Numerical Evaluation of Approximation Methods in Stochastic Programming

CHRISTIAN KÜCHLER and STEFAN VIGERSKE

Humboldt–Universität zu Berlin, Unter den Linden 6,
D–10099 Berlin, Germany,
`ckuechler@math.hu-berlin.de`, `stefan@math.hu-berlin.de`

Abstract

We study an approach for the evaluation of approximation and solution methods for multistage linear stochastic programs by measuring the performance of the obtained solutions on a set of out-of-sample scenarios. The main point of the approach is to restore the feasibility of solutions to an approximate problem along the out-of-sample scenarios. For this purpose, we consider and compare different feasibility and optimality based projection methods. With this at hand, we study the quality of solutions to different test models based on classical as well as recombining scenario trees.

Keywords. Multistage Stochastic Programming, Scenario Tree, Out-of-sample Evaluation

AMS subject classification. 90C15, 90C39, 90C59, 49M25

1 Introduction

In general, numerical solution methods for stochastic optimization problems require the underlying probability measures to have only a finite support. Thus, different techniques have been developed to approximate random variables or stochastic processes by a limited number of scenarios or finite scenario trees, respectively. These techniques follow different principles like random sampling [19], moment matching [8], probability metrics [3, 4, 6, 15], and Quasi Monte-Carlo sampling [14]. Convergence of optimal values and/or solution sets has been proved for specific techniques and properties of statistical estimates and bounds have been established (cf., e.g., [19] and the references therein). Stability analysis of stochastic programs yields further hints how approximations should look like, cf. [5, 11, 13, 16].

Unfortunately, on the one hand, these theoretical results may require the optimization problems and underlying random variables to fulfill specific regularity assumptions that may be hard to verify in some cases of practical interest. On the other hand, quantitative error bounds and statistical properties are not available for all problem classes. Furthermore, due to the numerical complexity of stochastic programming models, it is sometimes necessary to use approximations that are too rough to obtain meaningful error bounds or confidence intervals via asymptotic results.

In such cases, one has to resort to numerical methods to measure the performance and quality of approximation and solution methods. Since a main task of stochastic programming is to provide decision strategies that are robust enough to be applicable in real-world scenarios, it suggests itself to measure the quality of an approximation method by evaluating the (optimal) solutions obtained from solving the approximate problem. This can be done, e.g., by evaluating these solutions along *out-of-sample scenarios*, cf., e.g., [10] and [2, 7] for one- and multistage problems, respectively.

In this paper, we study how out-of-sample testing may be used to study the behaviour of approximations to linear multistage stochastic programs (MSP). Thereby, we aim for problems with many stages, where, due to numerical complexity, the thoroughly construction of out-of-sample strategies as in [7] and the second method of [2] do not apply. Furthermore, our framework differs since we abstain from a (relatively) complete recourse assumption. Then, in particular, optimal solutions of an approximate problem are not necessarily feasible along out-of-sample scenarios. Therefore, the generation of feasible solutions out of solutions of an approximate problem is an important issue. Furthermore, this question may be of interest whenever one is interested in obtaining practically applicable solutions. For this *feasibility restoration* we adopt different projection approaches.

Considering MSPs from power scheduling and finance, the proposed feasibility restoration approaches are applied to study the quality of solutions obtained by the decomposition approach proposed in [12], based on recombining scenario trees, and solutions induced by non-recombining trees that have been constructed using the *Forward Tree Construction Algorithm* of Heitsch and Römisch [4, Algorithm 4.5].

2 Problem Formulation

On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we consider an $\mathbb{R}^s$-valued discrete time stochastic process $\boldsymbol{\xi} = (\boldsymbol{\xi}_t)_{t=1,\ldots,T}$. For $t = 1, \ldots, T$, the vector $(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_t)$ is denoted by $\boldsymbol{\xi}_{[t]}$. We study the following linear multistage stochastic program:

$$(P) \quad v(\boldsymbol{\xi}) := \inf \left\{ \mathbb{E}\left[\varphi(\boldsymbol{\xi}, x(\boldsymbol{\xi}))\right] : \begin{array}{ll} x \in \mathcal{M}_m, \quad x_t \geq 0, & t = 1, ..., T \\ A_{t,0} x_t(\boldsymbol{\xi}_{[t]}) + A_{t,1} x_{t-1}(\boldsymbol{\xi}_{[t-1]}) = h_t(\boldsymbol{\xi}_t), & t = 2, ..., T \end{array} \right\},$$

with the cost function

$$\varphi(\xi, x(\xi)) := \sum_{t=1}^{T} \left\langle c_t(\xi_t), x_t(\xi_{[t]}) \right\rangle.$$

The set $\mathcal{M}_m$ consists of all tuples $x = (x_1, \ldots, x_T)$ of Borel-measurable mappings $x_t : \mathbb{R}^{s \cdot t} \to \mathbb{R}^m$. The *costs* $c_t(\cdot)$ and the *right-hand sides* $h_t(\cdot)$ are affine mappings from $\mathbb{R}^s$ to $\mathbb{R}^m$ and $\mathbb{R}^r$, respectively, for $t = 1, \ldots, T$. The assumption of non-random technology and recourse matrices is due to notational simplicity.

Whenever the optimization problem (P) is not solvable analytically, a usual approach is to replace the process $\boldsymbol{\xi}$ by a process $\bar{\boldsymbol{\xi}}$ taking only a finite number of scenarios $\bar{\xi}^j = (\bar{\xi}_t^j)_{t=1,\ldots,T}$, $j \in J$, with J being some index set. The approximate problem

$$(\bar{P}) \quad v(\bar{\boldsymbol{\xi}}) := \inf \left\{ \mathbb{E}\left[\varphi(\bar{\boldsymbol{\xi}}, x(\bar{\boldsymbol{\xi}}))\right] : \begin{array}{ll} x \in \mathcal{M}_m, \quad x_t \geq 0, & t = 1, ..., T \\ A_{t,0} x_t(\bar{\boldsymbol{\xi}}_{[t]}) + A_{t,1} x_{t-1}(\bar{\boldsymbol{\xi}}_{[t-1]}) = h_t(\bar{\boldsymbol{\xi}}_t), & t = 2, ..., T \end{array} \right\},$$

may then be solved by numerical methods (that may be again approximative). Denoting by $\bar{x} = (\bar{x}_t(\cdot))_{t=1,\ldots,T}$ a solution resulting from the particular approximation and solution procedure, the optimal value of the approximate problem

$$v(\bar{\boldsymbol{\xi}}) = \mathbb{E}[\varphi(\bar{\boldsymbol{\xi}}, \bar{x}(\bar{\boldsymbol{\xi}}))]$$

is often considered as an approximation of $v(\boldsymbol{\xi})$. However, specific regularity assumptions on the problem (P) and the processes $\boldsymbol{\xi}$ and $\bar{\boldsymbol{\xi}}$ are necessary to ensure certain approximation qualities, cf. [5], [11], and [13]. Indeed, without such conditions $\bar{\boldsymbol{\xi}}$ may be close to $\boldsymbol{\xi}$ in some sense, but passing from (P) to $(\bar{P})$ may lead to significant changes in the optimal value, e.g., by providing *arbitrage possibilities*, see [11, Example A.4].

