Threshold accepting approach to improve bound-based approximations for portfolio optimization

Daniel Kuhn, Panos Parpas, and Berç Rustem

Department of Computing

Imperial College of Science, Technology, and Medicine

180 Queen's Gate, London SW7 2BZ, UK.

8th July 2007

Abstract. We discuss a discretization scheme for a portfolio selection problem. The model is a benchmark relative, mean-variance optimization problem in continuous time. In order to make the model computationally tractable, we discretize it in time and space. Our approximation scheme is designed in such a way so that the optimal values of the approximate problems yield bounds on the optimal value of the original problem. We discuss convergence of the bounds as the granulariy of the dicsretization is increased. We also propose a threshold accepting algorithm that attempts to find the most accurate discretization among all discretizations of a given complexity, and we provide promising results of a numerical case study.

Key words. portfolio optimization, stochastic programming, time discretization, bounds, threshold accepting

1 Introduction

Portfolio problems are prototypical examples of optimization problems under uncertainty. Finance researchers and investment practitioners remain fascinated by numerous questions around the inherent risk-return tradeoff, which implies that invested money can yield higher profits only if it is subject to the possibility

of being lost. The modelling of financial risks and investor preferences as well as the solution of large-scale portfolio optimization problems are challenging tasks. A principal thesis of this article is the consensus that realistic portfolio problems must be cast in a multi-period setting and can only be tackled computationally. In fact, investment decisions are coupled in time through serial dependencies of returns or consumption needs and, to a certain extent, through transaction costs, administrative fees, and taxes. Analytical solutions of such time-coupled models are usually not available. Numerical solutions, on the other hand, always rely on suitable approximations. In this paper we pursue the following strategy. Starting from a continuous-time continuous-state model, we discretize both the underlying planning horizon as well as the probability space of the random asset returns. This joint time and space discretization allows us to approximate the original investment problem by a sequence of multistage stochastic programs defined on finite scenario trees. The trees are designed in such a way that the solutions of the corresponding multistage stochastic programs provide bounds on the solution of the original problem. The effectiveness of discretization schemes depends on our ability to solve the resulting deterministic equivalent problem. Clearly as the granularity of the descritization increases, an appropriate scheme should lead to a reduction in the error. From a practical point of view we cannot solve problems with a large number of stages, and with a large branching factor in each stage. For this reason we attempt to optimize the tree generation using a heuristic algorithm. The heuristic we use is based on a variant of simulated annealing called threshold accepting [3]. The basic idea is for a given complexity of the tree to perform a random walk on all the possible configurations using the heuristic. For our purposes, the complexity of the tree is defined as the number of nodes on the tree. This simple procedure allows us to perform both time and space discretization simultaneously. The advantage of threshold accepting is that it requires less tuning of its parameters in order for it compute an acceptable solution.

In the present paper we generalize the bounding approximation scheme proposed in [17,18] to stochastic programs with random technology matrices. Moreover, we develop a modelling framework in which the tradeoff between the granularity of the time and space approximations can be analyzed. Theoretical convergence results guarantee asymptotic consistency of our multistage stochastic programs as the cardinality of the underlying scenario trees increases. In this work we go one step further and devise an efficient strategy to control the approximation error in situations when computational resources such as storage space and CPU power are finite. One of the contributions of this paper is the use of deterministic bounds with a stochastic search for the optimal tree. We empirically show that attempting to compute optimal trees can yield a significant reduction to the error for trees of comparable complexity. We believe that the combination of deterministic and stochastic schemes is not well explored in stochastic programming. In this paper we propose a method to combine these two paradigms.

Since the pioneering work of Markowitz [20], literature on portfolio theory has grown dramatically. An overview of influential contributions up to 2001 is provided in [24]. Much of the earlier literature focuses on investment situations in a single period setting. Dynamic portfolio optimization in continuous time was popularized by Merton [21]. While most dynamic models maximize expected utility of terminal wealth and or intertemporal consumption, the traditional mean-variance approach due to Markowitz has received relatively little attention in the multi-period situation. A stochastic programming based approach to multi-period mean-variance analysis was proposed by Frauendorfer [8, 10]. Gülpınar *et. al.* [13, 14] refined this model and tested its performance for different scenario trees. Furthermore, they reformulated the model as a stochastic min-max problem, assuming that some parameters of the underlying return distributions are ambiguous [12]. Recently, dynamic mean-variance portfolio problems without transaction costs have also been studied in a continuous-time framework. Analyt-

ical solutions were first obtained by Zhou and Li [26] who employ an embedding technique and certain concepts of stochastic linear-quadratic control theory.

