Scenario tree approximation and risk aversion strategies for stochastic optimization of electricity production and trading

Andreas Eichhorn, Holger Heitsch, and Werner Römisch

Humboldt-University Berlin, Department of Mathematics, 10099 Berlin, Germany,
{eichhorn,heitsch,romisch}@math.hu-berlin.de,
http://www.math.hu-berlin.de/~{eichhorn,heitsch,romisch}

Summary. Dynamic stochastic optimization techniques are highly relevant for applications in electricity production and trading since there are uncertainty factors at different time stages (e.g., demand, spot prices) that can be described reasonably by statistical models. In this paper, two aspects of this approach are highlighted: scenario tree approximation and risk aversion. The former is a procedure to replace a general statistical model (probability distribution), which makes the optimization problem intractable, suitably by a finite discrete distribution. Our methods rest upon suitable stability results for stochastic optimization problems. With regard to risk aversion we present the approach of polyhedral risk measures. For stochastic optimization problems minimizing risk measures from this class it has been shown that numerical tractability as well as stability results known for classical (non-risk-averse) stochastic programs remain valid. In particular, the same scenario approximation methods can be used. Finally, we present illustrative numerical results from an electricity portfolio optimization model for a municipal power utility.

Key words: Stochastic programming, optimization, scenario tree approximation, risk management, polyhedral risk functionals, multiperiod risk, energy trading, power portfolio, electricity

1 Introduction

The deregulation of energy markets has led to several new challenges for electric power utilities. Electric power has to be generated in a competitive environment and, in addition, coordinated with several trading activities. Electricity portfolios for spot and derivative markets become important, and the electrical load as well as electricity prices become increasingly unpredictable. Hence, the number of uncertainty sources and the financial risk for electric utilities have increased. These facts initiated the development of stochastic optimization models for producing and trading electricity. We mention, for example, stochastic hydro-electric and trading models [13, 32] and stochastic hydro-thermal production and trading models [12, 18, 19, 28, 37, 38, 39]. For an overview on stochastic programming models in energy we refer to [40].

Typical stochastic optimization models for producing and trading electricity, however, are focused on (expected) profit maximization while *risk management* is considered as an extra task. Power utilities often separate the planning of their hydro-thermal electricity production versus a preliminary and simplified trading model from the risk management. However, alternatively, risk management may be integrated into the (hydro-thermal) power production and trading planning by maximizing expected profit and minimizing (or bounding) a certain risk functional simultaneously [3, 9, 26]. Such *integrated risk management* strategies promise additional overall efficiency for power utilities.

Mathematical modeling of integrated risk management of an electricity producing and trading utility leads to multi-stage stochastic programs with risk objectives or risk constraints. In the present paper, we discuss two basic aspects of implementing such models: (i) the approximate representation of the underlying probability distribution by a finite discrete distribution, i.e., by a finite number of scenarios with their probabilities, and (ii) modeling and minimization of risk.

The first is typically an indispensable first step towards a solution of a stochastic optimization model. On the other hand, this is a highly sensitive concern, in particular, if dynamic decision structures are involved (multi-stage stochastic programming [36]). Then, the scenarios of the approximate distribution must exhibit *tree structure*. Moreover, it is of interest to get by with a moderate number of scenarios to have the resulting problem tractable. We refer to the overview [6] and to several different approaches [4, 5, 23, 20, 25, 27, 31] for scenario tree generation.

In section 4 we assume that scenarios of the underlying stochastic load-price process are available, e.g., by sampling from a properly developed stochastic (time series) model or by some other approximation scheme. We desribe a methodology based on clustering and scenario reduction that produces a tree of scenarios and represents a good approximation of the stochastic process. The approach is based on suitable stability results ensuring that the obtained approximate problems are indeed related to the original (infinite dimensional) ones. For interested readers these stability results are presented in section 3. The methodology as well as the stability arguments are based on distances of random vectors that allow to decide about their closeness. Moreover, since multi-stage stochastic programs look for decisions that do not anticipate, but depend at each time period t only on information that is available at t, a distance measure for the information flow is needed. It is expressed by a distance of filtrations, since the information increase over time is modeled by σ-fields forming a filtration that is associated to the stochastic process.

The second topic requires the selection of appropriate risk functionals that allow to quantify risk in a meaningful way and preserve tractability of the optimization model. We argue that *polyhedral risk functionals* satisfy both demands. These are given as (the optimal values of) certain simple linear stochastic programs. Well-known risk functionals such as Average Value-at-Risk AVaR and expected polyhedral utility belong to this class and, moreover, multi-period risk functionals for multi-stage stochastic programs are suggested. For stochastic programs incorporating polyhedral risk functionals it has been shown that numerical tractability as well as stability results known for classical (non-risk-averse) stochastic programs remain valid. In particular, the same scenario tree approximation methods can be used.

In a case study, we present illustrative numerical results from an electricity portfolio optimization model for a municipal power utility. In particular, it is shown that the use of different risk objectives leads to different risk aversion strategies by trading at derivative markets. They require less than additional 1% of the optimal expected revenue.

2 Mathematical framework

Let a finite number of time steps $T \in \mathbb{N}$ as well as a multivariate discrete-time *stochastic process* $\xi = (\xi_1, ..., \xi_T)$ be given. This means that each ξ_t is a d-dimensional random vector (with some fixed dimension $d \in \mathbb{N}$) whose realization can be observed at time step $t = 1, ..., T$, respectively. Since $t = 1$ represents the present we require that ξ_1 is deterministic, i.e., $\xi_1 \in \mathbb{R}^d$. For $t \geq 2$ we require that each ξ_t has *statistical moments* of oder r with some number $r \geq 1$ (that will be specified later on), i.e., $\mathbb{E}[|\xi_t|^r] < \infty$ for $t = 1, ..., T$ where $\mathbb{E}[\,.\,]$ denotes the expected value functional and $|\,.\,|$ refers to the Euclidean norm in $\mathbb{R}^d$.

Mathematically, these requirements are typically expressed by means of the so-called L_r-spaces: $\xi_t \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d)$ where $(\Omega, \mathcal{F}, \mathbb{P})$ is a given *probability space*. Now, in multi-stage stochastic programming, decisions x_t can be made at each time step $t = 1, ..., T$ based on the observations until time t, respectively. This means that x_t may depend and may only depend on (the concrete realization of) $\xi^t := (\xi_1, ..., \xi_t)$, respectively. This *nonanticipativity* requirement can be expressed by $x_t \in L_{r'}(\Omega, \sigma(\xi^t), \mathbb{P}; \mathbb{R}^{m_t})$ with some moment order $r' \geq 1$ (specified later on) and some dimensions $m_t \in \mathbb{N}$ $(t = 1, ..., T)$. In other words: x_t must be a $\sigma(\xi^t)$-measurable random element where $\sigma(\xi^t)$ is the sub-σ-field of the original σ-field $\mathcal{F}$ generated by $\xi_1, ..., \xi_t$. The sequence of all σ-fields is increasing, i.e., $\{\emptyset, \Omega\} = \sigma(\xi^1) \subseteq \sigma(\xi^2) \subseteq ... \subseteq \sigma(\xi^T) = \mathcal{F}$ and thus forms a so-called *filtration*. Assume for the moment that the input random vector ξ is represented in the form of a scenario tree, where d real variables are associated to each node of the tree. Then the $\sigma(\xi^t)$-measurability of x_t for every $t \in \{1, \ldots, T\}$ means that the decision vector x is represented by the same tree (as ξ), but with m_t real variables associated to each node at time t.

In this presentation, we consider *linear* multi-stage stochastic program of the form

$$\min_{x_1,\ldots,x_T} \left\{ \mathbb{E}\left[\sum_{\tau=1}^{T} \langle b_t(\xi_t), x_t\rangle\right] \left| \begin{array}{l} x_t \in L_{r'}(\Omega, \sigma(\xi^t), \mathbb{P}; \mathbb{R}^{m_t}), \\ x_t \in X_t \ \ \mathbb{P}\text{-almost surely (a.s.)}, \\ A_{t,0}x_t + A_{t,1}(\xi_t)x_{t-1} = h_t(\xi_t) \text{ a.s.} \\ (t = 1, ..., T) \end{array} \right. \right\} \tag{1}$$

with some numbers $m_t, n_t \in \mathbb{N}$, given polyhedral sets $X_t \subseteq \mathbb{R}^{m_t}$, recourse matrices $A_{t,0} \in \mathbb{R}^{n_t \times m_t}$, technology matrices $A_{t,1} \in \mathbb{R}^{n_t \times m_{t-1}}$ (where we assume $A_{1,1} \equiv 0$), and vectors $h_t \in \mathbb{R}^{n_t}$ and $b_t \in \mathbb{R}^{m_t}$ (cost factors). The items $A_{t,1}$, h_t, and b_t may depend on ξ_t $(t = 1, ..., T)$. It is assumed that this dependence is affinely linear. This allows, for example, to model that some components of b_t, h_t and/or some elements of the matrix $A_{t,1}$ are stochastic and ξ denotes the vector of all such stochastic inputs.

