

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 [17] for multistage stochastic programs we show that their 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 a fan of individual scenarios 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

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 [4, 5], and to the recent books [35], [40] and the monograph [24] 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 ξ_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 (ξ_1, \dots, ξ_t) . This property is equivalent to the measurability of x_t with respect to the σ -field $\mathcal{F}_t(\xi) \subseteq \mathcal{F}$ which is generated by (ξ_1, \dots, ξ_t) . Clearly, we have

*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.

$\mathcal{F}_t(\xi) \subseteq \mathcal{F}_{t+1}(\xi)$ 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 ξ_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 [38, 39]) 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 ξ_1 of \mathbb{R}^d . This leads to linear programming models that are very large scale in most cases and can be solved by linear programming techniques, in particular, by decomposition methods that exploit specific structures of the model. We refer to [35, Chapter 3] for a recent survey.

Presently, there exist several approaches to generate scenario trees for multistage stochastic programs (see [7] for a survey). They are based on several different principles. We mention here (i) bound-based constructions [1, 9, 12, 26], (ii) Monte Carlo-based schemes [2, 37, 39] or Quasi Monte Carlo-based methods [29, 30], (iii) (EVPI-based) Sampling within decomposition schemes [3, 6, 18, 23], (iv) the target/moment-matching principle [21, 22], and (v) probability metric based approximations [14, 16, 19, 20, 31].

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 [1, 26] belonging to (i) also offer convergence arguments (restricted to linear models containing only stochasticity in right-hand sides in [1] and to convex models whose stochasticity is assumed to follow some linear block-diagonal autoregressive process with compact supports in [26]). 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 [37] and the complexity is discussed in [38]. The Quasi Monte Carlo-based methods in [29, 30] 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 [29] also applies to conditional sampling procedures, a general procedure for generating scenario trees of such time

series driven stochastic processes is developed in [30] by approximating each of the (independent) uniform random variables using Quasi Monte Carlo methods (see [28]). 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" [28, p. 7]. The approach of (v) is based on probability distances that are relevant for the stability of multistage models. While the papers [14, 20, 31] employ Fortet-Mourier or Wasserstein distances, our recent work [16] is based on the rigorous stability result for linear multistage stochastic programs in [17]. 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 [7, 25].

In the present paper we extend the theoretical results obtained in [17] by proving an existence result for solutions of (1) and a (qualitative) stability result for solutions of multistage models. In addition, we review the forward technique of [16] for generating scenario trees. Its idea is to start with a fan of individual scenarios which represents a good initial approximation of the underlying stochastic input process ξ . This scenario fan might be obtained by sampling or resampling techniques based on parametric or nonparametric stochastic models of ξ . Starting from such a scenario fan, a tree is constructed recursively by scenario reduction [8, 15] and bundling (Algorithm 3.2). We review an error estimate for Algorithm 3.2 in terms of the L_r -distance and an estimate of the filtration distance D_f (Theorem 3.4). Based on the latter estimate we present a modification of Algorithm 3.2, which allows to control the tree generation by tolerance levels for the two relevant distances (L_r and D_f). The resulting Algorithm 3.5 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 [16] on generating inflow-demand scenario trees based on real-life data provided by the French company EdF. Earlier or modified versions of the algorithms were used in [36] to generate scenario trees in power engineering models and in [27] on generating passenger demand scenario trees in airline revenue management based on real-life data provided by Lufthansa Systems.

Section 2 presents an extension of the stability result of [17] which provides the basis of our tree constructions. Section 3 reviews some results of [16], in particular, the forward tree construction and error estimates in terms of L_r - and filtration distances, respectively. 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

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

The definition of r' is justified by the proof of [17, 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 ξ has to be required. This fact is well known for the two-stage situation (see, e.g., Chapter 2 in [35]). 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 ξ 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

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

is the set of feasible elements of (1) and

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

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

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

the nonanticipativity subspace of ξ 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

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

denote the α -level set of the stochastic program (3) with input $\tilde{\xi}$. Since, for $\alpha = 0$, the level set coincides with the set solutions to (3), we will also use the notation

$$S(\tilde{\xi}) := l_0(F(\tilde{\xi}, \cdot)).$$

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 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 exists a $\delta > 0$ and a bounded subset B of $L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ such that $l_{\alpha}(F(\tilde{\xi}, \cdot))$ 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., [34, 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 $\times_{t=1}^T L_{\hat{r}}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{n_t})$, $\frac{1}{\hat{r}} + \frac{1}{\hat{r}'} = 1$, of the dual stochastic program to (3), i.e., it holds that

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

where X_t^* denotes the polar to the polyhedral cone X_t , $t = 1, \dots, T$, and superscript $*$ at matrices means transposition.

(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 level set $l_{\alpha}(F(\xi, \cdot))$ is compact with respect to \mathcal{T} . Since 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 and generated by a basis consisting of the sets

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

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 Ω 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. For these and related results we refer to [11, Sections 3 and 4].

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

Proposition 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. [33] and [34, 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

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

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

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

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 [33, 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 α -level set $l_\alpha(F(\xi, \cdot))$ 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 α -level set $l_\alpha(F(\xi, \cdot))$ 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 α -level set with respect to $\sigma(L_1, L_\infty)$ follows from [33, Theorem 3K] due to condition (A5). For $r' = \infty$, some α -level set is bounded due to (A3) and, hence, relatively compact with respect to $\sigma(L_\infty, L_1)$ due to Alaoglu's theorem [11, 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 α -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 first result we introduce the functional $D_f(\xi, \tilde{\xi})$ depending on the filtrations of ξ and of its approximation (or perturbation) $\tilde{\xi}$, respectively. It is defined by

$$D_f(\xi, \tilde{\xi}) := \sup_{\varepsilon \in (0, \alpha]} D_{f, \varepsilon}(\xi, \tilde{\xi}) \quad (5)$$

and $D_{f, \varepsilon}(\xi, \tilde{\xi})$ denotes the ε -filtration distance given by

$$D_{f, \varepsilon}(\xi, \tilde{\xi}) := \inf \sum_{t=2}^{T-1} \max\{\|x_t - \mathbb{E}[x_t | \mathcal{F}_t(\tilde{\xi})]\|_{r'}, \|\tilde{x}_t - \mathbb{E}[\tilde{x}_t | \mathcal{F}_t(\xi)]\|_{r'}\}, \quad (6)$$

where the infimum is taken with respect to all $x \in l_\varepsilon(F(\xi, \cdot))$ and $\tilde{x} \in l_\varepsilon(F(\tilde{\xi}, \cdot))$, respectively, i.e., with respect to all feasible decisions belonging to the ε -level sets of the original and perturbed programs. 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 ξ and $\tilde{\xi}$ exist, the filtration distance is of the simplified form

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

If Ω is finite, the conditions (A1)–(A3) imply the existence of solutions to (3) for each input $\tilde{\xi}$ with $\|\xi - \tilde{\xi}\|_r \leq \delta$. For general Ω solutions to (3) exist, for example, if the assumptions of Proposition 2.1 are satisfied. 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 taken from [17, Theorem 2.1].