In the presence of market frictions, portfolio constraints, and or parameter instability etc., the analytical treatment of dynamic mean-variance portfolio problems is out of the question, and one has to resort to algorithmic solution procedures. Below, we will use stochastic programming techniques [2, 16] to solve the portfolio problems under consideration. Suitable approximation schemes are necessary to keep the corresponding scenario trees of tractable size. In order to find solutions with provable performance guarantee, we use bounding methods to discretize the planning horizon and the probability space of the asset returns. Stage-aggregation and time-discretization methods with deterministic error bounds have previously been discussed in [17, 25]. Bounding methods for scenario generation have first been proposed by Madansky [19] and — in a game-theoretic setting — by Dupačovà [4]. Extensions to more general problem classes are due to Edirisinghe and Ziemba [5, 6], Frauendorfer [7, 9], and Kuhn [17, 18].

The general outline of this article is as follows. In Section 2 we formulate a mean-variance portfolio selection problem in continuous time with transaction costs and portfolio constraints. In order to achieve computational tractability, we develop a scenario tree based approximation for this problem. We proceed in two steps: Section 3 discusses discretization of the underlying planning horizon, while Sections 4 and 5 address discretization of the underlying probability space. In both cases we carefully estimate the approximation error. The time discretization scheme of Section 3 is tailored to the portfolio problem under consideration. In contrast, the space discretization scheme of Section 5 applies to a broad class of stochastic programs detailed in Section 4. We propose a randomized threshold accepting algorithm to minimize the approximation error over all scenario trees of a given cardinality. Section 6 reports on numerical results.

2 Portfolio problem

All random objects appearing in this article are defined on a complete probability space (Ω, Σ, P). By convention, random objects (i.e., random variables, random vectors, or stochastic processes) appear in boldface, while their realizations are denoted by the same symbols in normal face. Throughout our exposition, we consider a market of N investment opportunities or assets. We assume that the price $\boldsymbol{p}_n$ of asset n is driven by a geometric Brownian motion, that is,

$$\frac{\mathrm{d}\boldsymbol{p}_n(t)}{\boldsymbol{p}_n(t)} = \mu_n \, \mathrm{d}t + \mathrm{d}\boldsymbol{z}_n(t) \quad \text{for} \quad n = 1, \ldots, N \,,$$

where the $\boldsymbol{z}_n$ are correlated Wiener processes whose variance rates may differ from 1. Concretely speaking, we assume that

$$\mathrm{cov}(\mathrm{d}\boldsymbol{z}_m, \mathrm{d}\boldsymbol{z}_n) = \mathrm{E}(\mathrm{d}\boldsymbol{z}_m \, \mathrm{d}\boldsymbol{z}_n) = \sigma_{mn} \, \mathrm{d}t \,.$$

The limitation to geometric Brownian motions with a time-invariant correlation structure is merely for expositional convenience. Note that our approach easily extends to more general stochastic models. We denote by $\mathbb{A} := \{\mathcal{A}^t\}_{t \geq 0}$ the filtration generated by the asset prices. The underlying σ-fields are thus given by

$$\mathcal{A}^t := \sigma(\boldsymbol{p}_n(s) \,|\, s \in [0,t], \ n = 1, \ldots, N) \,.$$

It is convenient to interpret $\mathcal{A}^t$ as the time-t information set of an investor who continuously observes the asset prices. Let us assume that this investor holds a dynamically rebalanced portfolio of the assets in our model economy. This means that for each n, the investor continuously buys and sells assets of type n, pursuing the general goal of accumulating wealth over time. We denote by $\boldsymbol{b}_n$ and $\boldsymbol{s}_n$ the cumulative purchase and sale processes, respectively. To be precise, we assume that $\boldsymbol{b}_n$ and $\boldsymbol{s}_n$ are elements of $C(\mathbb{A})$, the convex cone of right-continuous non-decreasing stochastic processes adapted to $\mathbb{A}$ that vanish at $t = 0$. Thus, $\boldsymbol{b}_n(t)$ stands for the amount of money used up to time t to buy assets of type

