a finite set with sufficiently large sample size (see [19, Example 5.3]).

Next we describe an algorithmic procedure that starts from  $\hat{\xi}$  and ends up with a scenario tree process  $\xi_{\text{tr}}$  having the same root node  $\xi_1^*$ , less nodes than  $\hat{\xi}$  and allowing for constructive estimates of

$$\|\hat{\xi} - \xi_{\text{tr}}\|_r.$$

The idea of the algorithm consists in forming clusters of scenarios based on scenario reduction on the time horizon  $\{1, \dots, t\}$  recursively for increasing time  $t$ .

To this end, the  $L_r$ -seminorm  $\|\cdot\|_{r,t}$  on  $L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$  (with  $s = Td$ ) given by

$$\|\xi\|_{r,t} := (\mathbb{E}[|\xi|_t^r])^{\frac{1}{r}} \quad (18)$$

is used at step  $t$ , where  $|\cdot|_t$  is a seminorm on  $\mathbb{R}^s$  which, for each  $\xi = (\xi_1, \dots, \xi_T) \in \mathbb{R}^s$ , is given by  $|\xi|_t := |(\xi_1, \dots, \xi_t, 0, \dots, 0)|$ .

The following procedure determines recursively stochastic processes  $\hat{\xi}^t$  having scenarios  $\hat{\xi}^{t,i}$  endowed with probabilities  $p_i$ ,  $i \in I := \{1, \dots, N\}$ , and, in addition, partitions  $\mathcal{C}_t = \{C_t^1, \dots, C_t^{K_t}\}$  of the index set  $I$ , i.e.,

$$C_t^k \cap C_t^{k'} = \emptyset \quad (k \neq k') \quad \text{and} \quad \bigcup_{k=1}^{K_t} C_t^k = I. \quad (19)$$

The index sets  $C_t^k \in \mathcal{C}_t$ ,  $k = 1, \dots, K_t$ , characterize clusters of scenarios. 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  $t > 1$ ) we consider each cluster  $C_{t-1}^k$ , i.e., each scenario subset  $\{\hat{\xi}^{t-1,i}\}_{i \in C_{t-1}^k}$ , separately and delete scenarios  $\{\hat{\xi}^{t-1,j}\}_{j \in J_t^k}$  by the forward selection algorithm of [18] such that

$$\left( \sum_{k=1}^{K_{t-1}} \sum_{j \in J_t^k} p_j \min_{i \in I_t^k} |\hat{\xi}^{t-1,i} - \hat{\xi}^{t-1,j}|^r \right)^{\frac{1}{r}}$$

is bounded from above by some prescribed tolerance. Here, the index set  $I_t^k$  of remaining scenarios is given by

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

As in the general scenario reduction procedure in [18], the index set  $J_t^k$  is subdivided into index sets  $J_{t,i}^k$ ,  $i \in I_t^k$  such that

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

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

$$\alpha_t(j) = \begin{cases} i_t^k(j) & , j \in J_t^k, k = 1, \dots, K_{t-1} \\ j & , \text{otherwise.} \end{cases} \quad (20)$$

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

$$\hat{\xi}_\tau^{t,i} = \begin{cases} \xi_\tau^{\alpha_\tau(i)} & , \tau \leq t \\ \xi_\tau^i & , \text{otherwise} \end{cases} \quad (21)$$

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

$$\mathcal{C}_t = \{\alpha_t^{-1}(i) : i \in I_t^k, k = 1, \dots, K_{t-1}\}, \quad (22)$$

i.e., each element of the index sets  $I_t^k$  defines a new cluster and the new partition  $\mathcal{C}_t$  is a refinement of the former partition  $\mathcal{C}_{t-1}$ .