Theorem 2.2 *Let (A1), (A2) and (A3) be satisfied and $\mathcal{X}_1(\xi_1)$ be nonempty and bounded. Then there exists positive constants L and δ such that the estimate*

$$|v(\xi) - v(\tilde{\xi})| \leq L(\|\xi - \tilde{\xi}\|_r + D_f(\xi, \tilde{\xi})) \quad (7)$$

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 choice of r' and of $\mathcal{X}_1(\xi_1)$ is slightly more general than that considered in [17]. However, the proof of [17, Theorem 2.1] remains valid for the more general choice of r' and may be extended easily to constraints for x_1 that depend on ξ_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 solutions of (3), we need a slightly stronger version of the filtration distance D_f . For its definition we assume that solutions of (3) for the inputs ξ and $\tilde{\xi}$ exist (cf. Proposition 2.1). We define

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

Clearly, we have $D_f(\xi, \tilde{\xi}) \leq D_f^*(\xi, \tilde{\xi})$ and both functionals coincide if the solution sets $S(\xi)$ and $S(\tilde{\xi})$ are singletons.

Theorem 2.3 *Assume that only costs and right-hand sides are random in (3). Let (A1)–(A3) be satisfied for $r \in (1, \infty)$, $\frac{1}{r} + \frac{1}{r'} = 1$.*

If $(\xi^{(n)})$ is a sequence in $L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ converging to ξ 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 $x^{(n)} \in l_\alpha(F(\xi^{(n)}, \cdot))$ 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

$$\mathcal{X}(\xi) = \{x \in L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m) : x \in X, Ax = h(\xi)\}, \quad (8)$$

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., graph $\mathcal{X} = \{(x, \xi) \in L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m) \times L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s) \mid x \in \mathcal{X}(\xi)\}$ is closed and convex. Since $(\xi^{(n_k)})$ norm converges in $L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ to ξ and $(x^{(n_k)})$ weakly converges to x^* , the sequence $((x^{(n_k)}, \xi^{(n_k)}))$ of pairs in graph \mathcal{X} converges weakly to (x^*, ξ) . Due to the closedness and convexity of graph \mathcal{X} , Mazur's theorem [11, Theorem 3.19] implies that 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 have that

$$\begin{aligned} \left| \sum_{t=1}^T \langle y_t, x_t^* - \mathbb{E}[x_t^* | \mathcal{F}_t(\xi)] \rangle \right| &\leq \left| \sum_{t=1}^T \langle y_t, x_t^* - x_t^{(n_k)} \rangle \right| + \left| \sum_{t=1}^T \langle y_t, x_t^{(n_k)} - \mathbb{E}[x_t^{(n_k)} | \mathcal{F}_t(\xi)] \rangle \right| \\ &\quad + \left| \sum_{t=1}^T \langle y_t, \mathbb{E}[x_t^{(n_k)} | \mathcal{F}_t(\xi)] - \mathbb{E}[x_t^* | \mathcal{F}_t(\xi)] \rangle \right| \\ &\leq \left| \sum_{t=1}^T \langle y_t, x_t^* - x_t^{(n_k)} \rangle \right| + \|y\|_r D_f^*(\xi, \xi^{(n_k)}) \\ &\quad + \left| \sum_{t=1}^T \langle y_t, \mathbb{E}[x_t^{(n_k)} - x_t^* | \mathcal{F}_t(\xi)] \rangle \right|. \end{aligned}$$

The first term on the right-hand side converges to 0 for k tending to ∞ as the sequence $(x^{(n_k)})$ converges weakly to x^* . The third term converges to 0 as the mapping \mathcal{E}

$$\mathcal{E}(x) := (\mathbb{E}[x_1 | \mathcal{F}_1(\xi)], \dots, \mathbb{E}[x_T | \mathcal{F}_T(\xi)])$$

from $L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ into itself is linear and bounded and, hence, the sequence $(\mathcal{E}(x^{(n_k)}))$ converges weakly to $\mathcal{E}(x^*)$. Since $(D_f^*(\xi, \xi^{(n_k)}))$ also converges to 0, we conclude $x_t^* = \mathbb{E}[x_t^* | \mathcal{F}_t(\xi)]$ for each $t = 1, \dots, T$, and, hence, $x^* \in \mathcal{N}_{r'}(\xi)$. \square

Both Theorems 2.2 and 2.3 are valid for any choice of the underlying probability space such that there exists a version of ξ having its probability distribution. The right-hand side of (7) is minimal if the probability space is selected such that the distances $\|\xi - \tilde{\xi}\|_r$, $\|x_t - \mathbb{E}[x_t | \mathcal{F}_t(\tilde{\xi})]\|_{r'}$ and $\|\tilde{x}_t - \mathbb{E}[\tilde{x}_t | \mathcal{F}_t]\|_{r'}$ coincide with the corresponding L_r -minimal and $L_{r'}$ -minimal distances of the probability distributions of the corresponding random vectors. The L_r -minimal distance ℓ_r for $r \geq 1$ of probability distributions having r -th order absolute moments is given by

$$\ell_r(P, Q) := \inf \left\{ \int_{\Xi \times \Xi} |\xi - \tilde{\xi}|^r \eta(d\xi, d\tilde{\xi}) \mid \eta \in \mathcal{P}(\Xi \times \Xi), \pi_1 \eta = P, \pi_2 \eta = Q \right\}^{\frac{1}{r}} \quad (9)$$

(e.g., [32]). However, for deriving estimates of the filtration distance, specific probability spaces may be more appropriate (see Section 3).