Note that in (1) optimality of the stochastic costs $\langle b_t(\xi_t), x_t\rangle$ is determined in terms of the expected value, i.e., the objective is linear (risk-neutral). In section 5 and 6 we will consider the risk-averse alternative

$$\min_{x_1,\ldots,x_T} \left\{ \begin{array}{l} \gamma \cdot \rho(z_{t_1}, ..., z_{t_J}) \\ -(1-\gamma) \cdot \mathbb{E}[z_T] \end{array} \left| \begin{array}{l} x_t \in L_{r'}(\Omega, \sigma(\xi^t), \mathbb{P}; \mathbb{R}^{m_t}), \\ x_t \in X_t \text{ a.s.}, \\ A_{t,0}x_t + A_{t,1}(\xi_t)x_{t-1} = h_t(\xi_t) \text{ a.s.} \\ z_t := -\sum_{\tau=1}^{t} \langle b_\tau(\xi_\tau), x_\tau\rangle \text{ a.s.} \\ (t = 1, ..., T) \end{array} \right. \right\} \tag{2}$$

where the objective is supplemented with a (multi-period) risk functional ρ (risk measure). The number $\gamma \in [0,1]$ is a fixed weighting parameter. The random values z_t represent the accumulated revenues at each time t. Clearly, it holds that $z_t \in L_p(\Omega, \sigma(\xi^t), \mathbb{P})$ with $p \in [1, \infty]$ given by

$$\frac{1}{p} = \begin{cases} \frac{1}{r'} & , \text{ if all } b_t \text{ are non-random} \\ \frac{1}{r} + \frac{1}{r'} & , \text{ otherwise.} \end{cases}$$

The risk functional ρ is applied to a subset of J time steps $1 < t_1 < t_2 < ... < t_J = T$. Note that, since risk functionals are essentially nonlinear by nature, problem (2) is no longer linear. However, we will concentrate on the employment of risk functionals from the class of *polyhedral risk functionals* which exhibit a favorable sort of nonlinearity; cf. section 5.

3 Stability of multi-stage problems

Studying stability of the multi-stage stochastic programs (1) consists in regarding it as an optimization problems in the infinite dimensional linear space $\times_{t=1}^{T} L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{m_t})$. This is a Banach space when endowed with the norm

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

where $|\,.\,|$ denotes some norm on the relevant Euclidean spaces and $\operatorname{ess\,sup}|x_t|$ denotes the essential supremum of $|x_t|$, i.e., the smallest constant C such that $|x_t| \le C$ holds $\mathbb{P}$-almost surely. Analogously, ξ can be understood as an element of the Banach space $\times_{t=1}^T L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d)$ with norm $\|\xi\|_r$. For the integrability numbers r and r' it will be imposed that

$$r := \begin{cases} \in [1,\infty) \ , & \text{if only costs or only right-hand sides are random} \\ 2 \ , & \text{if only costs and right-hand sides are random} \\ T \ , & \text{if all technology matrices are random} \end{cases}$$
$$r' := \begin{cases} \frac{r}{r-1} \ , & \text{if only costs are random} \\ r \ , & \text{if only right-hand sides are random} \\ \infty \ , & \text{if all technology matrices are random} \end{cases} \tag{3}$$

with regard to problem (1). The choice of r and the definition of r' are motivated by the knowledge of existing moments of the input process ξ, by having the stochastic program well defined (in particular, such that $\langle b_t(\xi_t), x_t\rangle$ is integrable for every decision x_t and $t = 1, ..., T$), and by satisfying the conditions (A2) and (A3) (see below).

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., [36, Chapter 2]). If either right-hand sides or costs in a multi-stage model (1) are random, it is sufficient to require $r \ge 1$. 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 $r \ge T$ to have the model well defined and no flexibility w.r.t. r' remains.

3.1 Assumptions

Next we introduce some notation. We set $s := Td$ and $m := \sum_{t=1}^T m_t$. Let

$$F(\xi, x) := \mathbb{E}\big[\sum_{t=1}^T \langle b_t(\xi_t), x_t\rangle\big]$$

denote the objective function defined on $L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s) \times L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ and let

$$\mathcal{X}(\xi) := \left\{x \in \times_{t=1}^T L_{r'}(\Omega, \sigma(\xi^t), \mathbb{P}; \mathbb{R}^{m_t}) \,|\, x_t \in \mathcal{X}_t(x_{t-1}; \xi_t) \text{ a.s.} \ \ (t = 1, ..., T)\right\}$$

denote the set of feasible elements of (1) with $x_0 \equiv 0$ and

$$\mathcal{X}_t(x_{t-1};\xi_t) := \big\{x_t \in \mathbb{R}^{m_t} : x_t \in X_t, A_{t,0}x_t + A_{t,1}(\xi_t)x_{t-1} = h_t(\xi_t)\big\}$$

denoting the t-th feasibility set for every $t = 1, ..., T$. That allows to rewrite the stochastic program (1) in the short form

$$\min \big\{F(\xi, x) : x \in \mathcal{X}(\xi)\big\} \tag{4}$$

In the following, we need the optimal value

$$v(\xi) = \inf \big\{F(\xi, x) : x \in \mathcal{X}(\xi)\big\}$$

for every $\xi \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ and, for any $\varepsilon \geq 0$, the *ε-approximate solution set* (level-set)

$$S_\varepsilon(\xi) := \big\{x \in \mathcal{X}(\xi) : F(\xi, x) \leq v(\xi) + \varepsilon\big\}$$

of the stochastic program (4). Since, for $\varepsilon = 0$, the set $S_\varepsilon(\xi)$ coincides with the set solutions to (4), we will also use the notation $S(\xi) := S_0(\xi)$. The following conditions will be imposed on (4):

(A1) The numbers r, r' are chosen according to (3) and $\xi \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$.

(A2) There exists a $\delta > 0$ such that for any $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ satisfying $\|\tilde{\xi} - \xi\|_r \leq \delta$, any $t = 2, ..., T$ and any $x_\tau \in L_{r'}(\Omega, \sigma(\tilde{\xi}^\tau), \mathbb{P}; \mathbb{R}^{m_\tau})$ ($\tau = 1, ..., t-1$) satisfying $x_\tau \in \mathcal{X}_\tau(x_{\tau-1}; \tilde{\xi}_\tau)$ a.s. (where $x_0 = 0$), there exists $x_t \in L_{r'}(\Omega, \sigma(\tilde{\xi}^t), \mathbb{P}; \mathbb{R}^{m_t})$ satisfying $x_t \in \mathcal{X}_t(x_{t-1}; \tilde{\xi}_t)$ a.s. (*relatively complete recourse locally around* ξ).

(A3) The optimal values $v(\tilde{\xi})$ of (4) 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 $\varepsilon_0 > 0$ there exists a $\delta > 0$ and a bounded subset B of $L_{r'}(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^m)$ such that $S_{\varepsilon_0}(\tilde{\xi})$ 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 $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ sufficiently close to ξ in L_r, condition (A2) implies the existence of some feasible $\tilde{x}$ in $\mathcal{X}(\tilde{\xi})$ and (3) implies the finiteness of the objective $F(\tilde{\xi}, .)$ 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., $A_{t,0}X_t = \mathbb{R}^{n_t}$, $t = 1, ..., T$. The locally uniform level-boundedness of the objective function F is quite standard in perturbation results for optimization problems (see, e.g., [35, 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.

3.2 Optimal values

The first stability result for multi-stage stochastic programs represents a quantitative continuity property of the optimal values. Its main observation is that multi-stage models behave stable at some stochastic input process if both its probability distribution and its filtration are approximated with respect to the L_r-distance and the filtration distance

$$D_{\mathrm{f}}(\xi,\tilde{\xi}) := \sup_{\varepsilon>0} \inf_{\substack{x\in S_\varepsilon(\xi)\\ \tilde{x}\in S_\varepsilon(\tilde{\xi})}} \sum_{t=2}^{T-1} \max\Big\{ \big\| x_t - \mathbb{E}[x_t|\sigma(\tilde{\xi}^t)] \big\|_{r'}, \big\| \tilde{x}_t - \mathbb{E}[\tilde{x}_t|\sigma(\xi^t)] \big\|_{r'} \Big\} \quad (5)$$

where $\mathbb{E}[\,.\,|\sigma(\xi^t)]$ and $\mathbb{E}[\,.\,|\sigma(\tilde{\xi}^t)]$ $(t = 1, ..., T)$ are the corresponding conditional expectations, respectively. Note that for the supremum in (5) only small ε's are relevant and that the approximate solution sets are bounded for $\varepsilon \in (0, \varepsilon_0]$ according to (A3).

The following stability result for optimal values of program (4) is taken from [24, Theorem 2.1].

Theorem 1. *Let (A1), (A2) and (A3) be satisfied and the sets X_1 be nonempty and bounded. Then there exist positive constants L and δ such that the estimate*

$$\big|v(\xi) - v(\tilde{\xi})\big| \le L\left(\|\xi - \tilde{\xi}\|_r + D_{\mathrm{f}}(\xi,\tilde{\xi})\right) \quad (6)$$

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

The result states that the changes of optimal values are at most proportional to the errors in terms of L_r- and filtration distance when approximating ξ. The corresponding constant L depends on $\|\xi\|_r$ (i.e. the r-th moment of ξ), but is not known in general.

3.3 Approximate Solutions

To prove a stability result for (approximate) solutions of (4) a stronger version of the filtration distance D_{f} is needed, namely,

$$D_{\mathrm{f}}^*(\xi,\tilde{\xi}) := \sup_{x\in\mathcal{B}_\infty} \sum_{t=2}^{T} \big\| \mathbb{E}[x_t|\sigma(\xi^t)] - \mathbb{E}[x_t|\sigma(\tilde{\xi}^t)] \big\|_{r'}, \quad (7)$$

where $\mathcal{B}_\infty := \{x : \Omega \to \mathbb{R}^m : x \text{ is } \mathcal{F}\text{-measurable}, |x(\omega)| \le 1, \mathbb{P}\text{-almost surely}\}$. 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. If we require, in addition to assumption (A3), that for some $\varepsilon_0 > 0$ there exist constants $\delta > 0$ and $C > 0$ such that $|\tilde{x}(\omega)| \le C$ for $\mathbb{P}$-almost every $\omega \in \Omega$ and all $\tilde{x} \in S_{\varepsilon_0}(\tilde{\xi})$ with $\tilde{\xi} \in L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ and $\|\tilde{\xi} - \xi\|_r \le \delta$, we have

$$D_{\mathrm{f}}(\xi,\tilde{\xi}) \le C\, D_{\mathrm{f}}^*(\xi,\tilde{\xi}). \quad (8)$$

Sometimes it is sufficient to consider the unit ball in $L_{r'}$ rather than $\mathcal{B}$ (cf. [23, 22]). However, in contrast to D_{f} the distance D_{f}^* always satisfies the triangle inequality.