Being interested in a good approximation of the unknown value $v(\boldsymbol{\xi})$, it is thus reasonable rather to evaluate the approximate solution $\bar{x}$ with regard to the *original* data process $\boldsymbol{\xi}$, that is, to consider

$$\mathbb{E}[\varphi(\boldsymbol{\xi}, \bar{x}(\boldsymbol{\xi}))].$$

Furthermore, whenever the approximate solution $\bar{x}$ is not feasible (or even not defined) along the initial process $\boldsymbol{\xi}$, it may be appropriate to modify $\bar{x}$ to a feasible strategy $\tilde{x}$. Then the value

$$(1) \qquad\qquad\qquad \mathbb{E}[\varphi(\boldsymbol{\xi}, \tilde{x}(\boldsymbol{\xi}))]$$

provides an upper bound on $v(\boldsymbol{\xi})$ that can be realized by implementing the strategy $\tilde{x}$. The value (1) thus appears to be a more reliable approximation of $v(\boldsymbol{\xi})$ than $\mathbb{E}[\varphi(\boldsymbol{\xi}, \bar{x}(\boldsymbol{\xi}))]$.

To evaluate the integral (1), the law of large numbers suggests to draw independent samples ξ^i, $i \in I$, from the distribution of $\boldsymbol{\xi}$ and to consider the *out-of-sample value*

$$(2) \qquad \tilde{v}(\bar{\boldsymbol{\xi}}) := \frac{1}{|I|} \sum_{i \in I} \varphi(\xi^i, \tilde{x}(\xi^i)).$$

The value $\tilde{v}(\bar{\boldsymbol{\xi}})$ can be seen as the *real-world performance* of the approximative solution $\bar{x}(\cdot)$. Consequently, approximations $\bar{\boldsymbol{\xi}}'$ and $\bar{\boldsymbol{\xi}}''$ of different accuracy or constructed by different algorithms may be compared by means of their out-of-sample values $\tilde{v}(\bar{\boldsymbol{\xi}}')$ and $\tilde{v}(\bar{\boldsymbol{\xi}}'')$. Similarly, approximative solution algorithms can be compared by evaluating the resulting solutions.

3   Out-of-sample Evaluation

We consider a solution $\bar{x}(\cdot)$ to the approximate problem $(\bar{P})$ and denote the finite set of scenarios of the approximating process $\bar{\boldsymbol{\xi}}$ by $\{\bar{\xi}^j : j \in J\}$. Starting from $\bar{x}(\cdot)$, we aim to construct a strategy $\tilde{x}(\cdot)$ that is feasible along a set of out-of-sample scenarios $\{\xi^i, i \in I\} \subset \operatorname{supp} \mathbb{P}_{\boldsymbol{\xi}}$. In order to ensure that $\tilde{x}(\cdot)$ is implementable by a non-clairvoyant decision maker, this *feasibility restoration* has to be nonanticipative.

To this end, we consider a *nonanticipative* mapping $\pi : \{\xi^i : i \in I\} \to \{\bar{\xi}^j : j \in J\}$ that assigns every out-of-sample scenario ξ^i to some scenario of the approximated process that is close to ξ^i, in some sense. We say that π is nonanticipative if it can be written as $\pi(\xi^i) = (\pi_1(\xi^i_{[1]}), \ldots, \pi_T(\xi^i_{[T]}))$, where π_t are Borel measurable mappings from $\mathbb{R}^{s \cdot t} \to \{\bar{\xi}^j_t : j \in J\}$. Assuming a decision maker who has observed $\{\boldsymbol{\xi}_{[t]} = \xi^i_{[t]}\}$ until time t, the rule π suggests him a scenario $(\pi_1(\xi^i_{[1]}), \ldots, \pi_t(\xi^i_{[t]}))$ of the approximate model (and the corresponding strategy) that is close to his observation. The mapping π may be defined as a (conditional) projection, see the Appendix for a detailed construction.

The distance between the set of out-of-sample scenarios $\{\xi^i : i \in I\}$ and their associated tree scenarios $\pi(\xi^i)$ can be measured by the term

$$d_\pi(I, J) := \frac{1}{|I|} \sum_{i \in I} \frac{\sum_{t=1}^T \|\xi^i_t - \pi_t(\xi^i_{[t]})\|}{\sum_{t=1}^T \|\xi^i_t\|},$$

which is the relative euclidean distance between a scenario ξ^i and its assigned scenario $\pi(\xi^i)$, averaged over all scenarios $i \in I$.

Having related the out-of-sample scenarios ξ^i to the approximation scenarios $\bar{\xi}^j$ by the mapping π, we obtain that $\bar{x}(\pi(\cdot)) \in \mathcal{M}_m$, i.e., $\bar{x}(\pi(\boldsymbol{\xi}))$ is nonanticipative w.r.t. the process $\boldsymbol{\xi}$ and thus indeed a potential solution to the initial problem (P). Unfortunately, $\bar{x}(\pi(\xi^i))$

does not need to be *feasible* along the scenario ξ^i of the initial process $\boldsymbol{\xi}$, in general. In order to achieve this feasibility, different projection-based approaches to modify $\bar{x}(\pi(\cdot))$ are proposed in the next section.

In the following, we denote the decision $\bar{x}(\cdot)$ along the scenario $\pi(\xi^i)$ by $\bar{x}^i$ and refer to it as the *reference solution*. The modification of $\bar{x}^i$ along the out-of-sample scenario ξ^i is denoted by $\tilde{x}^i$.

3.1 Feasibility Restoration

Aiming for a (nonanticipative) solution $\tilde{x}^i$ that is feasible along the scenario ξ^i, we propose the following straightforward approach. Let $\tilde{x}_1^i := \bar{x}_1^i$. For $t = 2, \ldots, T$ and given $\tilde{x}_{t-1}^i$, we search for a feasible point $\tilde{x}_t^i$ that is *close to* $\bar{x}_t^i$.

Such a point $\tilde{x}_t^i$ may be found by projecting $\bar{x}_t^i$ on the feasible set at timestage t, i.e., on the set $\{x_t \in \mathbb{R}^m : A_{t,0} x_t + A_{t,1} \tilde{x}_{t-1}^i = h_t(\xi^i), \ x_t \geq 0\}$. However, in order to cope with possible future infeasibilities in models without relatively complete recourse, we further restrict the feasible set by incorporating information about future constraints. More precisely, we consider the value

$$
\begin{aligned}
(3) \qquad \Delta_t^i := \ &\min_{x_t} \|x_t - \bar{x}_t^i\|_\infty \\
&\text{s.t. } A_{t,0}\, x_t + A_{t,1}\, \tilde{x}_{t-1}^i = h_t(\xi_t^i), \\
&\qquad h_\tau^{\mathrm{low}} \leq A_{\tau,0}\, x_\tau + A_{\tau,1}\, x_{\tau-1} \leq h_\tau^{\mathrm{up}}, \qquad\qquad \tau = t+1, \ldots, T, \\
&\qquad x_\tau \geq 0, \qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \tau = t, \ldots, T,
\end{aligned}
$$

being the minimal distance from $\bar{x}_t^i$ onto the (reduced) feasible set at timestage t. The vectors h_τ^{low} and h_τ^{up} are chosen such that $h_\tau^{\mathrm{low}} \leq h_\tau(\xi_\tau^i) \leq h_\tau^{\mathrm{up}}$ holds true for all $i \in I$. The corresponding conditions in (3) are added to avoid decisions $\tilde{x}_t^i$ that may lead to future infeasibilities, at least to some degree. In particular, we set $h_{\tau,j}^{\mathrm{low}} = h_{\tau,j}^{\mathrm{up}}$ for those components j of $h_\tau(\cdot)$ that do not depend on ξ. Observe that this simple approach to avoid future infeasibilities relies on the assumption of non-random matrices $A_{\tau,0}$ and $A_{\tau,1}$. However, the approach can be extended, e.g., by demanding the existence of feasible decisions $x_\tau, \tau \geq t$, along all possible future realizations of the process $\bar{\boldsymbol{\xi}}$.