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

$$\xi_{\text{tr}}^k = (\xi_1^*, \xi_2^{\alpha_2(i)}, \dots, \xi_t^{\alpha_t(i)}, \dots, \xi_T^{\alpha_T(i)}) \quad \text{and} \quad \pi_T^k = \sum_{j \in C_T^k} p_j \quad \text{if } i \in C_T^k \quad (23)$$

for each  $k = 1, \dots, K_T$ . The index set  $I_t$  of realizations of  $\xi_{\text{tr}}$  is given by

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

For each  $t \in \{1, \dots, T\}$  and each  $i \in I$  there exists an unique index  $k_t(i) \in \{1, \dots, K_t\}$  such that  $i \in C_t^{k_t(i)}$ . Moreover, we have  $C_t^{k_t(i)} = \{i\} \cup J_{t,i}^{k_t(i)}$  for each  $i \in I_t$ . The

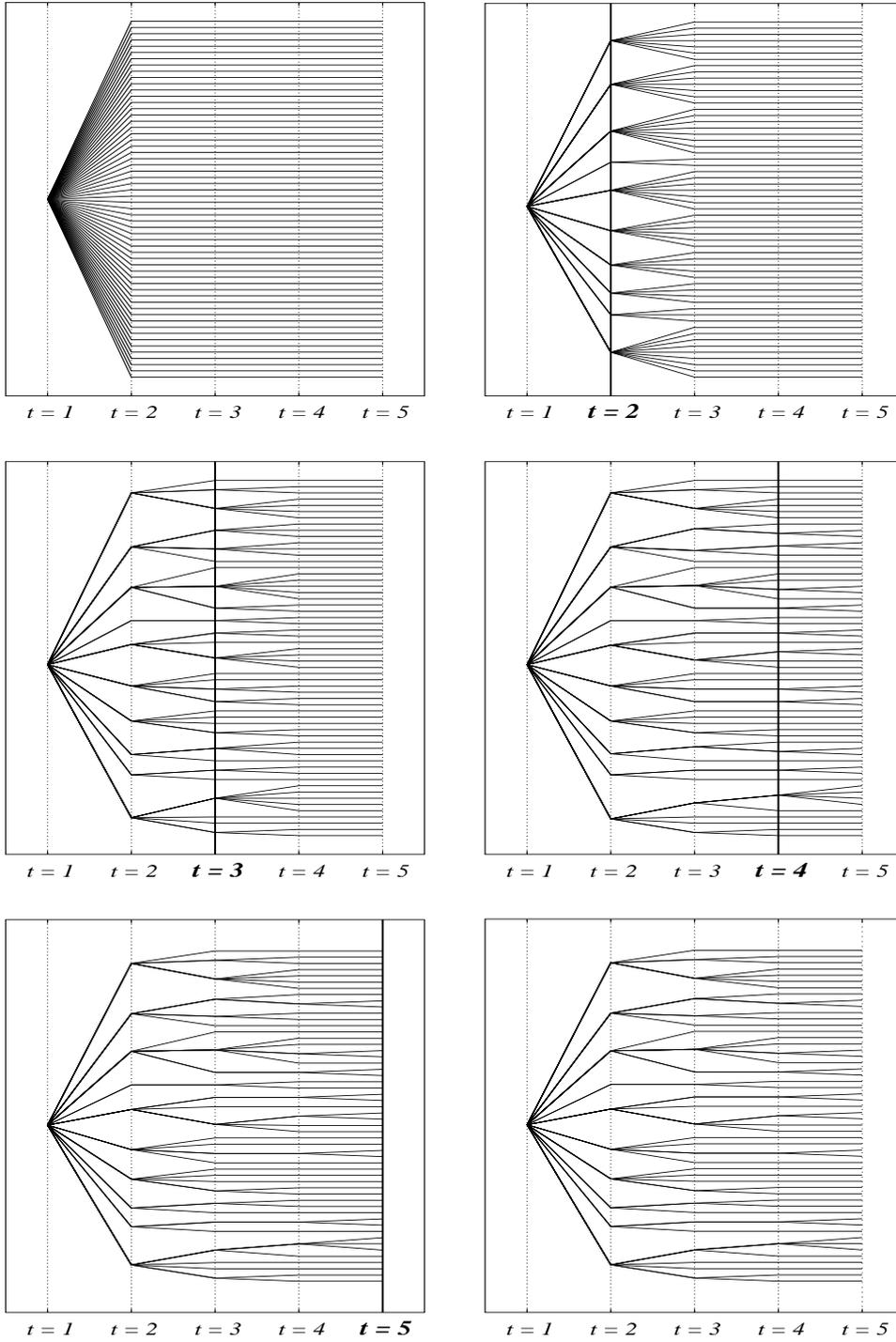


Figure 1: Illustration of the tree construction for an example with  $T=5$  time periods