Now, we state the second stability result that represents a Lipschitz property of approximate solution sets ([22, Theorem 2.4]).

Theorem 2. *Let (A1), (A2) and (A3) be satisfied with $r' \in [1,\infty)$ and the set X_1 be nonempty and bounded. Assume that the solution set $S(\xi)$ is nonempty. Then there exist $\bar{L} > 0$ and $\bar{\varepsilon} > 0$ such that*

$$\mathbb{d}_\infty\big(S_\varepsilon(\xi), S_\varepsilon(\tilde{\xi})\big) \leq \frac{\bar{L}}{\varepsilon}\Big(\|\xi - \tilde{\xi}\|_r + D_{\mathrm{f}}^*(\xi,\tilde{\xi})\Big) \tag{9}$$

holds for every $\tilde{\xi} \in L_r(\Omega,\mathcal{F},\mathbb{P};\mathbb{R}^s)$ with $\|\xi - \tilde{\xi}\|_r \leq \delta$ (with $\delta > 0$ from (A3)) and $S(\tilde{\xi}) \neq \emptyset$, and for any $\varepsilon \in (0,\bar{\varepsilon})$. Here, $\mathbb{d}_\infty$ denotes the Pompeiu-Hausdorff distance of closed bounded subsets of $L_{r'} = L_{r'}(\Omega.\mathcal{F},\mathbb{P};\mathbb{R}^m)$, which is given by

$$\mathbb{d}_\infty(B,\tilde{B}) = \sup_{x \in L_{r'}} \big|d_B(x) - d_{\tilde{B}}(x)\big|$$

with $d_B(x)$ denoting the distance of x to B, i.e., $d_B(x) = \inf_{y\in B}\|x-y\|_{r'}$.

The most restrictive assumption in Theorem 2 is the existence of solutions to both problems. Notice that solutions always exist if the underlying random vector has a finite number of scenarios or if $r' \in (1,\infty)$. For a more thorough discussion we refer to [22, Section 2]. Notice that the constant $\frac{\bar{L}}{\varepsilon}$ gets larger for decreasing ε and that, indeed, Theorem 2 does not remain true for the Pompeiu-Hausdorff distance of solution sets $S(\xi) = S_0(\xi)$ and $S(\tilde{\xi}) = S_0(\tilde{\xi})$, respectively.

4 Construction of scenario trees

In this section we want to introduce a general approach to generate appropriate scenario trees by making use of the stability theory of the previous section. To this end we assume that $r \geq 1$ and r' are selected such that ξ has a finite r-th moment and according to (3), respectively. Then we aim at generating a scenario tree ξ_{tr} such that the distances

$$\|\xi - \xi_{\mathrm{tr}}\|_r \quad \text{and} \quad D_{\mathrm{f}}^*(\xi,\xi_{\mathrm{tr}}) \tag{10}$$

are small, where the latter is given by (7). We conclude that the optimal values $v(\xi)$ and $v(\xi_{\mathrm{tr}})$, and the approximate solution sets $S_\varepsilon(\xi)$ and $S_\varepsilon(\xi_{\mathrm{tr}})$ are close to each other according to Theorem 1 and Theorem 2, respectively.

4.1 General Approach

The scenario tree construction method starts with a good initial scenario approximation consisting of a finite number of scenarios. These scenarios might be obtained by quantization techniques [16] or by sampling or resampling techniques based on parametric or nonparametric stochastic models of the input process ξ. Let us denote the initial approximation of ξ by $\hat{\xi}$ having scenarios

$\xi^i = (\xi^i_1, ..., \xi^i_T) \in \mathbb{R}^{Td}$ with probabilities $p_i > 0$, $i = 1, ..., N$, and a common root, i.e., $\xi^1_1 = ... = \xi^N_1 =: \xi^*_1$.

In the following we assume that

$$\|\xi - \hat{\xi}\|_r \ + \ D^*_{\mathrm{f}}(\xi, \hat{\xi}) \ \leq \ \varepsilon \tag{11}$$

holds for some given (initial) tolerance $\varepsilon > 0$. For example, condition (11) may be satisfied for D^*_{f} and any tolerance $\varepsilon > 0$ if $\hat{\xi}$ is obtained by sampling from a finite set with sufficiently large sample size (see [23, Example 5.3]). A more general case is discussed in [20], where the only assumption is that the initial set of scenarios provides a good approximation with respect to the L_r-distance.

Next we describe an algorithmic procedure that starts from $\hat{\xi}$ and ends up with a scenario tree process ξ_{tr} having the same root node ξ^*_1, less nodes than $\hat{\xi}$ and allowing for constructive estimates of $\|\hat{\xi} - \xi_{\mathrm{tr}}\|_r$. The idea of the algorithm consists in forming clusters of scenarios based on scenario reduction on the time horizon $\{1, ..., t\}$ recursively for increasing time t. To this end, the seminorm $\| \,.\, \|_{r,t}$ on $L_r(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^s)$ (with $s = Td$) given by

$$\|\xi\|_{r,t} := \left(\mathbb{E}\left[|\xi|^r_t\right]\right)^{1/r} \tag{12}$$

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

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

$$C^k_t \cap C^{k'}_t = \emptyset \quad (k \neq k') \quad \text{and} \quad \bigcup_{k=1}^{K_t} C^k_t = I. \tag{13}$$

The index sets $C^k_t \in \mathcal{C}_t$, $k = 1, ..., K_t$, represent clusters of scenarios (see Figure 1 for an illustration). To define these clusters we aim at aggregating similar scenarios at each time step.

The initialization of the scenario tree generation 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^k_{t-1}, i.e., each scenario subset $\{\hat{\xi}^{t-1,i}\}_{i \in C^k_{t-1}}$, separately and delete scenarios $\{\hat{\xi}^{t-1,j}\}_{j \in J^k_t}$ by the forward selection algorithm of [21] (see also [23, Section 2]) such that

$$\left(\sum_{k=1}^{K_{t-1}} \sum_{j \in J^k_t} p_j \min_{i \in I^k_t} \left|\hat{\xi}^{t-1,i} - \hat{\xi}^{t-1,j}\right|^r_t \right)^{1/r}$$

is bounded from above by some prescribed tolerance. Here, the index set I^k_t of remaining scenarios is given by

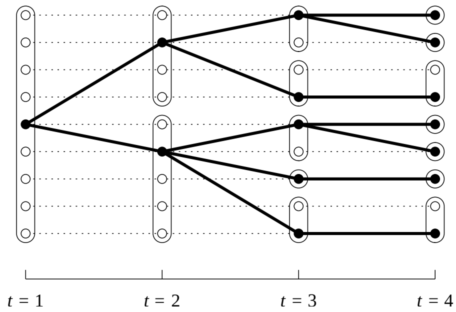

Fig. 1. Illustration of the tree construction by recursive scenario clustering

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

As in the general scenario reduction procedure in [21], the index set J_t^k is subdivided into index sets $J_{t,i}^k$, $i \in I_t^k$ such that

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

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

$$\alpha_t(j) = \begin{cases} i_t^k(j) & , j \in J_t^k,\, k = 1, ..., K_{t-1} \\ j & , \text{otherwise.} \end{cases} \tag{14}$$

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

$$\hat{\xi}_\tau^{t,i} = \begin{cases} \xi_\tau^{\alpha_\tau(i)} & , \tau \le t \\ \xi_\tau^i & , \text{otherwise} \end{cases} \tag{15}$$

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 $t = 1, ..., T$. The partition $\mathcal{C}_t$ at t is defined by

$$\mathcal{C}_t = \{\alpha_t^{-1}(i) : i \in I_t^k,\, k = 1, ..., K_{t-1}\}, \tag{16}$$

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

$$\xi_{\text{tr}}^k = \left(\xi_1^*, \xi_2^{\alpha_2(i)}, ..., \xi_t^{\alpha_t(i)}, ..., \xi_T^{\alpha_T(i)}\right) \quad \text{and} \quad \pi_T^k = \sum_{j \in C_T^k} p_j \quad \text{if } i \in C_T^k \tag{17}$$

for each $k = 1, ..., K_T$. 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 $t \in \{1, ..., T\}$ and each $i \in I$ there exists an unique index $k_t(i) \in \{1, ..., K_t\}$ 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 [23, Theorem 3.4].

Theorem 3. *Let the stochastic process $\hat{\xi}$ with fixed initial node ξ_1^*, scenarios ξ^i and probabilities p_i, $i = 1, ..., N$, be given. Let ξ_{tr} be the stochastic process with scenarios $\xi_{\mathrm{tr}}^k = (\xi_1^*, \xi_2^{\alpha_2(i)}, ..., \xi_t^{\alpha_t(i)}, ..., \xi_T^{\alpha_T(i)})$ and probabilities π_T^k for $i \in C_T^k$, $k = 1, ..., K_T$. Then we have*

$$\left\|\hat{\xi} - \xi_{\mathrm{tr}}\right\|_r \leq \sum_{t=2}^{T} \left(\sum_{k=1}^{K_{t-1}} \sum_{j \in J_t^k} p_j \min_{i \in I_t^k} |\xi_t^i - \xi_t^j|^r \right)^{1/r}. \tag{18}$$

4.2 Flexible algorithm

Summarizing the above ideas yields the following scenario tree construction algorithm that allows to control the tree structure as well as the approximation tolerance with respect to the L_r-distance.

Algorithm 1 (forward tree construction)
Let N scenarios ξ^i with probabilities p_i, $i = 1, ..., N$, fixed root $\xi_1^* \in \mathbb{R}^d$, $r \geq 1$ and tolerances ε_r, ε_t, $t = 2, ..., T$, be given such that $\sum_{t=2}^{T} \varepsilon_t \leq \varepsilon_r$.