Basic Restoration

One may think about several techniques for determining a feasible point $\tilde{x}_t^i$ based on previously computed values for $\tilde{x}_1^i, \ldots, \tilde{x}_{t-1}^i$. A *basic* method is to just stay as close as

possible to the reference solution $\bar{x}_t^i$ and to set $\tilde{x}_t^i$ to a solution point of problem (3):

$$\tilde{x}_t^i := \operatorname*{argmin}_{x_t} \|x_t - \bar{x}_t^i\|_\infty$$

$$\text{s.t. } A_{t,0}\, x_t + A_{t,1}\, \tilde{x}_{t-1}^i = h_t(\xi_t^i),$$

$$h_\tau^{\text{low}} \leq A_{\tau,0}\, x_\tau + A_{\tau,1}\, x_{\tau-1} \leq h_\tau^{\text{up}}, \qquad \tau = t+1, \ldots, T,$$

$$x_\tau \geq 0, \qquad\qquad\qquad\qquad \tau = t, \ldots, T.$$

Myopic Restoration

However, sometimes it may be reasonable to exchange some closeness to $\bar{x}_t^i$ by cost minimality along the out-of-sample scenario ξ^i. That is, we allow the decision $\tilde{x}_t^i$ to deviate from the set of closest feasible solutions by a relative fraction $\varepsilon_{\text{rel}} \geq 0$ or an absolute value $\varepsilon_{\text{abs}} \geq 0$ in order to minimizing the costs along ξ^i. Doing so in a *myopic* way means to minimize $\langle c_t(\xi^i), x_t \rangle$, i.e.,

$$\tilde{x}_t^i := \operatorname*{argmin}_{x_t} \langle c_t(\xi^i), x_t \rangle + \rho_t \|x_t - \bar{x}_t^i\|_\infty$$

$$\text{s.t. } A_{t,0}\, x_t + A_{t,1}\, \tilde{x}_{t-1}^i = h_t(\xi_t^i),$$

$$h_\tau^{\text{low}} \leq A_{\tau,0}\, x_\tau + A_{\tau,1}\, x_{\tau-1} \leq h_\tau^{\text{up}}, \qquad \tau = t+1, \ldots, T,$$

$$x_\tau \geq 0, \qquad\qquad\qquad\qquad \tau = t, \ldots, T,$$

$$\|x_t - \bar{x}_t^i\|_\infty \leq (1 + \varepsilon_{\text{rel}})\Delta_t^i + \varepsilon_{\text{abs}}.$$

where $\rho_t \geq 0$ is chosen very small, e.g., $\rho_t = 10^{-4}\|c_t(\xi^i)\|_\infty$.

Note that letting $\rho_t \to 0$ and $\varepsilon_{rel} + \varepsilon_{abs} \to \infty$ increases the emphasis on cost minimality along ξ^i by allowing larger deviation from the reference solution $\bar{x}_t^i$.

Farsighted Restoration

Due to the time-coupling constraints, a decision x_t at time t impacts the feasible sets for future decisions and thus the future costs. These future costs can be taken into account within the feasibility restoration by considering the *shadow prices* associated to the time-coupling constraints. For this purpose, we consider the dual problem to $(\bar{P})$,

$$(\bar{D}) \quad \max \left\{ \mathbb{E}\left[\sum_{t=1}^T \langle h_t(\bar{\boldsymbol{\xi}}_t), \mu_t \rangle\right] : \begin{array}{l} \mu \in \mathcal{M}_r, \quad \mu_1 A_{1,1} \leq c_1, \quad \mu_T(\bar{\boldsymbol{\xi}}_{[T]})A_{T,0} \leq c_T(\bar{\boldsymbol{\xi}}_T), \\ \mu_t(\bar{\boldsymbol{\xi}}_{[t]})A_{t,0} + \mathbb{E}\left[\mu_{t+1}(\bar{\boldsymbol{\xi}}_{[t+1]})A_{t+1,1}|\bar{\boldsymbol{\xi}}_{[t]}\right] \leq c_t(\bar{\boldsymbol{\xi}}_t), \\ \hspace{7cm} t = 2, .., T-1. \end{array} \right\},$$

and denote by $\bar{\mu}$ an optimal solution of $(\bar{D})$. The shadow price vector corresponding to the primal decision x_t is then equal to $\bar{\mu}_{t+1}(\bar{\boldsymbol{\xi}}_{[t+1]})A_{t+1,1}$. In particular, this value is a subgradient of the *cost-to-go function* at time $t+1$, that will be defined below. In order

to maintain the nonanticipativity of the feasibility restoration along the out-of-sample scenario ξ^i, we resort to the *expected* shadow price vector

$$\eta_{t+1} := \mathbb{E}\left[\mu_{t+1}(\bar{\boldsymbol{\xi}}_{[t+1]})A_{t+1,1} \,\big|\, \bar{\boldsymbol{\xi}}_{[t]} = \pi_t(\xi^i_{[t]})\right].$$

Using this dual information about future costs, we choose $\tilde{x}^i_t$ in a *farsighted* way by setting

$$\tilde{x}^i_t := \underset{x_t}{\mathrm{argmin}} \; \langle c_t(\xi^i) - \eta_{t+1}, x_t \rangle + \rho_t \|x_t - \bar{x}^i_t\|_\infty$$

$$\text{s.t. } A_{t,0}\, x_t + A_{t,1}\, \tilde{x}^i_{t-1} = h_t(\xi^i_t),$$
$$h^{\mathrm{low}}_\tau \le A_{\tau,0}\, x_\tau + A_{\tau,1}\, x_{\tau-1} \le h^{\mathrm{up}}_\tau, \qquad\qquad \tau = t+1,\dots,T,$$
$$x_\tau \ge 0, \qquad\qquad\qquad\qquad\qquad \tau = t,\dots,T,$$
$$\|x_t - \bar{x}^i_t\|_\infty \le (1 + \varepsilon_{\mathrm{rel}})\Delta^i_t + \varepsilon_{\mathrm{abs}}.$$