3 Constructing scenario trees

Let ξ be the original stochastic process on some 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 ξ^{tr}

such that

$$\|\xi - \xi^{\text{tr}}\|_r \quad \text{and} \quad D_f(\xi, \xi^{\text{tr}}) \quad (10)$$

are small and, hence, the optimal values $v(\xi)$ and $v(\xi^{\text{tr}})$ are close to each other according to Theorem 2.2. Since this problem is hardly solvable in general, we replace ξ by a finitely discrete approximation ξ^f such that $\|\xi - \xi^f\|_r$ is small and its scenarios form a *fan* of individual scenarios. Let its scenarios and probabilities be denoted by $\xi^i =$

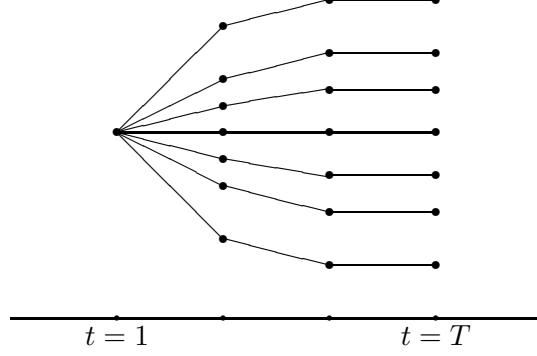


Figure 1: Example of a fan of individual scenarios with $T = 4$ and $N = 7$

$(\xi_1^i, \dots, \xi_T^i)$ and p_i , $i = 1, \dots, N$, respectively, and assume that all scenarios coincide at the first time period $t = 1$, i.e., $\xi_1^1 = \dots = \xi_1^N =: \xi_1^*$. Such a fan of individual scenarios may be obtained in many practical situations, e.g., by resampling techniques from historical data, by sampling from statistical models (e.g., time series models) or by quantization of the probability distribution of ξ [13]. Clearly, a fan may be regarded as a scenario tree with root node at $t = 1$ having N branches at the root and consisting of $1 + (T - 1)N$ nodes.

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

In this section we describe an algorithmic procedure that starts from the fan ξ^f and ends up with scenario trees ξ^{tr} having the same root node ξ_1^* , less nodes than ξ^f and allowing for constructive estimates of the L_r -norm $\|\xi^f - \xi^{\text{tr}}\|_r$ and of the corresponding filtration distance. Here, $r \geq 1$ is determined such that the optimal values of the underlying multistage stochastic program satisfy an estimate of the form (7) in Theorem 2.2. The idea of the algorithm consists in forming clusters of scenarios based on scenario reduction on the time horizon $\{1, \dots, t\}$ recursively for decreasing and increasing time t , respectively.

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

$$\|\xi\|_{r,t} := \left(\mathbb{E}[|\xi|_t^r] \right)^{\frac{1}{r}} = \left(\sum_{i=1}^N p_i |\xi^i|_t^r \right)^{\frac{1}{r}} \quad (11)$$

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

3.1 Forward tree construction

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 partitions $\mathcal{C}_t = \{C_t^1, \dots, C_t^{K_t}\}$ of I , i.e., such that

$$C_t^k \cap C_t^{k'} = \emptyset \quad \forall k \neq k' \quad \text{and} \quad \bigcup_{k=1}^{K_t} C_t^k = I. \quad (12)$$

The elements of such a partition \mathcal{C}_t are called (scenario) clusters. The initialization of the procedure consists in setting $\hat{\xi}^1 = \xi$, i.e., $\hat{\xi}^{1,i} = \xi^i$, $i \in I$, and $\mathcal{C}_1 = \{I\}$. At step t (with $t > 1$) 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 [15] such that

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

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 [15] the index set J_t^k is subdivided into index sets $J_{t,i}^k$, $i \in I_t^k$ such that

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

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

$$\alpha_t(j) = \begin{cases} i_t^k(j) & , j \in J_t^k, k = 1, \dots, K_{t-1}, \\ j & , \text{otherwise.} \end{cases} \quad (13)$$

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

$$\hat{\xi}_\tau^{t,i} = \begin{cases} \xi_\tau^{\alpha_t(i)} & , \tau \leq t, \\ \xi_\tau^i & , \text{otherwise,} \end{cases} \quad (14)$$

with probabilities p_i for each $i \in I$. The processes $\hat{\xi}^t$ are illustrated in Figure 2, where $\hat{\xi}^t$ corresponds to the t -th picture for $t = 1, \dots, T$. The partition \mathcal{C}_t at time t is defined by

$$\mathcal{C}_t = \{\alpha_t^{-1}(i) : i \in I_t^k, k = 1, \dots, K_{t-1}\}, \quad (15)$$

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

The 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

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

for each $k = 1, \dots, K_T$. The index set I_t of realizations of ξ_t^{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 a unique index $k_t(i) \in \{1, \dots, K_t\}$ such that $i \in C_t^{k_t(i)}$. Moreover, we have $C_t^{k_t(i)} = \{i\} \cup J_{t,i}^{k_t-1(i)}$ for each $i \in I_t$. The probability of the i -th realization of ξ_t^{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 [16, Theorem 3.4].

Theorem 3.1 *Let the stochastic process ξ with fixed initial node ξ_1^* , scenarios ξ^i and probabilities p_i , $i = 1, \dots, N$, be given. Let ξ^{tr} be the stochastic process with scenarios $\hat{\xi}^k = (\xi_1^*, \xi_2^{\alpha_2(i)}, \dots, \xi_t^{\alpha_t(i)}, \dots, \xi_T^{\alpha_T(i)})$ and probabilities π_T^k if $i \in C_T^k$, $k = 1, \dots, K_T$. Then we have*

$$\|\xi^f - \xi^{\text{tr}}\|_r = \left(\sum_{t=2}^T \sum_{k=1}^{K_{t-1}} \sum_{j \in J_t^k} p_j \min_{i \in I_t^k} |\xi_t^i - \xi_t^j|^r \right)^{\frac{1}{r}}. \quad (16)$$

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 ξ^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 ε , ε_t , $t = 1, \dots, T$, be given such that $\sum_{t=1}^T \varepsilon_t \leq \varepsilon$.

Step 1: Set $\hat{\xi}^1 := \xi$ 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 (13) and a stochastic process $\hat{\xi}^t$ having N scenarios $\hat{\xi}^{t,i}$ with probabilities p_i according to (14) and such that $\|\hat{\xi}^t - \hat{\xi}^{t-1}\|_{r,t} \leq \varepsilon_t$. Set $\mathcal{C}_t = \{\alpha_t^{-1}(i) : i \in I_t^k, k = 1, \dots, K_{t-1}\}$.

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

While the first picture in Figure 2 illustrates the original fan ξ^f , the second, third, fourth and fifth pictures correspond to the situation after Steps 2–5 of the algorithm. The final picture corresponds to Step 6 and illustrates the final scenario tree ξ^{tr} . The proof of the following corollary is also given in [16].

Corollary 3.3 *Let a scenario fan ξ^f with fixed initial node ξ_1^* , scenarios ξ^i and probabilities p_i , $i = 1, \dots, N$, be given. If ξ^{tr} is constructed by Algorithm 3.2 we have*

$$\|\xi^f - \xi^{\text{tr}}\|_r \leq \sum_{t=1}^T \varepsilon_t \leq \varepsilon.$$

When using Algorithm 3.2, the selection of $r > 1$ should be done according to Section 2. The tolerances ε_t should be nonincreasing for ε_t , $t = 2, \dots, T$. The smaller ε_t is, the more branchings occur at t .