Step 1: Set $\hat{\xi}^1 := \hat{\xi}$ and $\mathcal{C}_1 = \{\{1, ..., N\}\}$.
Step t: Let $\mathcal{C}_{t-1} = \{C_{t-1}^1, ..., C_{t-1}^{K_{t-1}}\}$. 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 $\alpha_t(\,.\,)$ according to (14) and a stochastic process $\hat{\xi}^t$ having N scenarios $\hat{\xi}^{t,i}$ with probabilities p_i according to (15) and such that

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

Set $\mathcal{C}_t = \{\alpha_t^{-1}(i) : i \in I_t^k,\ k = 1, ..., K_{t-1}\}$.
Step T+1: Let $\mathcal{C}_T = \{C_T^1, ..., C_T^{K_T}\}$. Construct a stochastic process ξ_{tr} having K_T scenarios ξ_{tr}^k such that $\xi_{\mathrm{tr},t}^k := \xi_t^{\alpha_t(i)}$, $t = 1, ..., T$, if $i \in C_T^k$ with probabilities π_T^k according to (17), $k = 1, ..., K_T$.

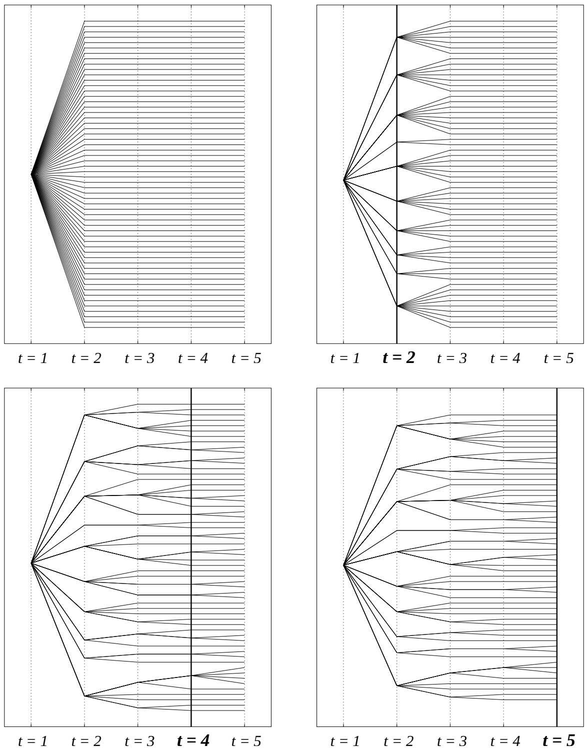

Fig. 2. Stepwise scenario tree construction for an example

While the first picture in Figure 2 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 $\xi_{\rm tr}$. The proof of the following corollary is also given in [23].

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

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

The next results states that the distance $|v(\xi) - v(\xi_{\rm tr})|$ of optimal values gets small if the initial tolerance ε in (11) as well as ε_r are small (cf. [22, Theorem 3.4].

Theorem 4. *Let (A1), (A2) and (A3) be satisfied with $r' \in [1, \infty)$ and the set X_1 be nonempty and bounded. Let $L > 0$, $\delta > 0$ and $C > 0$ be the constants appearing in Theorem 1 and (8), respectively. If $(\varepsilon_r^{(n)})$ is a sequence tending to*

0 *such that the corresponding tolerances* $\varepsilon_t^{(n)}$ *in Algorithm 1 are nonincreasing for all* $t = 2, ..., T$*, the corresponding sequence* $(\xi_{\mathrm{tr}}^{(n)})$ *has the property*

$$\limsup_{n\to\infty} |v(\xi) - v(\xi_{\mathrm{tr}}^{(n)})| \leq L \max\{1, C\}\varepsilon, \tag{19}$$

where $\varepsilon > 0$ *is the initial tolerance in (11).*

5 Polyhedral risk functionals

The results and methods from section 3 and section 4 rest upon the linearity of problem (1) to some extent. Hence, in general they are not valid for the risk-averse problem (2) incorporating a general (nonlinear) risk functionals ρ such as, e.g., Value-at-Risk ($\rho = \mathrm{VaR}_\alpha$) or standard deviation. Also algorithmic approaches for (1) might be destroyed by the incorporation of general risk functionals. However, in this section we consider the risk-averse problem (2) with ρ being chosen as a so-called *polyhedral risk functional*. This class of risk functionals has been introduced in [8, 7]. The key feature of these functionals is that they, though being non-linear, do not destroy mathematical structures of stochastic programs such as linearity or convexity.

5.1 Definition

The reason for the favorable behavior of polyhedral risk functionals in (2) is obvious from their definition: a polyhedral risk functional ρ is given by (the optimal value of) a linear stochastic minimization problem of the form

$$\rho(z) = \inf \left\{ \mathbb{E}\left[\sum_{j=0}^{J} \langle c_j, y_j \rangle\right] \middle| \begin{array}{l} y \in \times_{j=0}^{J} L_p(\Omega, \sigma(\xi^{t_j}), \mathbb{P}; \mathbb{R}^{k_j}) \\ y_j \in Y_j \ \mathbb{P}\text{-almost surely (a.s.)} \ (j = 0, ..., J), \\ \sum_{\tau=0}^{j} \langle w_{j,\tau}, y_{j-\tau} \rangle = z_{t_j} \ \text{ a.s. } (j = 1, ..., J), \\ \sum_{\tau=0}^{j} V_{j,\tau} y_{j-\tau} = r_j \ \text{ a.s. } (j = 0, ..., J) \end{array} \right\} \tag{20}$$

for every $z = (z_{t_1}, ..., z_{t_J}) \in \times_{j=1}^{J} L_p(\Omega, \sigma(\xi^{t_j}), \mathbb{P})$ with some $p \in [1, \infty)$. The numbers $k_j \in \mathbb{N}_0$, $d_j \in \mathbb{N}_0$ $(j = 0, ..., J)$, vectors $c_j \in \mathbb{R}^{k_j}$, $r_j \in \mathbb{R}^{d_j}$ $(j = 0, ..., J)$, $w_{j,\tau} \in \mathbb{R}^{k_{j-\tau}}$ $(j = 1, ..., J,\ \tau = 0, ..., j)$, matrices $V_{j,\tau} \in \mathbb{R}^{d_j \times k_{j-\tau}}$ $(j = 0, ..., J,\ \tau = 0, ..., j)$, and polyhedral cones $Y_j \subseteq \mathbb{R}^{k_j}$ $(j = 0, ..., J)$ have to be chosen in advance such that the resulting functional exhibits suitable risk functional properties. Clearly, if definition (20) is inserted into (2) with[1] $\gamma = 1$, one ends up with the problem

[1] The choice $\gamma = 1$ is not restrictive at all since the so-called *mean-risk objective* $\gamma \cdot \rho(z_{t_1}, ..., z_{t_J}) - (1 - \gamma) \cdot \mathbb{E}[z_T]$ can be expressed as another polyhedral risk functional of the form (20); cf. [8, 7].

$$\min\left\{\mathbb{E}\left[\sum_{j=0}^{J}\langle c_j, y_j\rangle\right]\left|\begin{array}{l} x\in\times_{t=1}^{T}L_{r'}(\Omega,\mathcal{A}_t,\mathbb{P};\mathbb{R}^{m_t}),\ x_t\in X_t \text{ a.s. } (t\geq 1),\\ y\in\times_{j=1}^{J}L_p(\Omega,\mathcal{A}_{t_j},\mathbb{P};\mathbb{R}^{k_j}),\ y_j\in Y_j \text{ a.s. } (j\geq 0),\\ A_{t,0}x_t+A_{t,1}(\xi_t)x_{t-1}=h_t(\xi_t) \text{ a.s. } (t=2,...,T),\\ z_t=z_t(x,\xi):=-\sum_{\tau=1}^{t}\langle b_\tau(\xi_\tau),x_\tau\rangle\ (t=1,...,T),\\ \sum_{\tau=0}^{j}\langle w_{j,\tau},y_{j-\tau}\rangle=z_{t_j} \text{ a.s. } (j=1,...,J),\\ \sum_{\tau=0}^{j}V_{j,\tau}y_{j-\tau}=r_j \text{ a.s. } (j=0,...,J)\end{array}\right.\right\} \tag{21}$$

i.e., the non-linearity of the functional ρ is transformed into a problem of the form (1) with additional decision variables y_j and additional linear constraints. This fact is not only useful for stability analysis (see below), it is also appreciated with regard to algorithmic issues. Note that this transformation is also possible if integer variables are incorporated into (1).

Most well-known risk functionals (e.g., VaR_α and standard deviation which are both not polyhedral) depend on a single random variable z only rather than on a finite sequence $z_{t_1},...,z_{t_J}$. In the framework of (2) this means $J=1$ and $t_1=T$. Several coherence axioms for such single-period risk functionals have been suggested in [1, 14, 30] which are broadly accepted. For medium- and long-term economic activities (such as the model in section 6) one may want to use multi-period risk functionals ($J>1$) that take into account the temporal development of profits and losses, e.g., to avoid liquidity problems at intermediate time steps. Also for this case coherence axioms are suggested [2, 15, 33]. In both the single- and the multi-period case such axioms give directions for the choice of the vectors and matrices in (20).

5.2 Properties

Because the arguments z_{t_j} in (20) appear on the right-hand sides of the constraints, it can be concluded that the functional ρ is always *convex* [8, 7]. Hence, the theory of convex duality can be applied. This yields dual representations for ρ which can be useful for interpretation and verification of coherence axioms, and for algorithmic approaches, too.