Extensive Restoration

An even more farsighted method is to use not only a single subgradient, but several. Such subgradients are available, e.g., whenever the approximate problem $(\bar{P})$ was solved by a Nested-Benders-style algorithm [17]. Thus, for $t = 1,\dots,T$, let

$$Q_t(x_{t-1}, \xi_{[t]}) := \min_{x_t} \; \langle c_t(\xi_t), x_t \rangle + \mathbb{E}\left[Q_{t+1}(x_t, \bar{\boldsymbol{\xi}}_{[t+1]}) \,\big|\, \bar{\boldsymbol{\xi}}_{[t]} = \xi_{[t]}\right]$$

$$\text{s.t. } A_{t,0}x_t + A_{t,1}x_{t-1} = h_t(\xi_t), \quad x_t \ge 0,$$

be the cost-to-go function at time t in state $\xi_{[t]}$ (with $Q_{T+1}(\cdot,\cdot) := 0$). Having subgradient information π'_t about $\mathbb{E}\left[Q_{t+1}(\cdot,\bar{\boldsymbol{\xi}}_{[t+1]}) \,\big|\, \bar{\boldsymbol{\xi}}_{[t]} = \xi_{[t]}\right]$ available in a set of supporting points x'_t, we can set

$$\tilde{x}^i_t := \underset{x_t}{\mathrm{argmin}} \; \langle c_t(\xi^i), x_t \rangle + \Theta_t + \rho_t \|x_t - \bar{x}^i_t\|_\infty$$

$$\text{s.t. } A_{t,0}\, x_t + A_{t,1}\, \tilde{x}^i_{t-1} = h_t(\xi^i_t),$$
$$h^{\mathrm{low}}_\tau \le A_{\tau,0}\, x_\tau + A_{\tau,1}\, x_{\tau-1} \le h^{\mathrm{up}}_\tau, \qquad\qquad\qquad \tau = t+1,\dots,T,$$
$$x_\tau \ge 0, \qquad\qquad\qquad\qquad\qquad\qquad \tau = t,\dots,T,$$
$$\Theta_t \ge \mathbb{E}\left[Q_{t+1}(x'_t, \bar{\boldsymbol{\xi}}_{[t+1]}) \,\big|\, \bar{\boldsymbol{\xi}}_{[t]} = \pi_t(\xi^i_{[t]})\right] + \langle \pi'_t, x_t - x'_t \rangle, \qquad \text{for all } x'_t,$$
$$\|x_t - \bar{x}^i_t\|_\infty \le (1 + \varepsilon_{\mathrm{rel}})\Delta^i_t + \varepsilon_{\mathrm{abs}}.$$

Remark 1. *The latter 'extensive' method is related to the first approach in [2], which has been proposed for MSPs with interstage independence or a weak type of interstage dependence.*

Remark 2. *It is also possible to apply a preprocessing step to the introduced restoration methods similar to the optimal basis prolongation in [1]. That is, having an optimal basis from the solution of $(\bar{P})$ and a feasible solution $\tilde{x}^i_{t-1}$ at hand, one can construct a*

corresponding primal solution $\bar{x}_t^i$ that satisfies $A_{t,0}\bar{x}_t^i + A_{t,1}\tilde{x}_{t-1}^i = h_t(\xi_t^i)$, but might violate the positivity constraint $x_t \geq 0$. Thus, a feasibility restoration step that uses $\bar{x}_t^i$ instead of $\tilde{x}_t^i$ is applied afterwards. However, within our numerical experiments this preprocessing affects the out-of-sample evaluation rather adversely.

3.2 Measuring Infeasibility

Without relatively complete recourse, the feasibility restoration might fail if problem (3) is infeasible. However, in some cases it might be possible to relax certain 'soft' constraints in order to obtain a feasible solution. Let S_t be a matrix that indicates the 'soft' dynamic constraints at timestage t (of course, the concrete choice of S_t depends on the considered model). In order to determine how much these constraints have to be relaxed to make (3) feasible, we solve the auxiliary problem

$$(4) \qquad \min_{x_t,\dots,x_T,\tilde{s}_t^i} |\tilde{s}_t^i|_1$$

$$\text{s.t. } A_{t,0}\, x_t + A_{t,1}\, \tilde{x}_{t-1}^i + S_t \tilde{s}_t^i = h_t(\xi_t^i).$$

$$h_\tau^{\text{low}} \leq A_{\tau,0}\, x_\tau + A_{\tau,1}\, x_{\tau-1} \leq h_\tau^{\text{up}}, \qquad\qquad \tau = t+1,\dots,T,$$

$$x_\tau \geq 0, \qquad\qquad \tau = t,\dots,T.$$

If problem (4) is feasible, we apply one of the feasibility restoration methods from above with $\varepsilon_{\text{rel}} = \varepsilon_{\text{abs}} = 0$ and the right side $h_t(\xi^i)$ replaced by $h_t(\xi^i) - \tilde{s}_t^i$, to obtain a 'minimal infeasible' solution $\tilde{x}_{t-1}^i$. With this solution at hand we can proceed to the next timestage. If the relaxed problem (4) is feasible for every $t = 2,\dots,T$, we say that the solution $\tilde{x}^i$ is *weakly infeasible*. The *minimal weak infeasibility of* $\tilde{x}^i$ along ξ^i is then defined as

$$(5) \qquad\qquad\qquad \tilde{s}^i := \sum_{t=1}^T |\tilde{s}_t^i|_1,$$

where we let $\tilde{s}_t^i = 0$ if (3) was feasible in timestage t.

If even the relaxed problem (4) is infeasible for some $t \geq 2$, $\tilde{x}^i$ is denoted as *strongly infeasible* and the feasibility restoration for the out-of-sample scenario ξ^i is abandoned.

Piecing this all together, it is reasonable to measure the quality of a solution by the following parameters. By I_+, I_-, and I_{--} we denote the sets of indices $i \in I$ such that $\tilde{x}^i$ is feasible, weakly infeasible, or strongly infeasible, respectively. Then we consider the

$$\text{(feasible) out-of-sample value } \tilde{v}(\bar{\boldsymbol{\xi}}) \;\; := \;\; \frac{1}{|I_+|} \sum_{i \in I_+} \varphi(\xi^i, \tilde{x}(\xi^i)),$$

$$\text{infeasibility rate } \tilde{r} \;\; := \;\; \frac{|I_- \cup I_{--}|}{|I|}, \text{ and}$$

$$\text{average (weak) infeasibility } \tilde{s} \;\; := \;\; \frac{1}{|I_-|} \sum_{i \in I_-} \tilde{s}^i,$$

where $\tilde{s}^i$ is defined by (5).

4 Numerical Examples

We apply the proposed out-of-sample evaluation method for two stochastic programming models using in both cases classical scenario trees as well as recombining scenario trees.

The classical scenario trees are generated using the *Forward Tree Construction Algorithm* of Heitsch and Römisch [4, Algorithm 4.5]. The optimal value $v(\bar{\boldsymbol{\xi}})$ of the scenario tree based problem is computed by solving the deterministic equivalent with CPLEX [9].

The recombining scenario trees are constructed by a modified version of [4, Alg. 4.5]. Recombination takes place every four time stages and the number of different subtrees per timeperiod (the time between two recombination points) is varied between two and eight. The tree based problem is solved by an extension of the Nested Benders Decomposition Algorithm as introduced in [12]. Note that the decomposition approach does not require the solution process to be recombining too. However, the recombining nature of the process ξ allows a "dynamic recombination" of solutions due to the re-use of Benders Cuts. Within this decomposition algorithm the quality of the approximated cost-to-go functions is controlled by an aggregation parameter $\gamma \in [0, 1]$. Numerical experiments with various values of γ have shown that the choice $\gamma = 0.01$ yields sufficiently accurate solutions. For details about the generation of recombining scenario trees and the decomposition approach, see also [12].

4.1 Power Scheduling

Model Description

We consider a power generating system consisting of several thermal units (index set I), a pumped hydro unit, and a wind power plant. The objective is to find cost-optimal operation levels of the thermal units and hydro units under uncertain production of electricity from wind. The model makes no claim to depict a real world situation, but is for studying purposes only.