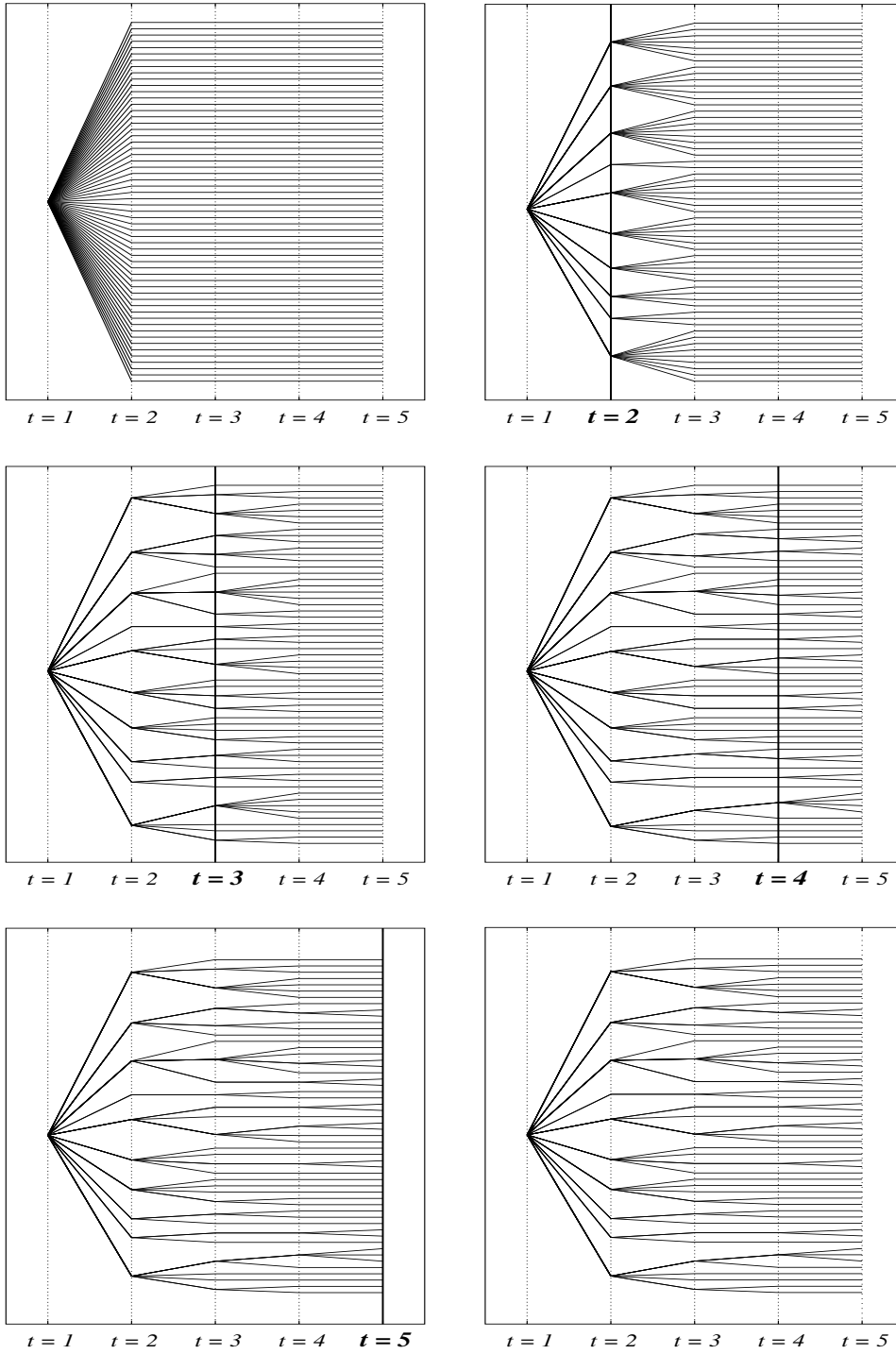


Figure 2: Illustration of the tree construction for an example with $T=5$ time periods

3.2 Estimating filtration distances

Let ξ^f be the (discrete) approximation of the original stochastic process ξ consisting of individual scenarios and ξ^{tr} be the process obtained by the forward tree construction approach of Section 3.1, respectively. Theorem 3.1 provides an estimate for $\|\xi^f - \xi^{\text{tr}}\|_r$. Of course, our aim is to generate the scenario tree ξ^{tr} such that the optimal values $v(\xi)$ and $v(\xi^{\text{tr}})$ and (eventually) the sets $S(\xi)$ and $S(\xi^{\text{tr}})$ of solutions are close. Making use of the stability estimate of Theorem 2.2, for example, we obtain

$$\begin{aligned} |v(\xi) - v(\xi^{\text{tr}})| &\leq L(\|\xi - \xi^{\text{tr}}\|_r + D_f(\xi, \xi^{\text{tr}})) \\ &\leq L(\|\xi - \xi^f\|_r + \|\xi^f - \xi^{\text{tr}}\|_r + D_f(\xi, \xi^{\text{tr}})). \end{aligned}$$

Since the fan ξ^f is assumed to be selected such that $\|\xi - \xi^f\|_r$ is small, it remains to estimate the filtration distance $D_f(\xi, \xi^{\text{tr}})$. The latter aim, however, is very challenging and hardly possible in general. Therefore, we replace ξ by ξ^f , derive bounds for

$$D_f(\xi^f, \xi^{\text{tr}})$$

and develop strategies for controlling the tree generation process by exploiting the latter bounds together with the bound in Corollary 3.3. Although this step induces an error, it is heuristically justified in our opinion as ξ^f approximates ξ in the L_r -sense and, hence, carries approximate information on the filtration of ξ . We stress again that the filtration distance is evaluated at solutions of (3) (if they exist) with inputs ξ^f and ξ^{tr} , respectively.