Theorem 5. *([8, 33, 7]) Let ρ be a polyhedral risk functional of the form (20) and let the following conditions be satisfied for Y_j, c_j, $w_{j,\tau}$, and $V_{j,\tau}$:*

- **complete recourse**: $\begin{pmatrix} V_{j,0}\\ w'_{j,0}\end{pmatrix} Y_j=\mathbb{R}^{d_j+1}$ $(j=1,...,J)$,
- **dual feasibility**: $\bigcap_{j=0}^{J}\mathcal{D}_{\rho,j}\neq\emptyset$ *with*

$$\mathcal{D}_{\rho,j}:=\left\{\begin{array}{l}(u_v,u_w)\in\mathbb{R}^J\times\mathbb{R}^{\sum d_j}:\\ c_j+\sum_{\nu=\max\{1,j\}}^{J}u_{v,\nu}w_{\nu,\nu-j}+\sum_{\nu=j}^{J}V^*_{\nu,\nu-j}u_{w,\nu}\in -Y_j^*\end{array}\right\}.$$

Then the functional ρ is finite, convex, and continuous on $\times_{j=1}^{J}L_p(\Omega,\sigma(\xi^{t_j}),\mathbb{P})$ and it is representable by

$$\rho(z) = \sup \left\{ -\mathbb{E}\left[\sum_{j=1}^{J}\left(\lambda_j z_{t_j} + \langle \mu_j, r_j \rangle\right)\right] \middle| \begin{array}{l} \lambda_j \in L_{p'}(\Omega, \sigma(\xi^{t_j}), \mathbb{P}), \\ \mu_j \in L_{p'}(\Omega, \sigma(\xi^{t_j}), \mathbb{P}; \mathbb{R}^{d_j}), \\ (\mathbb{E}[\lambda|\xi^{t_j}], \mathbb{E}[\mu|\xi^{t_j}]) \in \mathcal{D}_{\rho,j} \;\; a.s. \\ (j = 0, ..., J) \end{array} \right\}$$

with $p' \in (1, \infty]$ *being defined by* $1/p + 1/p' = 1$.

The above dual representation can be read as follows: the supremum operator aims at making λ large where z is small (in compliance with the respective constraints). Hence, $\rho(z)$ can be understood as a *worst case weighted expectation* of z (possibly biased by $\langle \mu_j, r_j \rangle$). If ρ satisfies the coherence axioms from [2], then (and only then) the constraints in the dual representation are such that all the λ multipliers are probability densities and $\langle \mu_j, r_j \rangle$ is always zero.

5.3 Single-period examples

For $J = 1$ and $t_1 = T$, i.e., for the single-period situation, polyhedral risk functionals can be found in economic literature.

Example 1. The *Conditional* or *Average Value-at-Risk* at level $\alpha \in (0, 1)$ (CVaR_α or AVaR_α, cf. [34] and [14, Chapter 4.4]) is given by

$$\mathrm{AVaR}_\alpha(z) := \tfrac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_{\bar{\alpha}}(z) d\bar{\alpha} = \inf_{y_0 \in \mathbb{R}} \left\{ y_0 + \tfrac{1}{\alpha}\mathbb{E}\left[(y_0 + z)^-\right]\right\} \tag{22}$$

where the representation on the right is due to [34]. By introducing variables for positive and negative parts of $y_0 + z$, respectively, AVaR_α can be rewritten in the form (20) with $J = 1$, $d_0 = d_1 = 0$, $k_0 = 1$, $k_1 = 2$, $c_0 = 1$, $c_1 = \left(0, \frac{1}{\alpha}\right)$, $w_{1,0} = (1, -1)$, $w_{1,1} = -1$, $Y_0 = \mathbb{R}$, and $Y_1 = \mathbb{R}^2_+$. Hence, AVaR_α is a polyhedral risk functional. Moreover, complete recourse and dual feasibility are satisfied and the dual representation of Theorem 5 reads

$$\mathrm{AVaR}_\alpha(z) = \sup\left\{-\mathbb{E}[\lambda z] : \lambda \in L_{p'}(\Omega, \mathcal{F}, \mathbb{P}),\ \lambda \in [0, \tfrac{1}{\alpha}] \text{ a.s.}, \mathbb{E}[\lambda] = 1\right\}$$

where the λ multipliers can be interpreted as densities. We note that AVaR_α is known to be a convex risk functional in the sense of [14], a coherent risk functional in the sense of [1], and it is 1st and 2nd order stochastic dominance consistent [30].

Example 2. Consider *expected utility* as a risk functional, i.e., $\rho_u(z) = -\mathbb{E}[u(z)]$ with some concave and non-decreasing utility function $u : \mathbb{R} \to \mathbb{R}$. This approach goes back to [29]. Typically, non-linear utility functions $u : \mathbb{R} \to \mathbb{R}$ are used within this framework. Of course, in this case ρ_u cannot be represented by a linear stochastic program. However, in cases when the domain of the outcome z can be bounded a priori, it makes sense to consider piecewise linear utility functions u. In that case, $-u$ is convex and piecewise linear, hence, according to [35, Example 3.54] there exist $k \in \mathbb{N}$, $w \in \mathbb{R}^k$, $c \in \mathbb{R}^k$, and $v \in \{0,1\}^k$ such that

$$-u(\mu) = \inf \left\{ \langle c, y \rangle \,\middle|\, y \in \mathbb{R}^k,\ y \geq 0\ \langle w, y \rangle = \mu,\ \langle v, y \rangle = 1 \right\}$$

for all $\mu \in \mathbb{R}$. For this case, the expected utility risk functional reads

$$\rho_u(z) = \inf \left\{ \mathbb{E}[\langle c, y_1 \rangle] \,\middle|\, \begin{array}{l} y_1 \in L_p(\Omega, \mathcal{A}, \mathbb{P}; \mathbb{R}^k),\ y_1 \geq 0 \ \text{ a.s.} \\ \langle w, y_1 \rangle = z \ \text{ a.s.},\ \langle v, y_1 \rangle = 1 \ \text{ a.s.} \end{array} \right\}$$

where [35, Theorem 14.60] is used to justify the interchange of infimum and expectation. Hence, ρ_u is a polyhedral risk functional with $k_0 = d_0 = 0$, $k_1 = k$, $d_1 = 1$, $c_1 = c$, $w_{1,0} = w$, $V_{1,0} = v'$, and $Y_1 = \mathbb{R}_+^{k_1}$. The special case of the *expected regret* (expected loss), i.e., the case that $\rho(z) = \mathbb{E}[(z - \gamma)^-]$ with some target $\gamma \in \mathbb{R}$, is obtained by setting $k = 3$, $w = (\gamma, 1, -1)$, $v = (1, 0, 0)$, and $c = (0, 0, -1)$.

5.4 Multi-period examples

For $J > 1$, i.e., for the multi-period situation, only few (polyhedral) risk functionals are suggested in economic literature. However, the framework of polyhedral risk functionals is *constructive*: various multi-period polyhedral risk functionals have been proposed in [8, 7, 33] that can be understood as multi-period extentions of AVaR_α. They all satisfy the basic risk coherence axioms from [2], but they differ with respect to the incorporation of the information dynamics. We present a selection of those in the following (keeping the original index numbers). It is assumed that the random variables z_t represent *accumulated* revenues as in problem (2).

Example 3. The functional

$$\rho_2(z_{t_1}, \dots, z_{t_J}) := \inf_{y_0 \in \mathbb{R}} \left\{ y_0 + \tfrac{1}{\alpha} \tfrac{1}{J} \textstyle\sum_{j=1}^{J} \mathbb{E}\left[\left(z_{t_j} + y_0 \right)^- \right] \right\}.$$

from [8] can be understood as AVaR_α applied to a compound lottery, i.e., applied to z_0 given by $z_0(\omega) := z_{\iota(\omega)}(\omega)$ with ι being uniformly distributed on $\{t_1, \dots, t_J\}$ and independent of $z_{t_1}, \dots, z_{t_J}$. Clearly, ρ_2 can be represented through (20) by introducing (stochastic) variables for the positive and the negative part of $z_{t_j} + y_0$, respectively, for $j = 1, \dots, J$. Hence, it is a polyhedral risk functional. It satisfies complete recourse and dual feasibility. The dual representation according to Theorem 5 given by

$$\rho_2(z) = \sup \left\{ -\mathbb{E}\left[\textstyle\sum_{j=1}^{J} \lambda_j z_{t_j} \right] \,\middle|\, \begin{array}{l} \lambda \in \times_{j=1}^{J} L_p(\Omega, \sigma(\xi^{t_j}), \mathbb{P}),\ \sum_{j=1}^{J} \mathbb{E}[\lambda_j] = 1 \\ \lambda_j \in [0, \frac{1}{\alpha}] \text{ a.s. } (j = 1, \dots, J), \end{array} \right\}$$

aims at placing the available probability mass of λ to stages where $z = (z_{t_1}, \dots, z_{t_J})$ attains low values.

Example 4. The polyhedral risk functional ρ_4 from [8], though being defined via an infimum representation of the form (20), is easier to catch by its dual representation according to Theorem 5 given by

$$\rho_4(z) = \sup\left\{ -\mathbb{E}\left[\sum_{j=1}^{J} \lambda_j z_{t_j}\right] \left| \begin{array}{l} \lambda \in \times_{j=1}^{J} L_p(\Omega, \sigma(\xi^{t_j}), \mathbb{P}), \\ \lambda_j \in [0, \frac{1}{\alpha}] \text{ a.s., } \mathbb{E}[\lambda_j] = \frac{1}{J} \ (j = 1, ..., J) \\ \lambda_j = \mathbb{E}[\lambda_{j+1} | \sigma(\xi^{t_j})] \text{ a.s. } (j = 1, ..., J-1) \end{array} \right. \right\}$$

with $z = (z_{t_1}, ..., z_{t_J})$. Here, the multiplier process λ has to be a martingale and, hence, all time steps are weighted equally.