probability of the  $i$ -th realization of  $\xi_t^{\text{tr}}$  is  $\pi_t^i = \sum_{j \in C_t^{k_t(i)}} p_j$ . The branching degree of scenario  $i \in I_{t-1}$  coincides with the cardinality of  $I_t^{k_t(i)}$ .

The next result quantifies the relative error of the  $t$ -th construction step and is proved in [19, Theorem 3.4].

**Theorem 3.1** *Let the stochastic process  $\hat{\xi}$  with fixed initial node  $\xi_1^*$ , scenarios  $\xi^i$  and probabilities  $p_i$ ,  $i = 1, \dots, N$ , be given. Let  $\xi_{\text{tr}}$  be the stochastic process with scenarios  $\xi_{\text{tr}}^k = (\xi_1^*, \xi_2^{\alpha_2(i)}, \dots, \xi_t^{\alpha_t(i)}, \dots, \xi_T^{\alpha_T(i)})$  and probabilities  $\pi_T^k$  if  $i \in C_T^k$ ,  $k = 1, \dots, K_T$ . Then we have*

$$\|\hat{\xi} - \xi_{\text{tr}}\|_r \leq \sum_{t=2}^T \left( \sum_{k=1}^{K_{t-1}} \sum_{j \in J_t^k} p_j \min_{i \in I_t^k} |\xi_t^i - \xi_t^j|^r \right)^{\frac{1}{r}}. \quad (24)$$

Next, 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 3.2 (forward tree construction)**

Let  $N$  scenarios  $\xi^i$  with probabilities  $p_i$ ,  $i = 1, \dots, N$ , fixed root  $\xi_1^* \in \mathbb{R}^d$  and probability distribution  $P$ ,  $r \geq 1$  and tolerances  $\varepsilon_r, \varepsilon_t$ ,  $t = 2, \dots, T$ , be given such that  $\sum_{t=2}^T \varepsilon_t \leq \varepsilon_r$ .

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

**Step  $t$ :** Let  $\mathcal{C}_{t-1} = \{C_{t-1}^1, \dots, C_{t-1}^{K_{t-1}}\}$ . 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 (20) and a stochastic process  $\hat{\xi}^t$  having  $N$  scenarios  $\hat{\xi}^{t,i}$  with probabilities  $p_i$  according to (21) and such that

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

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

**Step  $T+1$ :** Let  $\mathcal{C}_T = \{C_T^1, \dots, C_T^{K_T}\}$ . Construct a stochastic process  $\xi_{\text{tr}}$  having  $K_T$  scenarios  $\xi_{\text{tr}}^k$  such that  $\xi_{\text{tr},t}^k := \xi_t^{\alpha_t(i)}$ ,  $t = 1, \dots, T$ , if  $i \in C_T^k$  with probabilities  $\pi_T^k$  according to (23),  $k = 1, \dots, K_T$ .

While the first picture in Figure 1 illustrates the process  $\hat{\xi}$ , the  $t$ -th picture corresponds to the situation after Step  $t$ ,  $t = 2, 3, 4, 5$  of the algorithm. The final picture corresponds to Step 6 and illustrates the final scenario tree  $\xi_{\text{tr}}$ . The proof of the following corollary is also given in [19].

**Corollary 3.3** *Let a stochastic process  $\hat{\xi}$  with fixed initial node  $\xi_1^*$ , scenarios  $\xi^i$  and probabilities  $p_i$ ,  $i = 1, \dots, N$ , be given. If  $\xi_{\text{tr}}$  is constructed by Algorithm 3.2, we have*

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

The next results states that the distance  $|v(\xi) - v(\xi_{\text{tr}})|$  of optimal values gets small if the initial tolerance  $\varepsilon$  in (17) as well as  $\varepsilon_r$  are small.