Now, let ξ^f and ξ^{tr} be defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is defined by $\Omega = \{\omega_1, \dots, \omega_N\}$, \mathcal{F} denotes the power set of Ω and $\mathbb{P}(\omega_i) = p_i$, $i = 1, \dots, N$. We assume that conditions (A2) and (A3) of Section 2 are satisfied and know that solutions $x \in S(\xi^f)$ and $\tilde{x} \in S(\xi^{\text{tr}})$ exist. Since the filtration $(\mathcal{F}_t)_{t=1}^T$ of ξ^f satisfies $\mathcal{F}_t = \mathcal{F}$ for $t = 2, \dots, T$, we obtain the following estimate

$$D_f(\xi^f, \xi^{\text{tr}}) \leq \begin{cases} \sum_{t=2}^{T-1} (\mathbb{E}[|x_t - \mathbb{E}[x_t | \mathcal{F}_t(\xi^{\text{tr}})]|^{r'}])^{\frac{1}{r'}} & , 1 \leq r' < \infty \\ \sum_{t=2}^{T-1} \|x_t - \mathbb{E}[x_t | \mathcal{F}_t(\xi^{\text{tr}})]\|_{\infty} & , r' = \infty \end{cases} \quad (17)$$

for the filtration distance of ξ^f and ξ^{tr} , respectively.

As in Section 3.1 we denote by I_t again the index set of realizations of the scenario tree ξ^{tr} at time t , by $C_t^{k_t(i)}$ the cluster $C_t^{k_t(i)} = \{i\} \cup J_{t,i}^{k_t-1(i)}$ for each $i \in I_t$ and by π_t^i the cluster probability for each $i \in I_t$. Then we obtain

$$D_f(\xi^f, \xi^{\text{tr}}) \leq \begin{cases} \sum_{t=2}^{T-1} \left(\sum_{i \in I_t} \sum_{j \in C_t^{k_t(i)}} p_j \left| x_t^j - \frac{1}{\pi_t^i} \sum_{l \in C_t^{k_t(i)}} p_l x_t^l \right|^{r'} \right)^{\frac{1}{r'}} & (1 \leq r' < \infty) \\ \sum_{t=2}^{T-1} \max_{i \in I_t} \max_{j \in C_t^{k_t(i)}} \left| x_t^j - \frac{1}{\pi_t^i} \sum_{l \in C_t^{k_t(i)}} p_l x_t^l \right| & (r' = \infty), \end{cases} \quad (18)$$

where x_t^i , $i = 1, \dots, N$, are the t -th components of solution scenarios of the two-stage model with input ξ^f having scenarios ξ^j , $j = 1, \dots, N$.

Starting from (18) we are interested in bounds on D_f that are based on input information only. This requires to derive estimates for the distance of any two scenarios of some solution $x \in S(\xi^f)$. To this end, we assume that only costs and right-hand sides are random in (3) and consider the scenario-based stochastic program

$$\min \left\{ \langle b_1(\xi_1^*), x_1 \rangle + \sum_{i=1}^N p_i \sum_{t=2}^T \langle b_t(\xi_t^i), x_t^i \rangle \left| \begin{array}{l} x_t^i \in X_t, t = 1, \dots, T, A_{1,0}x_1 = h_1(\xi_1), \\ A_{t,0}x_t^i + A_{t,1}x_{t-1}^i = h_t(\xi_t^i), \\ t = 2, \dots, T, i = 1, \dots, N \end{array} \right. \right\}, \quad (19)$$

which is indeed two-stage and represents a linear program. It is well known that the minimization decomposes into first- and second-stage variables leading to the following form of the two-stage program (19)

$$\min \left\{ \langle b_1(\xi_1^*), x_1 \rangle + \sum_{i=1}^N p_i \Phi(\xi^i, x_1) \mid x_1 \in X_1 \right\}, \quad (20)$$

where the optimal value function Φ is extended real-valued and defined on $\Xi \times \mathcal{X}_1(\xi_1^*)$ by

$$\Phi(z, x_1) := \inf \left\{ \sum_{t=2}^T \langle b_t(z_t), x_t \rangle \mid x_t \in X_t, A_{t,0}x_t + A_{t,1}x_{t-1} = h_t(z_t), t = 2, \dots, T \right\} \quad (21)$$

for any pair $(z, x_1) \in \Xi \times \mathcal{X}_1(\xi_1^*)$. Exploiting Lipschitz stability properties of solutions to the linear program on the right-hand side of (21), the following result is proved as Theorem 4.8 in [16].

Theorem 3.4 *Assume that only costs and right-hand sides are random in (3) and that (A2) and (A3) are satisfied. Then there exists a constant $\hat{L} > 0$ such that the filtration distance allows the estimate*

$$D_f(\xi^f, \xi^{\text{tr}}) \leq \hat{L} \begin{cases} \left(\sum_{i \in I_2} \sum_{j \in C_2^{k_2(i)}} p_j |\xi^j - \xi^i|^{r'} \right)^{\frac{1}{r'}}, & 1 \leq r' < \infty \\ \max_{i \in I_2} \max_{j \in C_2^{k_2(i)}} |\xi^j - \xi^i|, & r' = \infty. \end{cases} \quad (22)$$

The estimate (22) shows that the first time period after the deterministic first stage plays a major role. The estimate advises that every cluster $C_2^{k_2(i)}$ has to be chosen such that the term

$$B_{2,i} := \begin{cases} \sum_{j \in C_2^{k_2(i)}} p_j |\xi^j - \xi^i|^{r'} & , 1 \leq r' < \infty \\ \max_{j \in C_2^{k_2(i)}} |\xi^j - \xi^i| & , r' = \infty \end{cases} \quad (23)$$

is small enough. This means that scenario $i \in I_2$ should admit branching at $t = 2$ only if the distance of ξ^i and each scenario ξ^j , $j \in C_2^{k_2(i)}$ that branches from i is not too large, i.e., the bound $B_{2,i}$ defined in (23) is small. In many practical cases the latter condition will imply that the cardinality of $C_2^{k_2(i)}$ remains relatively small and that of I_2 large. Notice that the distance of scenarios in (22) is measured with respect to the

whole time horizon. The latter fact represents the main difference to the estimates of the L_r -distance in Theorem 3.1.

Accordingly, Algorithm 3.2 has to be modified such that clusters $C_2^{k_2(i)}$, $i \in I_2$, are determined at step $t = 2$ satisfying the condition

$$B_{r'}^* := \begin{cases} \left(\sum_{i \in I_2} B_{2,i} \right)^{\frac{1}{r'}} \leq \varepsilon_f & , 1 \leq r' < \infty \\ \max_{i \in I_2} B_{2,i} \leq \varepsilon_f & , r' = \infty \end{cases} \quad (24)$$

for some filtration tolerance ε_f . If (24) is satisfied, the further branching behavior at time periods t with $2 < t \leq T$ is controlled via the existing tests in Algorithm 3.2. Hence, the modified algorithm is of the form:

Algorithm 3.5 (modified forward tree construction)

Let N scenarios ξ^i with probabilities p_i , $i = 1, \dots, N$, root $\xi_1^* \in \mathbb{R}^d$, $r \geq 1$ and filtration level ε_f , reduction level ε_r and ε_t , $t = 1, \dots, T$, be given such that $\sum_{t=1}^T \varepsilon_t \leq \varepsilon_r$.