Let us denote by $p_{i,t}$ the operation level of the thermal unit $i \in I$ at time t, by l_t the fill level of the water reservoir, and by w_t and v_t the operation level of the water pump and turbine, respectively. Deterministic parameters of the problem are operation ranges for the thermal units $\underline{p}_i < \bar{p}_i$, $i \in I$, the pump $\bar{w} > 0$, and the turbines $\bar{v} > 0$, the capacity of the water reservoirs $\bar{l} > 0$, the fill levels l_{in} and l_{end} of the reservoirs at the beginning and at the end of the considered time horizon, the efficiency of the pump η, the fuel costs b_i, $i \in I$, and the energy demand d_t. As stochastic parameter we consider the wind power

production κ_t. The complete model has the form

$$\min \quad \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i\in I}b_i p_{i,t}\right]$$

(6) $\qquad$ s.t. $\quad l_0 = l_{in}, \quad l_T \geq l_{end},$

(7) $\qquad\qquad l_t = l_{t-1} - (v_t - \eta w_t), \qquad\qquad\qquad t = 1,\ldots,T,$

(8) $\qquad\qquad |p_{i,t} - p_{i,t-1}| \leq \dfrac{1}{2}(\bar{p}_i - \underline{p}_i), \qquad\qquad i \in I,\, t = 2,\ldots,T,$

(9) $\qquad\qquad \displaystyle\sum_{i\in I} p_{i,t} + (v_t - w_t) + \kappa_t \geq d_t, \qquad\qquad t = 1,\ldots,T,$

(10) $\qquad\qquad \displaystyle\sum_{i\in I} p_{i,t} \leq \sum_{i\in I}\bar{p}_i - \dfrac{1}{10}d_t, \qquad\qquad t = 1,\ldots,T,$

$$\underline{p}_i \leq p_{i,t} \leq \bar{p}_i,\ 0 \leq v_t \leq \bar{v},\ 0 \leq w_t \leq \bar{w},\ 0 \leq l_t \leq \bar{l}, \quad i \in I,\, t = 1,\ldots,T.$$

Constraint (6) models the initial and final fill level of the reservoir, (7) couples the fill level of the reservoir between successive time stages, (8) bounds the change in the operation of the thermal units between successive time stages, (9) ensures that the electricity demand is covered, and (10) is a reserve requirement. The model parameters are given by Table 1 in the Appendix.

The time horizon is $T = 48$ hours. The random nature of future wind power output is modeled by adding a discrete time Brownian motion to the predicted wind speed curve. Then 1000 trajectories of wind speed are simulated and transformed into wind power output using a (piecewise linear) wind park *power curve*, c.f. Figure 1. The resulting wind power output scenarios are used for the construction of scenario trees. For the bounds on κ_t (as required for h_t^{low} and h_t^{up} in (3)) we use $\kappa_t^{\text{low}} \equiv 0$ and $\kappa_t^{\text{up}} \equiv \max_\tau \|\boldsymbol{\kappa}_\tau\|_\infty$ for $t = 1,\ldots,T$.
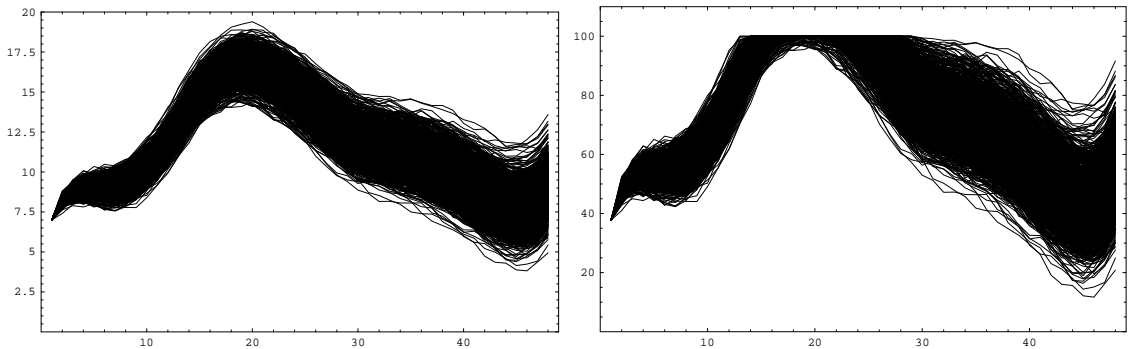

Figure 1: *Left side:* 1000 simulated trajectories from wind speed in m/s. *Right side:* The corresponding wind power output trajectories (percentage of maximal capacity). The maximal capacity of the wind power park is attained for a wind speed above 15 m/s.

While the capacity of the thermal units is sufficient to cover the maximal load, this

model does not possess relatively complete recourse due to the condition (6) on the minimal final fill level of the water storage and the reserve requirements (10). Since a violation of the reserve requirement (10) does not prohibit the power plant to operate inside its operational bounds and such that the demand is satisfied, we have chosen these constraints as soft constraints (cf. Section 3.1), i.e., weakly infeasible solutions are allowed to violate the reserve requirement.

Out-of-sample Evaluation

Figure 2 presents the results of the out-of-sample evaluation of classical and two types of recombining scenario trees using the farsighted feasibility restoration. As one can observe, the out-of-sample values $\tilde{v}(\bar{\boldsymbol{\xi}})$ are higher than the minimal costs $v(\bar{\boldsymbol{\xi}})$ of the tree based stochastic programs. While the optimal values $v(\bar{\boldsymbol{\xi}})$ of the approximate problems do not significantly differ for scenario trees having different numbers of nodes, the out-of-sample values are decreasing with a growing number of nodes in the same manner as the distance d_π between the scenario trees and the set of out-of-sample scenarios. This indicates that the quality of the tree based solutions with regard to their usefulness for real-world decision making improves with increasing accuracy of the scenario tree approximation.

Next, we observe that the out-of-sample values $\tilde{v}(\bar{\boldsymbol{\xi}})$ computed from solutions of *recombining tree* based stochastic programs are better than when classical scenario trees on the same number of nodes are used. This is probably due to the much higher number of scenarios that can be used within recombining trees. Comparing the out-of-sample values for recombining scenario trees with two and four different subtrees per timeperiod shows that a too extensive recombination (i.e., using only two subtrees) worsen the results, even though the number of nodes is the same.

Further, we compare the out-of-sample values that are obtained with varying values of $\varepsilon_{\mathrm{rel}}$, i.e., allowing the modified solution to differ more or less from the tree solution in favor of cost minimality. The results for the farsighted feasibility restoration are shown on the left side of Figure 2. The choice $\varepsilon_{\mathrm{rel}} = 0$ is equivalent to the basic method. As we can see, the out-of-sample values decrease with increasing values of $\varepsilon_{\mathrm{rel}}$, i.e., with an increasing freedom to move solutions towards optimality per timestage. However, we also observed that the number of out-of-sample scenarios that lead to an infeasible problem (3) in the feasibility restoration phase increases. While for $\varepsilon_{\mathrm{rel}} \leq 0.1$ the percentage of infeasible scenarios $\tilde{r}$ (c.f. Section 3.2) is at most 0.1% for all trees, this value increased to 0.4% for $\varepsilon_{\mathrm{rel}} = 0.5$. For $\varepsilon_{\mathrm{rel}} = 1$ the fraction of infeasible scenarios raised to 5%, where the infeasibilities resulting from recombining scenario tree solutions were always of the weak type. Both types of infeasibilities could be observed when using classical scenario trees.