**Theorem 3.4** *Let (A1), (A2) and (A3) be satisfied with  $r' \in [1, \infty)$  and the sets  $\mathcal{X}_1(\tilde{\xi}_1)$  be nonempty and uniformly bounded in  $\mathbb{R}^{m_1}$  if  $|\tilde{\xi}_1 - \xi_1| \leq \delta$ . Let  $L > 0$ ,  $\delta > 0$  and  $C > 0$  be the constants appearing in Theorem 2.2 and (8).*

*If  $(\varepsilon_r^{(n)})$  is a sequence tending to 0 such that the corresponding tolerances  $\varepsilon_t^{(n)}$  in Algorithm 3.2 are nonincreasing for all  $t = 2, \dots, T$ , the corresponding sequence  $(\xi_{\text{tr}}^{(n)})$  has the property*

$$\limsup_{n \rightarrow \infty} |v(\xi) - v(\xi_{\text{tr}}^{(n)})| \leq L \max\{1, C\} \varepsilon, \quad (25)$$

where  $\varepsilon > 0$  is the initial tolerance in (17).

**Proof:** It is shown in [19, Proposition 5.2] that the estimate

$$|v(\xi) - v(\xi_{\text{tr}}^{(n)})| \leq L(\varepsilon_r^{(n)} + \|\xi - \hat{\xi}\|_r + C D_f^*(\xi, \hat{\xi}) + C D_f^*(\hat{\xi}, \xi_{\text{tr}}^{(n)})) \quad (26)$$

is valid and that  $D_f^*(\hat{\xi}, \xi_{\text{tr}}^{(n)})$  tends to 0 as  $n \rightarrow \infty$ . We conclude that the estimate (26) implies (25).  $\square$

## 4 Numerical experience

We consider a mean-risk optimization model for electricity portfolios of a German municipal electricity company which consist of the own (thermal) electricity production, the spot market contracts, supply contracts and electricity futures. Stochasticity enters the model via the electricity demand, heat demand, spot prices, and future prices (cf. [13]). Our approach of generating input scenarios in form of a scenario tree consists in developing a statistical model for all stochastic components and in using Algorithm 3.2 started with a finite number of scenarios which are simulated from the statistical model.

### 4.1 Adapting a statistical model

For the stochastic input data of the optimization model (namely, electricity demand, heat demand, and electricity spot prices), we had access to historical data (from a yearly period of hourly observations). Due to climatic influences the demands are characterized by typical yearly cycles with high (low) demand during winter (summer) time. Furthermore, the demands contain weekly cycles due to varying consumption behavior of private and industrial customers on working days and weekends. The intra-day profiles reflect a characteristic consumption behavior of the customers with seasonal differences. Outliers can be observed on public holidays, on days between holidays, and on days with extreme climatic conditions. Spot prices are affected by climatic conditions, economic activities, local power producers, customer behavior etc. An all-embracing modeling is hardly possible. However, spot prices are also characterized by typical yearly cycles with high (lower) prices during winter (summer) time, and they show weekly and daily cycles, too. Hence, the (price and demand) data was de-