n, while $\boldsymbol{s}_n(t)$ represents the amount of money obtained up to time t from sales of asset n. Moreover, we let $\boldsymbol{w}_n$ be the portfolio weight process of asset n, i.e., $\boldsymbol{w}_n(t)$ denotes the amount of money invested in asset n at time t. The processes $\boldsymbol{w}_n$, $\boldsymbol{b}_n$, and $\boldsymbol{s}_n$ are related through the asset balance equations

$$\mathrm{d}\boldsymbol{w}_n(t) = \boldsymbol{w}_n(t)\,\frac{\mathrm{d}\boldsymbol{p}_n(t)}{\boldsymbol{p}_n(t)} + \mathrm{d}\boldsymbol{b}_n(t) - \mathrm{d}\boldsymbol{s}_n(t)\,. \tag{2.1}$$

The total wealth process $\boldsymbol{\pi}$ is naturally defined via $\boldsymbol{\pi}(t) = \sum_{n=1}^{N} \boldsymbol{w}_n(t)$. Throughout this article we assume that proportional transaction costs are incurred whenever shares of the assets are traded. We denote by c_b and c_s the transaction costs per dollar of shares bought and sold, respectively. The change of total wealth over a time interval of length $\mathrm{d}t$ is thus given by capital gains minus transaction costs, i.e.,

$$\mathrm{d}\boldsymbol{\pi}(t) = \sum_{n=1}^{N} \boldsymbol{w}_n(t)\,\frac{\mathrm{d}\boldsymbol{p}_n(t)}{\boldsymbol{p}_n(t)} - c_b\,\mathrm{d}\boldsymbol{b}_n(t) - c_s\,\mathrm{d}\boldsymbol{s}_n(t)\,. \tag{2.2}$$

Plugging the definition of $\boldsymbol{\pi}$ into (2.2) and using (2.1) we obtain

$$0 = \sum_{n=1}^{N} (1 + c_b)\,\mathrm{d}\boldsymbol{b}_n(t) - (1 - c_s)\,\mathrm{d}\boldsymbol{s}_n(t)\,. \tag{2.3}$$

The equations (2.1) and (2.3) are sufficient to describe the portfolio dynamics. Hence, the SDE (2.2) is not needed, and the process $\boldsymbol{\pi}$ can principally be eliminated. We will always impose the restrictions $\boldsymbol{w}_n(t) \geq 0$, which preclude the possibility to take short positions in the assets. These no short sales restrictions further guarantee that total wealth will always remain positive.

The performance of an investment portfolio of the above kind is usually not measured in absolute terms but rather relative to a benchmark portfolio. A natural choice for a benchmark is the market portfolio $\boldsymbol{\pi}^*(t) := \sum_{n=1}^{N} \boldsymbol{w}_n^*(t)$ with normalized initial value. By definition, the monetary portfolio weight of asset n in the market portfolio is given by $\boldsymbol{w}_n^*(t) := \boldsymbol{w}_n(0)\,\boldsymbol{p}_n(t)/\boldsymbol{p}_n(0)$. We will assume that the risk associated with a portfolio process $\boldsymbol{\pi}$, as perceived by our investor,

amounts to

$$\int_0^T \mathrm{Var}\left(\boldsymbol{\pi}(t) - \boldsymbol{\pi}^*(t)\right)\, \alpha(\mathrm{d}t)\,. \tag{2.4}$$

The involved probability measure $\alpha : \mathcal{B}([0,T]) \to [0,1]$ characterizes the temporal risk weighting.[1] Portfolio performance, on the other hand, is measured by expected terminal wealth. As in the celebrated single-period Markowitz model [20], our investor faces a multi-objective decision problem. The conflicting objectives are to minimize risk while maximizing expected terminal wealth. Such problems are usually addressed by determining the set of Pareto optimal solutions, which in the current setting represent a family of *efficient portfolios*. We can find all efficient portfolios by minimizing risk under a parametric constraint of the form

$$\mathrm{E}\left(\boldsymbol{\pi}(T)\right) \geq e^{\varrho\, T}\, \mathrm{E}(\boldsymbol{\pi}^*(T))\,, \tag{2.5}$$

where the performance parameter ϱ is swept within a suitable interval. The investor will then pick the ρ corresponding to that efficient portfolio which is best suited to his or her individual risk preferences. From the above discussion we conclude that the investment strategy associated with a particular efficient portfolio can be found by solving the following stochastic optimization problem.