We also performed computations with the myopic feasibility restoration and observed that the out-of-sample values are only slightly worse than with the farsighted approach.
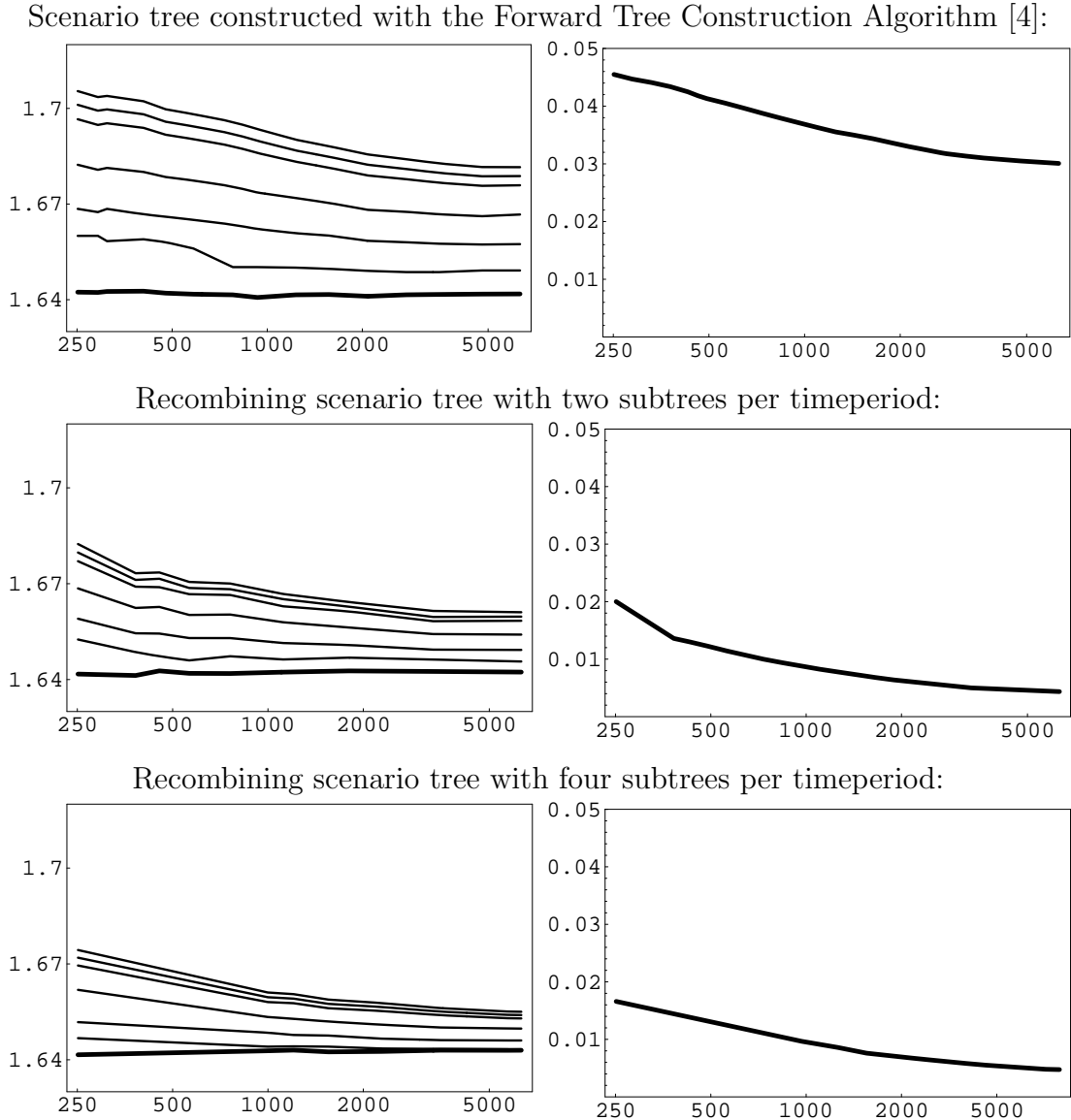
Figure 2: Results for the power scheduling model. *Left side:* Optimal value $v(\bar{\boldsymbol{\xi}})$ of the tree based problems (in bold) and corresponding out-of-sample values $\tilde{v}(\bar{\boldsymbol{\xi}})$ for $\varepsilon_{\text{abs}} = 0$ and several values of ε_{rel} using a series of scenario trees with increasing number of nodes (horizontal axis). The values for ε_{rel} are (with decreasing position of the $\tilde{v}(\bar{\boldsymbol{\xi}})$ curve) 0, 0.05, 0.10, 0.25, 0.5, and 1.0, where the farsighted feasibility restoration was used. *Right side:* The distance d_{π} between the scenario trees and the set of out-of-sample scenarios.

Hence, the additional information on (future) costs related to changes in the fill level of the water reservoir does not change the out-of-sample values considerably.

4.2 Swing Option Exercising

Model Description

The following problem of evaluating a swing option basically coincides with Problem 1 considered in [7]. A swing option allows its holder to purchase a total amount of U units of energy during the time $[1, T]$, for a fixed strike price K per unit. The amount of energy purchased at time $t \in [1, T]$ is denoted by x_t and has to lie in some interval $[l, u]$. Assuming that the purchased energy is immediately sold on the spot market, the holder of the swing option is interested in finding a purchase strategy $x_t, t = 1, \ldots, T$, that maximizes the expected accumulated wealth. This problem can be written as follows:

$$
(11) \qquad \min \left\{ \mathbb{E} \left[\sum_{t=1}^{T} \langle K - \boldsymbol{\xi}_t, x_t \rangle \right] : \begin{array}{l} x \in \mathcal{M}_1, \quad x_t \in [l, u], \quad t = 1, \ldots, T, \\ \sum_{t=1}^{T} x_t \leq U, \end{array} \right\}
$$

with positive constants l, u, U, and K. The stochastic process $\boldsymbol{\xi} = (\boldsymbol{\xi}_t)_{t=1}^{T}$ describes the spot market price per energy unit and is assumed to follow a discrete time geometric Brownian motion, i.e., $\boldsymbol{\xi}_1 = 1$ and

$$
(12) \qquad \boldsymbol{\xi}_t = \boldsymbol{\xi}_{t-1} \exp(\varepsilon_t - \frac{1}{2}\sigma^2), \quad t = 2, \ldots, T.
$$

Thereby ε_t, $t = 2, \ldots, T$, are independent, normally distributed random variables with expectation μ and variance σ^2. In the following, we assume for the sake of notational simplicity $l = 0$, $u = 1$, and $U \in \mathbb{N}$.