Example 5. In [2] it was suggested to apply a single-period risk functional to the pointwise minimum of $z = (z_{t_1}, ..., z_{t_J})$, i.e., to z_0 given by $z_0(\omega) := \min\{z_{t_1}(\omega), ..., z_{t_J}(\omega)\}$. Doing so by using AVaR_α yields the functional

$$\begin{aligned} \rho_6(z) &= \inf_{y_0 \in \mathbb{R}} \left(y_0 + \tfrac{1}{\alpha}\mathbb{E}[\,(y_0 + z_0)^-\,]\right) \\ &= \inf_{y_0 \in \mathbb{R}} \left(y_0 + \tfrac{1}{\alpha}\mathbb{E}[\,\max\{0, -y_0 - z_{t_1}, ..., -y_0 - z_{t_J}\}]\right) \end{aligned}$$

which can also be represented in the form (20) by introducing (stochastic) variables $y_{j,2} = \max\{0, -y_0 - z_{t_1}, ..., -y_0 - z_{t_j}\} = \max\{y_{j-1,2}, -y_0 - z_{t_j}\}$ for $j = 1, ..., J$; cf. [7]. Then, complete recourse and dual feasibility are satisfied and there is also a dual representation according to Theorem 5.

5.5 Stability

At the first glance it seems as if stability of problem (2) with ρ being chosen as a polyhedral risk functional (20) were covered by the results from section 3 due to the reformulation (21). However, a closer look to the latter problem reveals that it is not completely of the form (1): the resulting recourse matrices become stochastic when the dynamic constraints in (21) are integrated. Hence, Theorem 1 and Theorem 2 are not valid for problem (21) and cannot be suitably modified easily.

For this reason, stability of (2) is analyzed in [10, 7] systematically. Starting with the finding of further continuity properties of ρ (stronger than plain continuity as stated in Theorem 5), a stability theorem for the optimal values (corresponding to Theorem 1) can be proven. However, the filtration distance there is even more involved than D_f in (5) from Theorem 1.

For the justification of the scenario tree generation methods in section 4, it is necessary to estimate these problem dependent objects by problem independent ones as in (8). In order to get a similar estimate for the involved filtration distance for problem (2), it turns out to be necessary to impose further technical conditions on ρ (beside complete recourse and dual feasibility). However, these conditions can be shown to be satisfied for all known polyhedral risk functionals from [8, 7, 33] as long as the integrability number p is set to 1. We conclude that there is a theoretical basis for the scenario tree approximation methods from section 4 also in the situation of the risk-averse problem (2) if ρ is chosen as a suitable polyhedral risk functional.

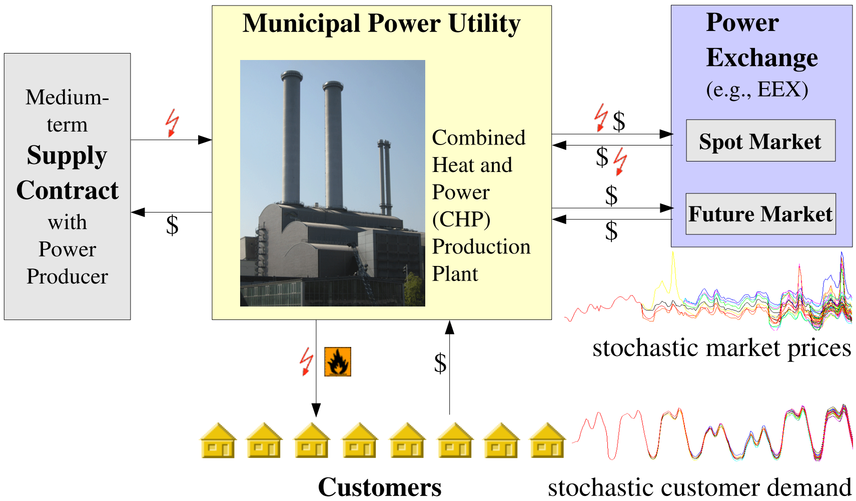

Fig. 3. Schematic diagram for the optimization model components

6 Case study

In this final section we demonstrate the use of the above theoretical results by presenting some simulation results from a power portolio optimization model; cf. Fig. 3. For motivation and for a detailed technical description of this model see [11, 9]; in the following, we describe its components on a more abstract level only. Its numerical output shall then illustrate the usage of scenario trees as well as the effect of different polyhedral risk functionals.

6.1 Model

Taking into account uncertainties in power portfolio optimization yields quite automatically to stochastic programming; see, e.g., [40]. The optimization model here is a mean-risk multi-stage stochastic program of the form (2). It is tailored to the one year planning situation of a certain (German) municipal power utility serving an electricity demand and a heat demand for certain customers; see Fig. 3. The (German) power market induces an hourly time discretization, hence, we have $T = 365 \cdot 24 = 8760$ time steps. Energy demands as well as market prices for each hour in the future are unknown at previous time steps. These uncertainties can be described reasonably by stochastic time series models; cf. [11]. It is assumed that the power utility is sufficiently small such that it can be considered as a *price-taker*, i.e., its decisions do not affect market prices or demands.

The concrete situation of the power utility is supposed to be as follows: It features a *combined heat and power* (CHP) production plant that can serve the heat demand completely but the electricity demand only in part. Hence,

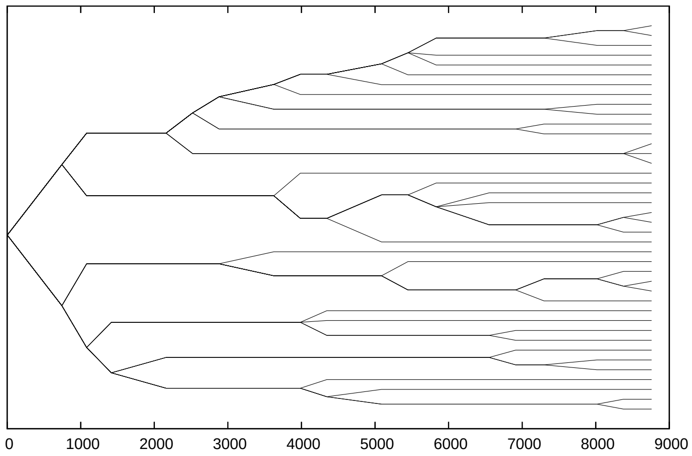

Fig. 4. Branching structure of the input scenario tree of 40 scenarios ($T = 8760$)

additional sources of electricity have to be used. Electricity can be obtained from the spot market of a power exchange (such as the European Energy Exchange EEX in Germany), or by purchasing a bilateral supply contract from a larger power producer. The latter possibility is suspected to be more expensive, but relying on the spot market only is known to be extremely risky. Spot price risk, however, may be reduced (hedged) by means of derivative products. Here, we consider futures from EEX (Phelix-futures, purely financial contracts).

The original practical purpose of this model was to evaluate given supply contracts in comparison with the possibility of relying on spot and future market only [9, 11]. In the presentation here, however, we focus on the qualitative output with respect to the effect of the different polyhedral risk functionals from section 5. Therefore, no such supply contracts are considered in the portfolio here.

The stochastic input process $\xi = (\xi_1, ..., \xi_T)$, modeled by an appropriate time series model (cf. [11]), is approximated by a scenario tree (cf. Fig. 4) according to the methods from section 4. Each random vector ξ_t consists of 27 components: electricity demand ξ_t^e, heat demand ξ_t^h, EEX spot prices ξ_t^s, as well as base and peak future prices ξ_t^{fbm} and ξ_t^{fpm} (for each month $m = 1, ..., 12$). However, to avoid technical problems related to arbitrage, the future prices are calculated as *fair prices* from the spot prices in the scenario tree, i.e., the methods from section 4 are applied only to the first three components ξ_t^e, ξ_t^h, and ξ_t^s $(t = 1, ..., T)$.

The decisions at each time t consist of CHP production amounts, EEX spot market volumes (electricity may be bought or sold), future stock, and contract flexibility (if there is any). The CHP production is subject to several technical (dynamic) constraints which are slightly simplified such that no integer variables come into play, i.e., everything is linear. There are no particular constraints for spot and future trading, but the pricing rules for EEX futures (initial margin, variation margin, transaction costs) make it necessary to introduce some auxiliary variables and constraints. Finally, there are the demand satisfaction constraints requiring that electricity demand and heat demand are always met. For further details we refer to [9]. The overall model (incorporating a polyhedral risk functional) is linear, i.e., it is of the form (2) resp. (21). Of course, the latter formulation is used for implementation.

6.2 Simulations results

Together with a fixed scenario tree (cf. Fig. 4) the overall optimization model is a (large-scale) linear program. For the simulation results presented here, we used a scenario tree of 40 scenarios and approx. $150,000$ nodes. The decision variables are defined on the nodes of the tree. For solving the linear program the ILOG CPLEX 9.1 software was employed. We restrict the presentation here to the case that no additional supply contracts are involved (beside EEX futures). Then, the different effects of the polyhedral risk functionals from section 5 can be observed best.

In Fig. 5 the accumulated revenues z_t over time for each scenario, i.e., the temporal developments of the company's wealth, are shown after optimization with different polyhedral risk functionals. Of course, the tree structure of the input scenario tree can also be found in these outputs since the (optimal) revenues are stochastic in the same manner as the inputs. Optimizing the expected overall revenue $\mathbb{E}[z_T]$ only (without any risk functional) yields large dispersion (spread) at time T (cf. top of Fig. 5). The incorporation of the (single-period) AVaR applied to z_T reduces this spread considerably, but yields high spread and very low values for z_t at earlier time steps $t < T$. Clearly, this behavior is not acceptable for a (small) power utility. The multi-period polyhedral risk functionals from section 5 are effective such that dispersion is somehow better distributed over all time steps.

The graphs in Fig. 5 suggest that the effect of ρ_2, ρ_4, and ρ_6 is more or less the same. However, Fig. 6 reveals that there are further differences among these multi-period risk functionals. For the calculation of these graphs, the fuel costs for the CHP plant have been slightly augmented in order to give the cash value curves a different direction. The difference between the multi-period functionals is, roughly speaking, that ρ_4 aims at equal spread at all times, whereas ρ_2 and ρ_6 try to find a maximal level that is rarely underrun.