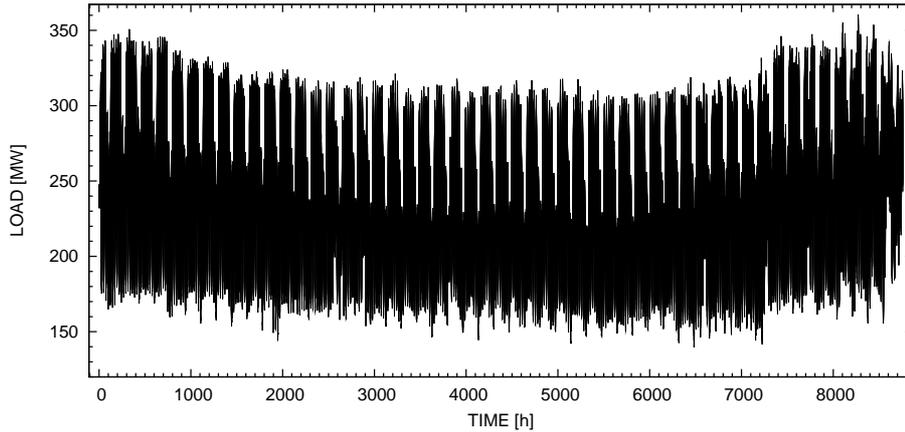


Figure 2: Time plot of load profile for one year

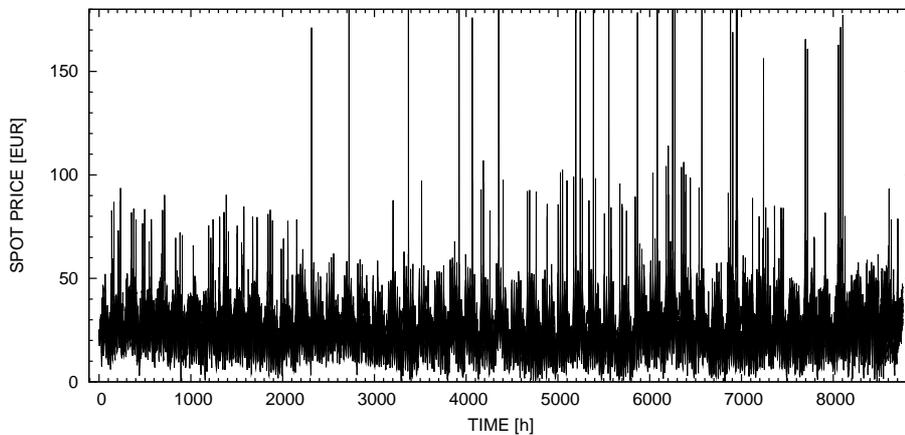


Figure 3: Time plot of spot price profile for one year

composed into intra-day profiles and daily average values. While the intra-day profiles are modeled by a distribution-free resampling procedure based on standard clustering algorithms, a three dimensional time series model was developed for the daily average values. The latter consists of deterministic trend functions and a trivariate autoregressive moving-average (ARMA) model for the (stationary) residual time series (see [13] for details). Then an arbitrary number of three dimensional scenarios can easily be obtained by simulating white noise processes for the ARMA model and by adding on afterwards the trend functions, the matched intra-day profiles from the clusters and extreme price outliers modeled by a discrete jump-diffusion process with time-varying jump parameters. Future price scenarios are directly derived from those for the spot prices.

## 4.2 Construction of input scenario trees

The three dimensional (electricity demand, heat demand, spot price) scenarios form the initial scenario set and serve as inputs for the forward tree construction (Algorithm

3.2). In our test series we started with a total number of 100 sample scenarios for a one year time horizon with hourly discretization. Table 1 displays the dimension of the

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

Table 1: Dimension of simulated input scenarios

simulated input scenarios. Due to the fact that electricity future products can only be traded monthly, branching was allowed only at the end of each month. Scenario trees were generated by Algorithm 3.2 for  $r = r' = 2$  and different relative reduction levels  $\varepsilon_{\text{rel}}$ . The relative levels are given by