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

Step 2: Determine disjoint index sets I_2^1 and J_2^1 such that $I_2^1 \cup J_2^1 = I$, the mapping $\alpha_2(\cdot)$ according to (13) and a stochastic process $\hat{\xi}^2$ having N scenarios $\hat{\xi}^{2,i}$ with probabilities p_i according to (14) such that $B_{r'}^* \leq \varepsilon_f$ (see (24)) and $\|\hat{\xi}^2 - \hat{\xi}^1\|_{r,2} \leq \varepsilon_2$. Set $\mathcal{C}_2 = \{\alpha_2^{-1}(i) : i \in I_2^1\}$.

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 (13) and a stochastic process $\hat{\xi}^t$ having N scenarios $\hat{\xi}^{t,i}$ with probabilities p_i according to (14) and such that $\|\hat{\xi}^t - \hat{\xi}^{t-1}\|_{r,t} \leq \varepsilon_t$. Set $\mathcal{C}_t = \{\alpha_t^{-1}(i) : i \in I_t^k, k = 1, \dots, K_{t-1}\}$.

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

Some numerical experience for Algorithm 3.5 applied to certain practical models in power engineering is reported in [16] and in the next section.

4 Numerical experience

We consider a mean-risk optimization model for electricity portfolios of a German municipal power 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. [10]). 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 3.5 started with a finite number of (individual) 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-

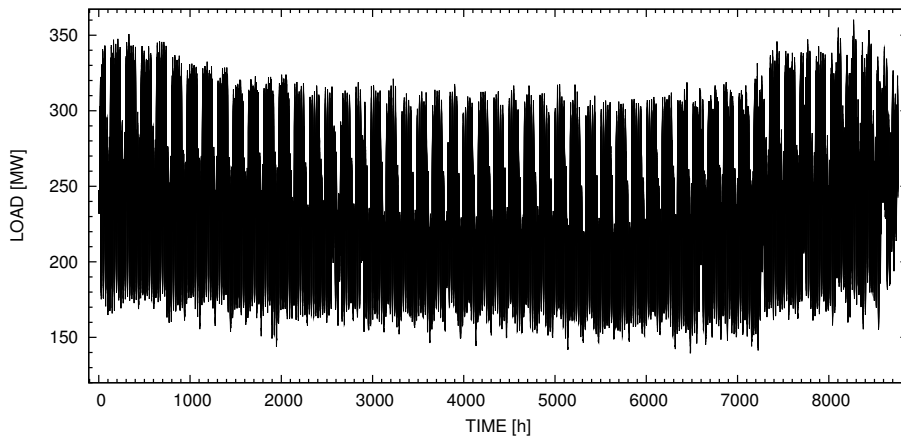


Figure 3: Time plot of load profile for one year

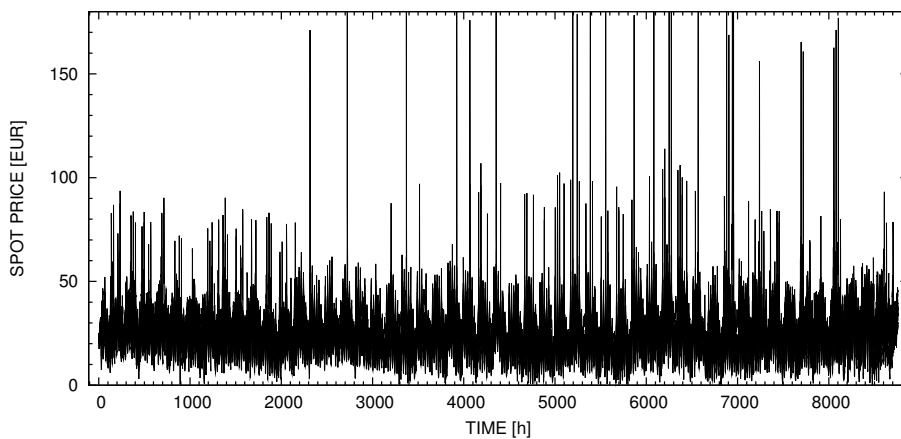


Figure 4: Time plot of spot price profile for one year

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 [10] 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 a scenario fan and serve as inputs for the modified forward tree construction (Algorithm 3.5). 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 3.5 for $r = r' = 2$ and different relative reduction and filtration levels $\varepsilon_{\text{rel},r}$ and $\varepsilon_{\text{rel},f}$, respectively. The relative levels are given by

$$\varepsilon_{\text{rel},r} := \frac{\varepsilon_r}{\varepsilon_{\text{max}}^r} \quad \text{and} \quad \varepsilon_{\text{rel},f} := \frac{\varepsilon_f}{\varepsilon_{\text{max}}^{r'}},$$

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

$$\varepsilon_t = \frac{\varepsilon}{T} \left[1 + \bar{q} \left(\frac{1}{2} - \frac{t}{T} \right) \right], \quad t = 2, \dots, T, \quad (25)$$

where $\bar{q} \in [0, 1]$ is a parameter that affects the branching structure of the constructed trees. For the test runs we used $\bar{q} = 0.6$. 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 and filtration levels. For varying reduction levels two series of scenario trees were computed where greater or smaller importance was attached to the filtration distance. As expected, the results show that a higher relative filtration level, i.e. more tolerance between the scenario fan and the scenario tree with respect to the filtration distance, leads to more branching regardless which reduction level is selected whenever the reduction level is not too small. On the other hand, a smaller relative filtration level

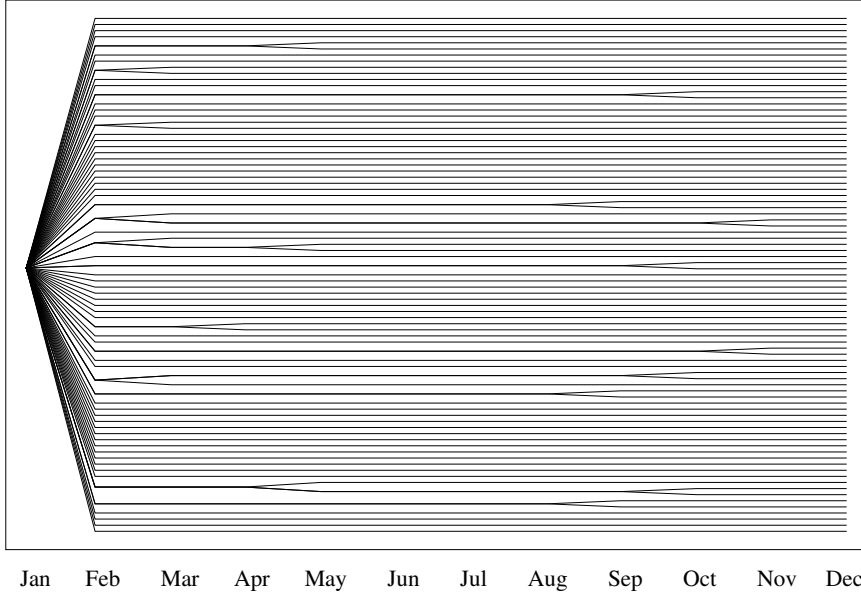
leads to less branching. Nevertheless, the number of nodes for some reduction level does not differ much when varying the filtration level. Clearly, for very small reduction levels, the reduction affects only a few scenarios. Furthermore, the number of nodes decreases considerably if the reduction level 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. Figures 5 and 6 show the scenario trees obtained for reduction levels of 25 percent and 50 percent, respectively.