In particular, this model allows for an analytic solution whenever the drift μ of the spot price process $\boldsymbol{\xi}$ is nonnegative. Indeed, using the (negative) payoff function $\varphi(\boldsymbol{\xi}_t) := (K - \boldsymbol{\xi}_t)^-$, problem (11) can be written as

$$
(13) \qquad \min \left\{ \mathbb{E} \left[\sum_{t=1}^{T} x_t \varphi(\boldsymbol{\xi}_t) \right] : \begin{array}{l} x_t \in \mathcal{M}_1, \quad x_t \in [0, 1], \quad t = 1, \ldots, T, \\ \sum_{t=1}^{T} x_t \leq U \end{array} \right\}.
$$

If $\mu \geq 0$, the process $\boldsymbol{\xi}$ is a submartingale. Consequently, due to the negativity, monotonicity, and concavity of φ, the process $(\varphi(\boldsymbol{\xi}_t))_{t=1,\ldots,T}$ is a supermartingale. It is thus no surprise that an early exercise of the the swing option is not optimal. Indeed, one easily shows that x^* defined by

$$
(14) \qquad x_t^*(\boldsymbol{\xi}_t) := \begin{cases} 0, & \text{if } t \leq T - U, \\ 1, & \text{if } t > T - U \text{ and } \boldsymbol{\xi}_t > K, \end{cases}
$$

is an optimal solution for (11). Observe that the algorithms we have used for the scenario tree construction do not maintain the martingale property, in general. Hence, the tree based optimal solutions are not necessarily of type (14).

Out-of-sample Evaluation

For the following computations, we used the parameter values $T = 52$, $U = 20$, $K = 1$, $\mu = 0$, and $\sigma = 0.07$. We reformulated the model by adding auxiliary variables $u_t = \sum_{\tau=1}^{t} x_t$ and the (redundant) constraints $u_t \leq U$, $t = 1, \ldots, T - 1$. Thus, by virtue of these modifications, the model has relatively complete recourse. This property and the small number of variables and constraints might let this model appear 'simple'. However, while in the power scheduling model uncertainty appeared only in one component of the power generating system, the decisions in the swing option model are mainly driven by the stochastic price process. Thus, one can expect that the quality of the price process approximations $\bar{\boldsymbol{\xi}}$ has a major impact on the optimal value of the stochastic program and the out-of-sample values. Indeed, Figure 3 shows that the relative deviation between the tree based optimal values $v(\bar{\boldsymbol{\xi}})$ and the out-of-sample values $\tilde{v}(\bar{\boldsymbol{\xi}})$ is much larger than in the power scheduling problem, see Figure 2.

For the considered classical scenario trees, the values $v(\bar{\boldsymbol{\xi}})$ and $\tilde{v}(\bar{\boldsymbol{\xi}})$ appear to be far from convergent. This illustrates that the number of scenarios used to approximate the price process over $T = 52$ timestages does not lead to a reliable solution yet. In contrast, scenario trees of similar size have shown to be adequate for the power scheduling model (48 timestages, cf. Section 4.1).

Using recombining scenario trees seems to slightly improve the situation. The optimal value $v(\bar{\boldsymbol{\xi}})$ and out-of-sample values $\tilde{v}(\bar{\boldsymbol{\xi}})$ are less fluctuating here. Again, we can observe that using only two subtrees per timeperiod is not sufficient and yields a too optimistic value $v(\bar{\boldsymbol{\xi}})$, while the use of eight scenario trees shifts both the tree value $v(\bar{\boldsymbol{\xi}})$ and the out-of-sample value $\tilde{v}(\bar{\boldsymbol{\xi}})$ closer to the exact value $v(\boldsymbol{\xi})$.

Next, we compare different feasibility restoration strategies. Since uncertainty appears only in the objective function coefficients, the scenario-wise solutions $\bar{x}^i$, $i \in I$, are always feasible for the corresponding scenario ξ^i, that is $\Delta_t^i = 0$. Hence, the basic feasibility restoration method coincides with evaluating the scenario solution $\bar{x}^i$ along the costs $c_t(\xi_t^i)$, $t = 1, \ldots, T$. We compare a basic, a myopic, and a farsighted strategy using $\varepsilon_{\text{abs}} = 1$. In difference to the power scheduling model, here the basic approach yields the best results. For the myopic feasibility restoration, the choice $\varepsilon_{\text{abs}} = 1$ yields a solution $\tilde{x}_t^i$ that is independent of the scenario tree solution $\bar{x}$. In particular, the out-of-sample values based on the myopic approach do not depend on the scenario tree. The farsighted feasibility restoration with $\varepsilon_{\text{abs}} = 1$ additionally utilizes in each timestage information from the scenario tree solution in form of shadow prices. As one can observe, this allows to improve the myopic results, even though the out-of-sample value gets worse with a growing number of nodes.

Scenario tree constructed with the Forward Tree Construction Algorithm [4]:

Recombining scenario tree with two subtrees per timeperiod:

Recombining scenario tree with eight subtrees per timeperiod:
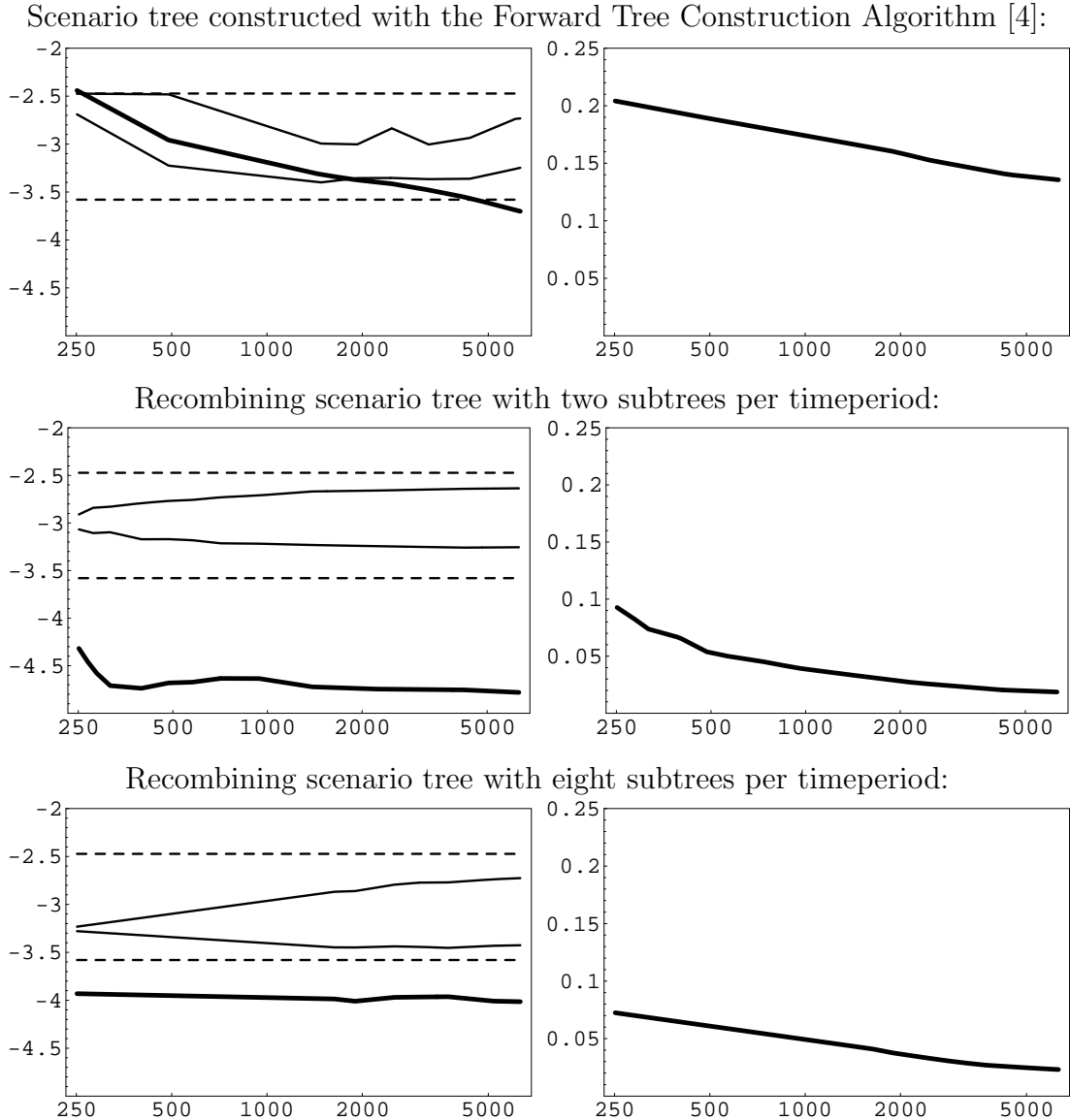