The different shapes of the cash value curves are achieved by different policies of future trading. Future trading is revealed through the jumps in the cash value curves and is explicitely shown in Fig. 7. These graphs display the overall future stock volumes (in Euro) at each time step. If no risk is considered then there is no future trading at all since, due to the fair-price assumption, there is no benefit from futures in terms of the expected revenue.

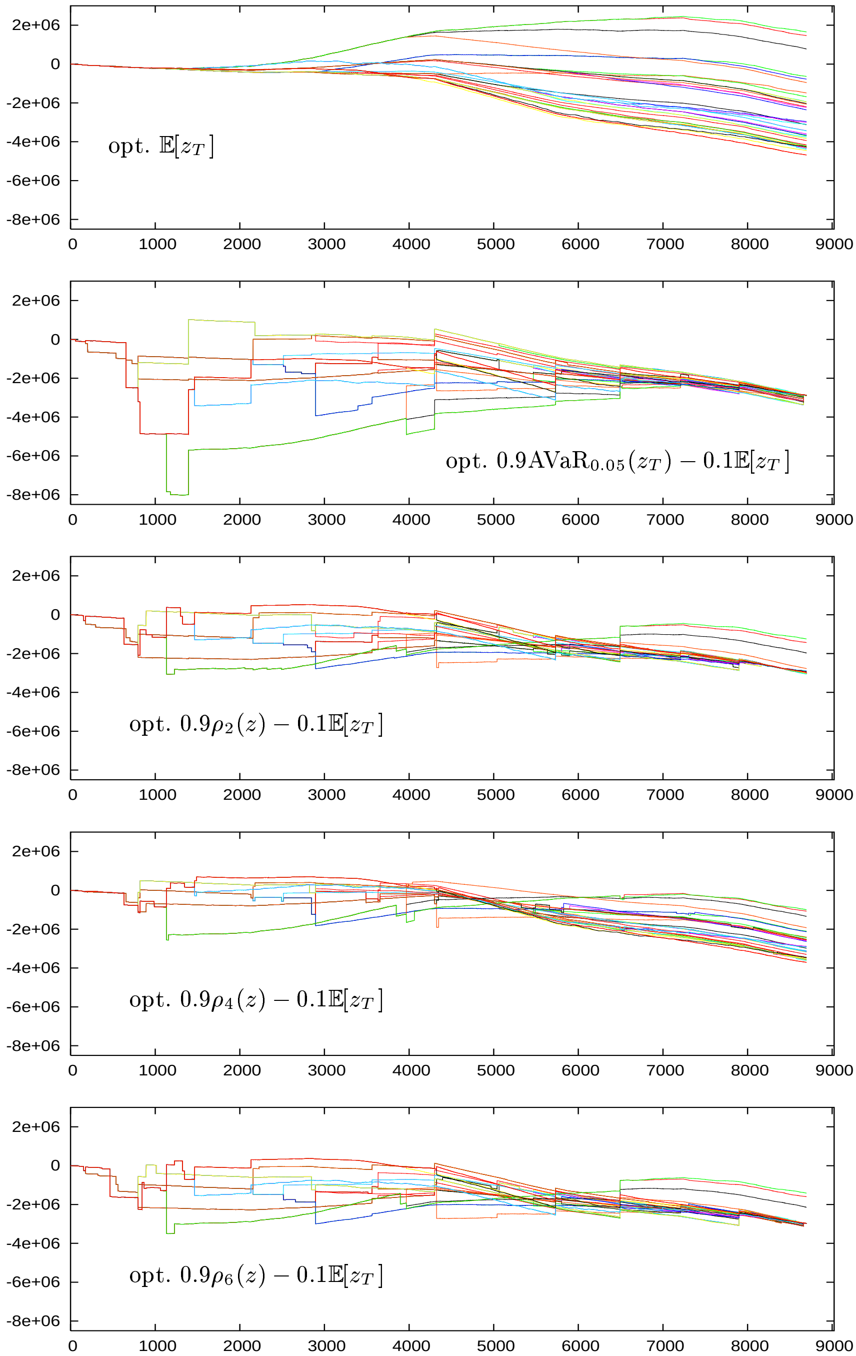

Fig. 5. Optimal cash values z_t (wealth) over time ($t = 1, ..., T$) for each scenario

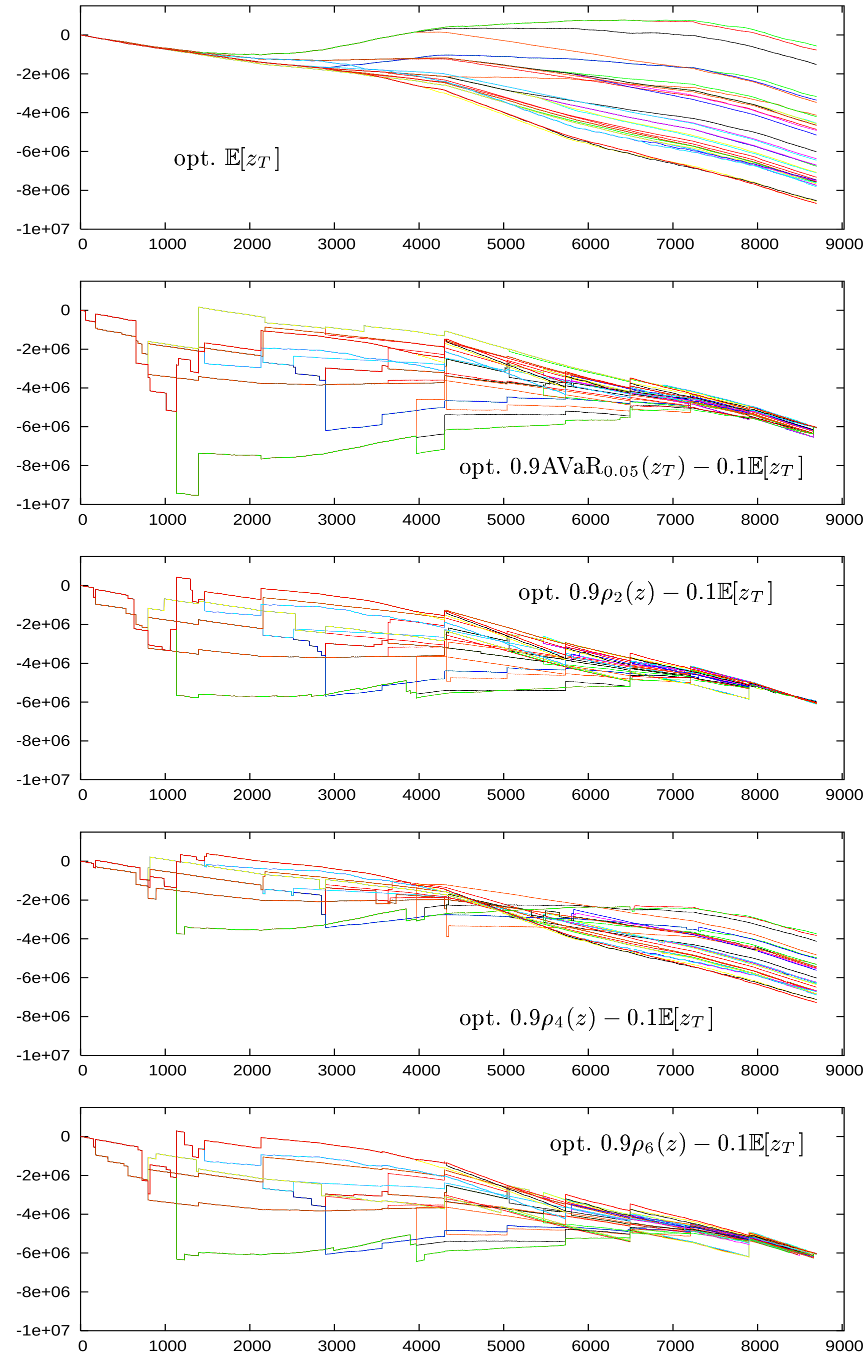

Fig. 6. Optimal cash values z_t (wealth) over time ($t = 1, ..., T$), **high fuel costs**