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

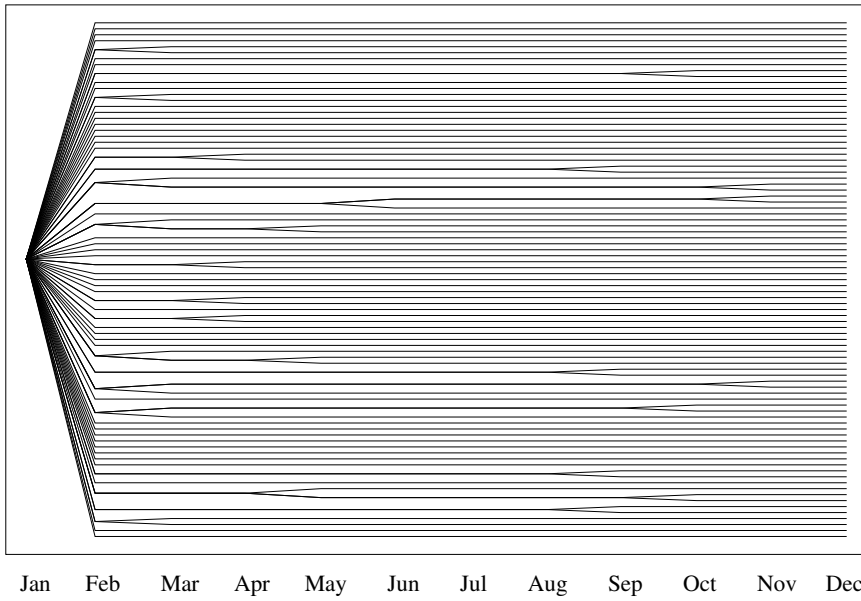
Table 2: Numerical results of Algorithm 3.5 for yearly demand-price scenario trees

References

- [1] Casey, M., Sen, S.: The scenario generation algorithm for multistage stochastic linear programming, *Mathematics of Operations Research* 30 (2005), 615–631.
- [2] Chiralaksanakul, A., Morton, D.P.: Assessing policy quality in multi-stage stochastic programming, *Stochastic Programming E-Print Series* 12–2004 (<www.speps.org>).
- [3] Corvera Poiré, X.: *Model Generation and Sampling Algorithms for Dynamic Stochastic Programming*, PhD Thesis, Department of Mathematics, University of Essex, 1995.

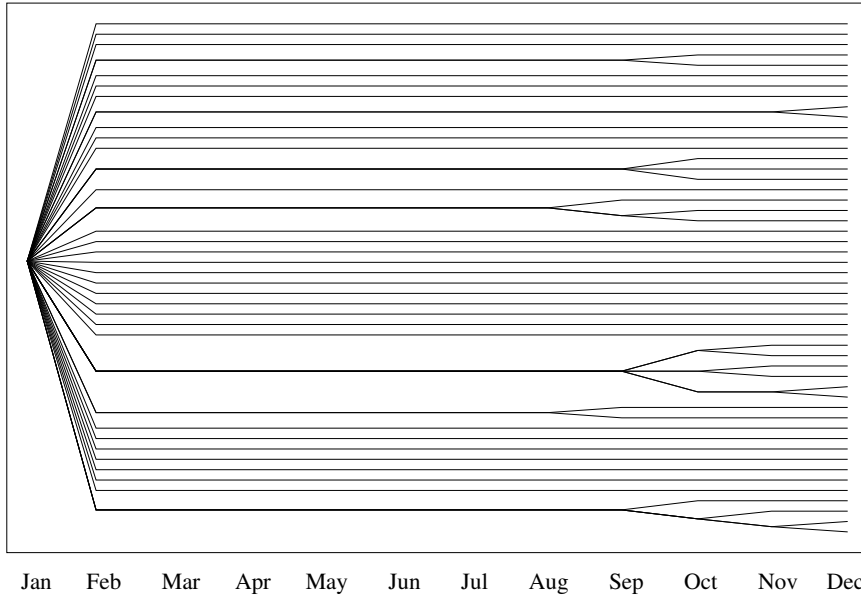


a) Modified forward tree construction with high filtration level $\varepsilon_{\text{rel},f}$

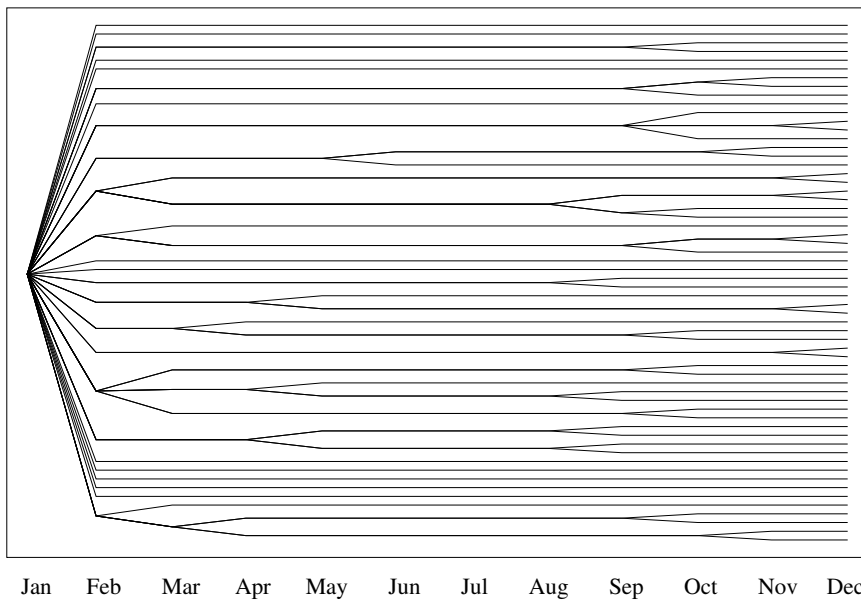


b) Modified forward tree construction with small filtration level $\varepsilon_{\text{rel},f}$