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

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

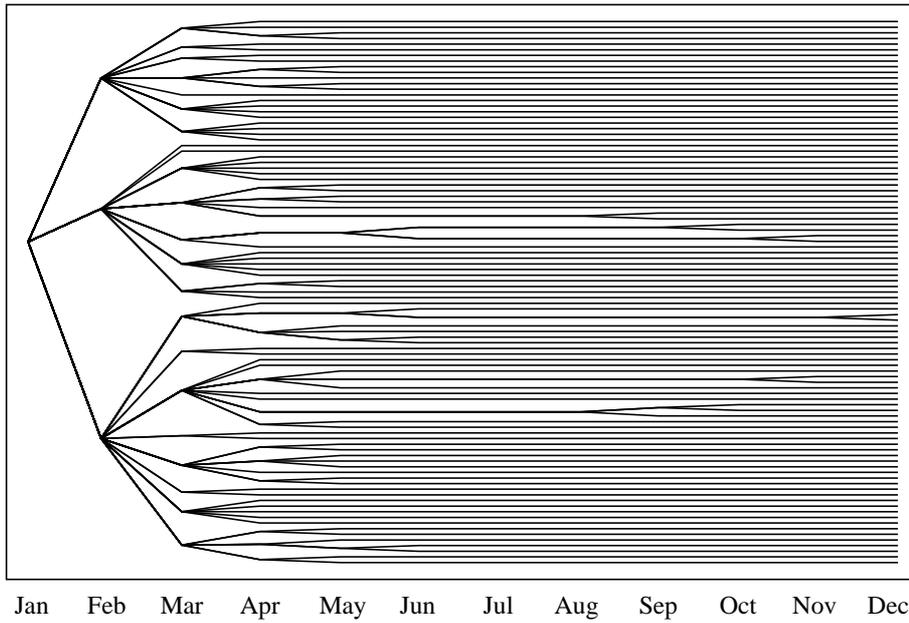
$$\varepsilon_t^r = \frac{\varepsilon^r}{T} \left[ 1 + \bar{q} \left( \frac{1}{2} - \frac{t}{T} \right) \right], \quad t = 2, \dots, T, \quad r = 2, \quad (27)$$

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

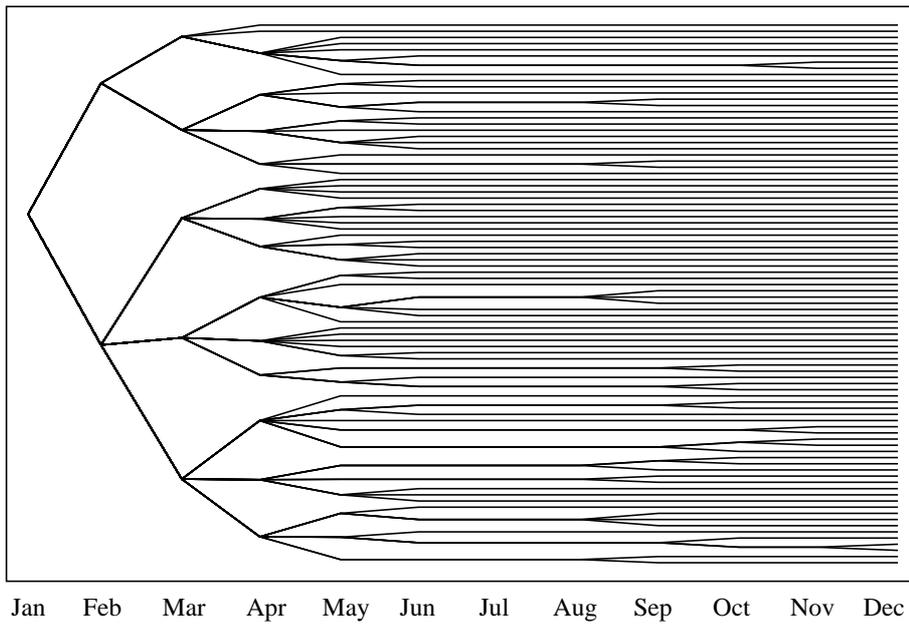
Table 2 displays the results of our test runs with different relative reduction levels. As expected, for very small reduction levels, the reduction affects only a few scenarios. Furthermore, the number of nodes decreases considerably if the reduction level is increased. The computing times of less than 30 seconds already include approximately 20 seconds for computing distances of all scenario pairs that are needed in all calculations. Figure 4 illustrates the scenario trees obtained for reduction levels of 40 percent and 55 percent, respectively.