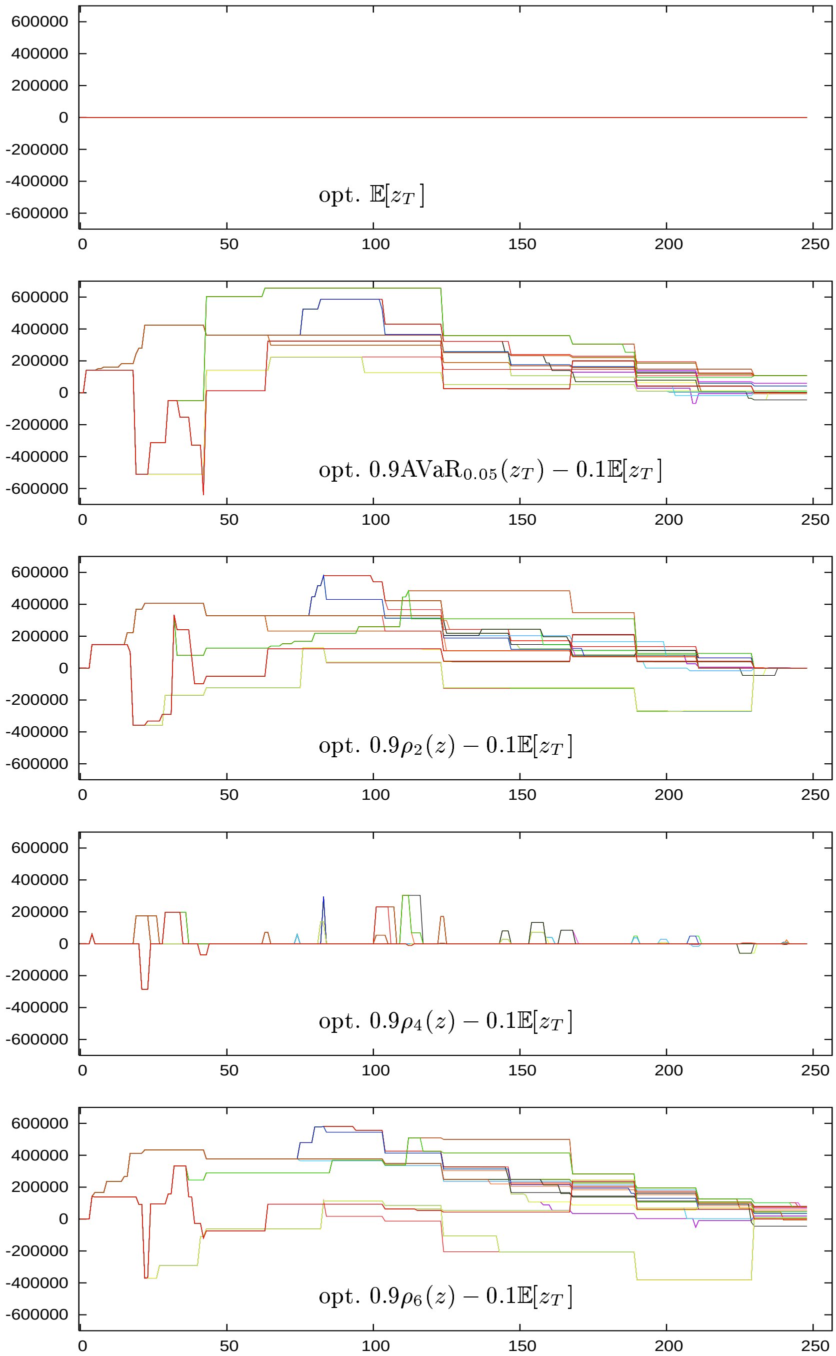

Fig. 7. Overall future stock over time (248 trading days), **high fuel costs**

Using AVaR, ρ_2, or ρ_6 leads to extensive future trading activity, whereas the application of ρ_4 yields more moderate future trading activity.

Finally, we mention that, within this application model, the incorporation of a polyhedral risk functional into the objective reduces the expected overall revenue $\mathbb{E}[z_T]$ only by approx. 1%. The additional computational effort arising from the risk measure is also very moderate.

7 Conclusion

We have presented a capacious theory for the framework of multi-stage stochastic programming. Though appearing rather technical and abstract at the first glance, these results are highly relevant in practice: Problems become numerically tractable by finite scenario tree approximation of the underlying stochastic input data. Moreover, risk-averion requirements can be incorporated without significant increase of complexity by means of polyhedral risk functionals. In particular, there is a theoretical basis for the scenario tree approximation methods in both cases, the risk-neutral and the risk-averse case. For illustration, we have presented an exemplary model for mean-risk optimization of an electricity portfolio.

Acknowledgement. The presented work was supported by the DFG Research Center MATHEON "Mathematics for Key Technologies" in Berlin (`http://www.matheon.de`) and by the "Wiener Wissenschafts-, Forschungs- und Technologiefonds" in Vienna (`http://www.univie.ac.at/crm/simopt`).

References

1. P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.
2. P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, and H. Ku. Coherent multiperiod risk adjusted values and Bellman's principle. *Annals of Operations Research*, 152:5–22, 2007.
3. B. Blaesig. *Risikomanagement in der Stromerzeugungs- und Handelsplanung*, volume 113 of *Aachener Beiträge zur Energieversorgung*. Klinkenberg, Aachen, Germany, 2007. PhD Thesis.
4. M. S. Casey and S. Sen. The scenario generation algorithm for multistage stochastic linear programming. *Mathematics of Operations Research*, 30:615–631, 2005.
5. M. A. H. Dempster. Sequential importance sampling algorithms for dynamic stochastic programming. *Zapiski Nauchnykh Seminarov POMI*, 312:94–129, 2004.
6. J. Dupačová, G. Consigli, and S. W. Wallace. Scenarios for multistage stochastic programs. *Annals of Operations Research*, 100:25–53, 2000.

7. A. Eichhorn. *Stochastic Programming Recourse Models: Approximation, Risk Aversion, Applications in Energy*. PhD thesis, Department of Mathematics, Humboldt University, Berlin, 2007.
8. A. Eichhorn and W. Römisch. Polyhedral risk measures in stochastic programming. *SIAM Journal on Optimization*, 16:69–95, 2005.
9. A. Eichhorn and W. Römisch. Mean-risk optimization models for electricity portfolio management. In *Proceedings of the 9th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, Stockholm, Sweden, 2006.
10. A. Eichhorn and W. Römisch. Stability of multistage stochastic programs incorporating polyhedral risk measures. *Optimization*, to appear, 2008.
11. A. Eichhorn, W. Römisch, and I. Wegner. Mean-risk optimization of electricity portfolios using multiperiod polyhedral risk measures. In *IEEE St. Petersburg PowerTech Proceedings*, 2005.
12. S.-E. Fleten and T. K. Kristoffersen. Short-term hydropower production planning by stochastic programming. *Computers and Operations Research*, 35:2656–2671, 2008.
13. S.-E. Fleten, S. W. Wallace, and W. T. Ziemba. Hedging electricity portfolios via stochastic programming. In Greengard and Ruszczyński [17], pages 71–93.
14. H. Föllmer and A. Schied. *Stochastic Finance. An Introduction in Discrete Time*, volume 27 of *De Gruyter Studies in Mathematics*. Walter de Gruyter, Berlin, 2nd edition, 2004.
15. M. Frittelli and G. Scandolo. Risk measures and capital requirements for processes. *Mathematical Finance*, 16:589–612, 2005.
16. S. Graf and H. Luschgy. *Foundations of Quantization for Probability Distributions*, volume 1730 of *Lecture Notes in Mathematics*. Springer, Berlin, 2000.
17. C. Greengard and A. Ruszczyński, editors. *Decision Making under Uncertainty: Energy and Power*, volume 128 of *IMA Volumes in Mathematics and its Applications*. Springer, New York, 2002.
18. N. Gröwe-Kuska, K. C. Kiwiel, M. P. Nowak, W. Römisch, and I. Wegner. Power management in a hydro-thermal system under uncertainty by Lagrangian relaxation. In Greengard and Ruszczyński [17], pages 39–70.
19. N. Gröwe-Kuska and W. Römisch. Stochastic unit commitment in hydrothermal power production planning. In S. W. Wallace and W. T. Ziemba, editors, *Applications of Stochastic Programming*, MPS/SIAM Series on Optimization, pages 633–653. SIAM, Philadelphia, PA, USA, 2005.
20. H. Heitsch. *Stabilität und Approximation stochastischer Optimierungsprobleme*. PhD thesis, Department of Mathematics, Humboldt University, Berlin, 2007.
21. H. Heitsch and W. Römisch. Scenario reduction algorithms in stochastic programming. *Computational Optimization and Applications*, 24:187–206, 2003.
22. H. Heitsch and W. Römisch. Stability and scenario trees for multistage stochastic programs. In G. Infanger, editor, *Stochastic Programming - The State of the Art*. 2007. to appear.
23. H. Heitsch and W. Römisch. Scenario tree modeling for multistage stochastic programs. *Mathematical Programming*, to appear, 2008.
24. H. Heitsch, W. Römisch, and C. Strugarek. Stability of multistage stochastic programs. *SIAM Journal on Optimization*, 17:511–525, 2006.
25. R. Hochreiter and G. Ch. Pflug. Financial scenario generation for stochastic multi-stage decision processes as facility location problems. *Annals of Operations Research*, 152:257–272, 2007.

26. R. Hochreiter, G. Ch. Pflug, and D. Wozabal. Multi-stage stochastic electricity portfolio optimization in liberalized energy markets. In *System Modeling and Optimization*, IFIP International Federation for Information Processing, pages 219–226. Springer, Boston, MA, USA, 2006.
27. K. Høyland, M. Kaut, and S. W. Wallace. A heuristic for moment-matching scenario generation. *Computational Optimization and Applications*, 24:169–185, 2003.
28. B. Krasenbrink. *Integrierte Jahresplanung von Stromerzeugung und -handel*, volume 81 of *Aachener Beiträge zur Energieversorgung*. Klinkenberg, Aachen, Germany, 2002. PhD Thesis.
29. J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ, USA, 1944.
30. W. Ogryczak and A. Ruszczyński. On consistency of stochastic dominance and mean-semideviation models. *Mathematical Programming*, 89:217–232, 2001.
31. T. Pennanen. Epi-convergent discretizations of multistage stochastic programs via integration quadratures. *Mathematical Programming*, to appear, 2008.
32. M. V. F. Pereira and L. M. V. G. Pinto. Multi-stage stochastic optimization applied to energy planning. *Mathematical Programming*, 52:359–375, 1991.
33. G. Ch. Pflug and W. Römisch. *Modeling, Measuring, and Managing Risk*. World Scientific, Singapore, 2007.
34. R. T. Rockafellar and S. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26:1443–1471, 2002.
35. R. T. Rockafellar and R. J-B. Wets. *Variational Analysis*, volume 317 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, 1st edition, 1998. (Corr. 2nd printing 2004).
36. A. Ruszczyński and A. Shapiro, editors. *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*. Elsevier, Amsterdam, 1st edition, 2003.
37. H. K. Schmöller. *Modellierung von Unsicherheiten bei der mittelfristigen Stromerzeugungs- und Handelsplanung*, volume 103 of *Aachener Beiträge zur Energieversorgung*. Klinkenberg, Aachen, Germany, 2005. PhD Thesis.
38. S. Sen, L. Yu, and T. Genc. A stochastic programming approach to power portfolio optimization. *Operations Research*, 54:55–72, 2006.
39. S. Takriti, B. Krasenbrink, and L. S.-Y. Wu. Incorporating fuel constraints and electricity spot prices into the stochastic unit commitment problem. *Operations Research*, 48:268–280, 2000.
40. S. W. Wallace and S.-E. Fleten. Stochastic programming models in energy. In Ruszczyński and Shapiro [36], chapter 10, pages 637–677.