Figure 5: Yearly demand-price scenario trees with reduction level $\varepsilon_{\text{rel},r} = 0.25$



a) Modified forward tree construction with high filtration level $\varepsilon_{\text{rel},f}$



b) Modified forward tree construction with small filtration level $\varepsilon_{\text{rel},f}$

Figure 6: Yearly demand-price scenario trees with reduction level $\varepsilon_{\text{rel},r} = 0.50$

- [4] Dantzig, G. B.: Linear programming under uncertainty, *Management Science* 1 (1955), 197–206.
- [5] Dantzig, G. B.: *Linear Programming and Extensions*, Princeton University Press, Princeton 1963.
- [6] Dempster, M.A.H.: Sequential importance sampling algorithms for dynamic stochastic programming, *Zap. Nauchn. Semin. POMI* 312 (2004), 94–129.
- [7] Dupačová, J.; Consigli, G.; Wallace, S. W.: Scenarios for multistage stochastic programs, *Annals of Operations Research* 100 (2000), 25–53.
- [8] Dupačová, J.; Gröwe-Kuska, N.; Römisch, W.: Scenario reduction in stochastic programming: An approach using probability metrics, *Mathematical Programming Ser. A* 95 (2003), 493–511.
- [9] Edirisinghe, N.C.P.: Bound-based approximations in multistage stochastic programming: Nonanticipativity aggregation, *Annals of Operations Research* 85 (1999), 103–127.
- [10] Eichhorn, A., Römisch, W., Wegner, I.: Mean-risk optimization of electricity portfolios using multiperiod polyhedral risk measures, *IEEE St. Petersburg Power Tech* 2005.
- [11] 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.
- [12] Frauendorfer, K.: Barycentric scenario trees in convex multistage stochastic programming, *Mathematical Programming Ser. B*, 75 (1996), 277–293.
- [13] Graf, S.; Luschgy, H.: *Foundations of quantization for probability distributions*, Lecture Notes in Mathematics, Vol. 1730, Springer, Berlin, 2000.
- [14] 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.
- [15] Heitsch, H.; Römisch, W.: Scenario reduction algorithms in stochastic programming, *Computational Optimization and Applications* 24 (2003), 187–206.
- [16] Heitsch, H.; Römisch, W.: Scenario tree modelling for multistage stochastic programs, Preprint 296, DFG Research Center MATHEON "Mathematics for key technologies", 2005 and submitted.
- [17] Heitsch, H.; Römisch, W.; Strugarek, C.: Stability of multistage stochastic programs, Preprint 255, DFG Research Center MATHEON "Mathematics for key technologies", 2005 and *SIAM Journal on Optimization* (to appear).
- [18] Higle, J.L.; Rayco, B.; Sen, S.: Stochastic scenario decomposition for multistage stochastic programs, submitted to *Annals of Operations Research*.
- [19] Hochreiter, R.: Computational Optimal Management Decisions – The case of Stochastic Programming for Financial Management, Ph.D. Thesis, University of Vienna, 2005.

- [20] Hochreiter, R., Pflug, G.: Financial scenario generation for stochastic multi-stage decision processes as facility location problem, *Annals of Operations Research* (to appear).
- [21] Høyland, K.; Wallace, S. W.: Generating scenario trees for multi-stage decision problems, *Management Science* **47** (2001), 295–307.
- [22] Høyland, K., Kaut, M., Wallace, S.W.: A heuristic for moment-matching scenario generation, *Computational Optimization and Applications* **24** (2003), 169–185.
- [23] Infanger, G.: *Planning under Uncertainty – Solving Large-Scale Stochastic Linear Programs*, Boyd& Fraser Publ. Comp., 1994.
- [24] Kall, P.; Mayer, J.: *Stochastic Linear Programming*, Springer, 2005.
- [25] Kaut, M., Wallace, S.W.: Evaluation of scenario-generation methods for stochastic programming, *Stochastic Programming E-Print Series* 14-2003 (<www.speps.org>).
- [26] Kuhn, D.: *Generalized bounds for convex multistage stochastic programs*, Lecture Notes in Economics and Mathematical Systems, Vol. 548, Springer, Berlin, 2005.
- [27] Möller, A., Römisch, W., Weber, K.: A new approach to O&D revenue management based on scenario trees, *Journal of Revenue and Pricing Management* **3** (2004), 265–276.
- [28] Niederreiter, H.: *Random Number Generation and Quasi-Monte Carlo Methods*, CMBS-NSF Regional Conference Series in Applied Mathematics, Vol. 63, SIAM, Philadelphia 1992.
- [29] Pennanen, T.: Epi-convergent discretizations of multistage stochastic programs, *Mathematics of Operations Research* **30** (2005), 245–256.
- [30] Pennanen, T.: Epi-convergent discretizations of multistage stochastic programs via integration quadratures, *Stochastic Programming E-Print Series* 19-2004 (<www.speps.org>) and *Mathematical Programming* Ser. B (to appear).
- [31] Pflug, G. Ch.: Scenario tree generation for multiperiod financial optimization by optimal discretization, *Mathematical Programming* **89** (2001), 251–271.
- [32] Rachev, S. T.; Rüschendorf, L.: *Mass Transportation Problems*, Vol. I and II, Springer, Berlin 1998.
- [33] 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.
- [34] Rockafellar, R. T.; Wets, R. J-B: *Variational Analysis*, Springer-Verlag, Berlin, 1998.
- [35] Ruszczyński, A.; Shapiro, A. (Eds.): *Stochastic Programming*, Handbooks in Operations Research and Management Science, Volume 10, Elsevier, Amsterdam 2003.
- [36] Schmöller, H.: Modellierung von Unsicherheiten bei der mittelfristigen Stromerzeugungs- und Handelsplanung, *Aachener Beiträge zur Energieversorgung*, Band 103, Aachen 2005.

- [37] Shapiro, A.: Inference of statistical bounds for multistage stochastic programming problems, *Math. Meth. Oper. Res.* 58 (2003), 57–68.
- [38] Shapiro, A.: On complexity of multistage stochastic programs, *Operations Research Letters* 34 (2006), 1–8.
- [39] Shapiro, A.: Stochastic programming approach to optimization under uncertainty, *Mathematical Programming* (to appear).
- [40] Wallace, S. W.; Ziemba, W. T. (Eds.): *Applications of Stochastic Programming*, MPS-SIAM Series in Optimization, 2005.