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

Table 2: Numerical results of Algorithm 3.2 for yearly demand-price scenario trees



a) Forward constructed scenario tree with reduction level  $\varepsilon_{rel} = 0.4$



b) Forward constructed scenario tree with reduction level  $\varepsilon_{rel} = 0.55$

Figure 4: Yearly demand-price scenario trees obtained by Algorithm 3.2

## Acknowledgement:

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.

## References

- [1] Attouch, H., Wets, R. J-B: Quantitative stability of variational systems: III.  $\varepsilon$ -approximate solutions, *Mathematical Programming* 61 (1993), 197–214.
- [2] Casey, M., Sen, S.: The scenario generation algorithm for multistage stochastic linear programming, *Mathematics of Operations Research* 30 (2005), 615–631.
- [3] Chiralaksanakul, A., Morton, D. P.: Assessing policy quality in multi-stage stochastic programming, *Stochastic Programming E-Print Series* 12–2004 (<[www.speps.org](http://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] Dantzig, G. B.: Linear programming under uncertainty, *Management Science* 1 (1955), 197–206.
- [6] Dantzig, G. B.: *Linear Programming and Extensions*, Princeton University Press, Princeton 1963.
- [7] Dempster, M.A.H.: Sequential importance sampling algorithms for dynamic stochastic programming, *Zap. Nauchn. Semin. POMI* 312 (2004), 94–129.
- [8] Dupačová, J., Consigli, G., Wallace, S. W.: Scenarios for multistage stochastic programs, *Annals of Operations Research* 100 (2000), 25–53.
- [9] Dupačová, J., Gröwe-Kuska, N., Römisch, W.: Scenario reduction in stochastic programming: An approach using probability metrics, *Mathematical Programming* 95 (2003), 493–511.
- [10] Edirisinghe, N.C.P.: Bound-based approximations in multistage stochastic programming: Nonanticipativity aggregation, *Annals of Operations Research* 85 (1999), 103–127.
- [11] Eichhorn, A., Römisch, W.: Polyhedral risk measures in stochastic programming, *SIAM Journal on Optimization* 16 (2005), 69–95.
- [12] Eichhorn, A.; Römisch, W.: Stability of multistage stochastic programs incorporating polyhedral risk measures, *Optimization* 57 (2008), 295–318.
- [13] Eichhorn, A., Römisch, W., Wegner, I.: Mean-risk optimization of electricity portfolios using multiperiod polyhedral risk measures, *IEEE St. Petersburg Power Tech* 2005.
- [14] Fabian, M., Habala, P., Hájek, P., Montesinos Santalucia, V., Pelant, J., Zizler, V.: *Functional Analysis and Infinite-Dimensional Geometry*, CMS Books in Mathematics, Springer, New York 2001.