Figure 3: Results for the swing option model. *Left side:* Optimal value $v(\bar{\xi})$ of the stochastic program (in bold) and the corresponding out-of-sample values $\tilde{v}(\bar{\xi})$ for basic (lower solid line) and farsighted (upper solid line) feasibility restoration using a series of scenario trees with increasing number of nodes (horizontal axis). The myopic value and the exact value $v(\xi)$ are plotted as the upper and lower dashed lines, respectively. The parameters $\varepsilon_{\text{rel}} = 0$ and $\varepsilon_{\text{abs}} = 1$ were used. *Right side:* The distance d_π between the scenario trees and the set of out-of-sample scenarios.

Appendix

For the sake of completeness, we sketch how the mapping $\pi : \{\xi^i : i \in I\} \rightarrow \{\bar{\xi}^j : j \in J\}$, as introduced in Section 3, may be constructed. Given a scenario tree consisting of the

scenarios $\bar{\xi}^j$, $j \in J$, a *node* n_t at time t is a subset of J, such that the scenarios $\bar{\xi}^j$, $j \in n_t$ are indistinguishable until time t. The set of nodes at time t is denoted by $\mathcal{N}_t$. The set of nodes at time $t+1$ *succeeding* from the node n_t is denoted by $\text{succ}(n_t) \subset \mathcal{N}_{t+1}$ and we have $\cup_{n_{t+1} \in \text{succ}(n_t)} n_{t+1} = n_t$. The realization of the (tree) process $\bar{\boldsymbol{\xi}}$ on a node n_t is denoted by $\bar{\xi}_t^{n_t}$. Recall that there is only one node at time $t = 1$, i.e. $\mathcal{N}_1 = \{n_1\} = \{J\}$, and there are $|J|$ nodes at time horizon T, each of it containing a different singular scenario index.

Now, the mapping π is constructed recursively as follows. Given some out-of-sample scenario ξ^i, we set $\pi_1(\xi_{[1]}^i) := \bar{\xi}_1^{n_1}$. Assume that for some $t \in \{2, \ldots, T\}$ we have $\pi_{t-1}(\xi_{[t-1]}^i) = \bar{\xi}_{t-1}^{n_{t-1}}$ for some node n_{t-1}. Then we define

$$\pi_t(\xi_{[t]}^i) := \underset{\bar{\xi}^{n_t} : n_t \in \text{succ}(n_{t-1})}{\operatorname{argmin}} \|\xi_t^i - \bar{\xi}_t^{n_t}\|.$$

	parameter	value
b_1	fuel cost coal	21
b_2	fuel cost gas & steam	48
b_3	fuel cost gas	154
$\bar{p}_1$	capacity coal	1,000
$\bar{p}_2$	capacity gas & steam	500
$\bar{p}_3$	capacity gas	500
$\bar{k}$	capacity wind	1,000
$\bar{v}$	capacity hydro turbine	2,000
$\bar{w}$	capacity hydro pump	2,000
η	pump efficiency	0.75
$\bar{l}$	capacity hydro storage	12,000
l_{in}, l_{end}	initial/final storage level	6,000

Table 1: Parameter of power scheduling model. Maximum demand per hour is $2,000$.

References

[1] M.S. Casey and S. Sen. The scenario generation algorithm for multistage stochastic linear programming. *Mathematics of Operations Research*, 30:615–631, 2005.

[2] A. Chiralaksanakul and D. Morton. Assessing policy quality in multi-stage, stochastic programming. *Stochastic Programming E-Print Series*, 12, 2004.

[3] J. Dupačová, N. Gröwe-Kuska, and W. Römisch. Scenarios reduction in stochastic programming: An approach using probability metrics. *Mathematical Programming*, 95(A):493–511, 2003.

[4] H. Heitsch and W. Römisch. Scenario tree modeling for multistage stochastic programs. *Mathematical Programming*, 118:371–406, 2009.

[5] H. Heitsch, W. Römisch, and C. Strugarek. Stability of multistage stochastic programs. *SIAM Journal on Optimization*, 17:511–525, 2006.

[6] R. Henrion, C. Küchler, and W. Römisch. Scenario reduction in stochastic programming with respect to discrepancy distances. *Computational Optimization and Applications*, 43:67–93, 2009.

[7] P. Hilli and T. Pennanen. Numerical study of discretizations of multistage stochastic programs. *Kybernetika*, 44:185–204, 2008.

[8] K. Høyland and S.W. Wallace. Generating scenario trees for multistage decision problems. *Management Science*, 47(2):295–307, 2001.

[9] ILOG, Inc. (now IBM). CPLEX 10.0. `http://www.ilog.com/products/cplex`.

[10] M. Kaut and S.W. Wallace. Evaluation of scenario-generation methods for stochastic programming. *Pacific Journal of Optimization*, 3(2):257–271, 2007.

[11] C. Küchler. On stability of multistage stochastic programs. *SIAM Journal on Optimization*, 19(2):952–968, 2008.

[12] C. Küchler and S. Vigerske. Decomposition of multistage stochastic programs with recombining scenario trees. *Stochastic Programming E-Print Series*, 9, 2007. `http://www.speps.org`.

[13] R. Mirkov and G.Ch. Pflug. Tree approximations of dynamic stochastic programs. *SIAM Journal on Optimization*, 18(3):1082–1105, 2007.

[14] T. Pennanen. Epi-convergent discretizations of multistage stochastic programs. *Mathematics of Operations Research*, 30(1):245–256, 2005.

[15] G.Ch. Pflug. Scenario tree generation for multiperiod financial optimization by optimal discretization. *Mathematical Programming*, 89:251–271, 2001.

[16] W. Römisch. *Stability of Stochastic Programming Problems*, chapter 8, pages 483–554. Volume 10 of Ruszczyński and Shapiro [18], 2003.

[17] A. Ruszczyński. *Decomposition Methods*, chapter 3, pages 141–221. Volume 10 of Ruszczyński and Shapiro [18], 2003.

[18] A. Ruszczyński and A. Shapiro, editors. *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*. Elsevier, Amsterdam, 2003.

[19] A. Shapiro. *Monte Carlo Sampling Methods*, chapter 6, pages 353–425. Volume 10 of Ruszczyński and Shapiro [18], 2003.