- [15] Frauendorfer, K.: Barycentric scenario trees in convex multistage stochastic programming, *Mathematical Programming* 75 (1996), 277–293.
- [16] Graf, S., Luschgy, H.: *Foundations of quantization for probability distributions*, Lecture Notes in Mathematics, Vol. 1730, Springer, Berlin, 2000.
- [17] Gröwe-Kuska, N., Heitsch, H., Römisch, W.: Scenario reduction and scenario tree construction for power management problems, *IEEE Bologna Power Tech Proceedings* (Borghetti, A., Nucci, C. A., Paolone, M. eds.), 2003 IEEE.
- [18] Heitsch, H., Römisch, W.: Scenario reduction algorithms in stochastic programming, *Computational Optimization and Applications* 24 (2003), 187–206.
- [19] Heitsch, H., Römisch, W.: Scenario tree modelling for multistage stochastic programs, *Mathematical Programming* 118 (2009), 371–406.
- [20] Heitsch, H., Römisch, W., Strugarek, C.: Stability of multistage stochastic programs, *SIAM Journal on Optimization* 17 (2006), 511–525.
- [21] Higle, J. L., Rayco, B., Sen, S.: Stochastic scenario decomposition for multistage stochastic programs, *IMA Journal of Management Mathematics* (to appear).
- [22] Hochreiter, R.: Computational Optimal Management Decisions – The case of Stochastic Programming for Financial Management, Ph.D. Thesis, University of Vienna, 2005.
- [23] Hochreiter, R., Pflug, G.: Financial scenario generation for stochastic multi-stage decision processes as facility location problem, *Annals of Operations Research* 152 (2007), 257–272.
- [24] Høyland, K., Wallace, S. W.: Generating scenario trees for multi-stage decision problems, *Management Science* 47 (2001), 295–307.
- [25] Høyland, K., Kaut, M., Wallace, S. W.: A heuristic for moment-matching scenario generation, *Computational Optimization and Applications* 24 (2003), 169–185.
- [26] Infanger, G.: *Planning under Uncertainty – Solving Large-Scale Stochastic Linear Programs*, Boyd & Fraser Publ. Comp., 1994.
- [27] Kall, P., Mayer, J.: *Stochastic Linear Programming*, Springer, 2005.
- [28] Kaut, M., Wallace, S.W.: Evaluation of scenario-generation methods for stochastic programming, *Pacific Journal of Optimization* 3 (2007), 257–271.
- [29] Kuhn, D.: *Generalized bounds for convex multistage stochastic programs*, Lecture Notes in Economics and Mathematical Systems, Vol. 548, Springer, Berlin, 2005.
- [30] Möller, A., Römisch, W., Weber, K.: Airline network revenue management by multistage stochastic programming, *Computational Management Science* 5 (2008), 355–377.
- [31] Niederreiter, H.: *Random Number Generation and Quasi-Monte Carlo Methods*, CMBS-NSF Regional Conference Series in Applied Mathematics, Vol. 63, SIAM, Philadelphia 1992.

- [32] Pennanen, T.: Epi-convergent discretizations of multistage stochastic programs, *Mathematics of Operations Research* 30 (2005), 245–256.
- [33] Pennanen, T.: Epi-convergent discretizations of multistage stochastic programs via integration quadratures, *Mathematical Programming* 116 (2009), 461–479.
- [34] Pflug, G. Ch.: Scenario tree generation for multiperiod financial optimization by optimal discretization, *Mathematical Programming* 89 (2001), 251–271.
- [35] Rachev, S. T., Römisch, W.: Quantitative stability in stochastic programming: The method of probability metrics, *Mathematics of Operations Research* 27 (2002), 792–818.
- [36] Rockafellar, R. T.: Integral functionals, normal integrands and measurable selections, in: *Nonlinear Operators and the Calculus of Variations* (J. P Gossez et al. eds.), Lecture Notes in Mathematics, Vol. 543, Springer, Berlin 1976, 157–207.
- [37] Rockafellar, R. T., Wets, R. J-B: *Variational Analysis*, Springer-Verlag, Berlin, 1998.
- [38] Römisch, W., Wets, R. J-B: Stability of  $\varepsilon$ -approximate solutions to convex stochastic programs, *SIAM Journal on Optimization* 18 (2007), 961–979.
- [39] Ruszczyński, A., Shapiro, A. (Eds.): *Stochastic Programming*, Handbooks in Operations Research and Management Science, Volume 10, Elsevier, Amsterdam 2003.
- [40] Schmöller, H.: Modellierung von Unsicherheiten bei der mittelfristigen Stromerzeugungs- und Handelsplanung, *Aachener Beiträge zur Energieversorgung*, Band 103, Aachen 2005.
- [41] Shapiro, A.: Inference of statistical bounds for multistage stochastic programming problems, *Math. Meth. Oper. Res.* 58 (2003), 57–68.
- [42] Shapiro, A.: On complexity of multistage stochastic programs, *Operations Research Letters* 34 (2006), 1–8.
- [43] Shapiro, A.: Stochastic programming approach to optimization under uncertainty, *Mathematical Programming* 112 (2008), 183–220.
- [44] Wallace, S. W., Ziemba, W. T. (Eds.): *Applications of Stochastic Programming*, MPS-SIAM Series in Optimization, 2005.
- [45] Zaanen, A. C.: *Linear Analysis*, North-Holland, Amsterdam 1953.