$$\begin{aligned}
&\underset{\boldsymbol{b}_n, \boldsymbol{s}_n \in C(\mathbb{A})}{\text{minimize}} \int_0^T \mathrm{Var}\left(\sum_{n=1}^N \boldsymbol{w}_n(t) - \boldsymbol{w}_n^*(t)\right)\, \alpha(\mathrm{d}t) \\
\text{s.t.}\quad & \left.\begin{array}{ll}
\mathrm{d}\boldsymbol{w}_n(t) = \boldsymbol{w}_n(t)\,\mathrm{d}\boldsymbol{p}_n(t)/\boldsymbol{p}_n(t) + \mathrm{d}\boldsymbol{b}_n(t) - \mathrm{d}\boldsymbol{s}_n(t) & \forall\, t, n \\
0 = \sum_{n=1}^N (1+c_b)\,\mathrm{d}\boldsymbol{b}_n(t) - (1-c_s)\,\mathrm{d}\boldsymbol{s}_n(t) & \forall\, t \\
0 \leq \mathrm{E}(\sum_{n=1}^N \boldsymbol{w}_n(T) - e^{\varrho\, T}\, \boldsymbol{w}_n^*(T)) & \\
0 \leq \boldsymbol{w}_n(t)\,, & \forall\, t, n
\end{array}\right\} P\text{-a.s.}
\end{aligned} \tag{$\mathcal{P}_\mathrm{c}$}$$

The presence of portfolio constraints and transaction costs severely limits analytical tractability of problem $\mathcal{P}_\mathrm{c}$, which we consider as a prototype model that will

[1] By using a deterministic benchmark and defining α as the Dirac measure concentrated at T, the risk functional (2.4) reduces to variance of terminal wealth.

eventually be generalized in several directions. For example, we plan to incorporate additional portfolio constraints and alternative *one-sided* risk measures. Moreover, we intend to replace our crude asset price model by a more sophisticated one that accounts for heteroscedasticity, fat tails, parameter instability, etc. Against this background, we should seek a computational approach for solving problem $\mathcal{P}_{\mathrm{c}}$. This requires discretization of the problem with respect to time and (probability) space.

3 Time discretization

To work towards computational tractability of problem $\mathcal{P}_{\mathrm{c}}$, we first simplify its temporal complexity. The goal of this section is to elaborate a stochastic optimization problem in discrete time that approximates the continuous-time problem $\mathcal{P}_{\mathrm{c}}$. To this end, we select a set of ordered time points[2] $0 = t_1 < \cdots < t_H < t_{H+1} = T$ and assume that portfolio rebalancing is restricted to these discrete dates. The optimal value of the approximate problem $\mathcal{P}$ with this extra restriction will provide an upper bound on the optimal value of the original problem $\mathcal{P}_{\mathrm{c}}$.

In order to formulate $\mathcal{P}$ as a discrete-time problem, we have to introduce additional notation. Let $\boldsymbol{w}_{n,h}^{-}$ be the capital invested in asset n at time $t_h - 0$ before reallocation of funds, and let $\boldsymbol{w}_{n,h}^{+}$ be the capital invested in asset n at time $t_h + 0$ after portfolio rebalancing. Furthermore, denote by $\boldsymbol{b}_{n,h}$ and $\boldsymbol{s}_{n,h}$ the amount of money used to buy and sell assets of type n at time t_h, respectively. In agreement with these definitions, $\boldsymbol{w}_{n,h}^{*}$ stands for the capitalization of asset n in the market portfolio at time t_h — note that the market portfolio is never rebalanced. By convention, initial wealth is normalized, i.e., $\sum_{n=1}^{N} \boldsymbol{w}_{n,1}^{-} = 1$. Recall also that the deterministic initial endowments satisfy $\boldsymbol{w}_{n,1}^{-} = \boldsymbol{w}_{n,1}^{*}$. Finally, it is convenient to introduce separate variables for the expected values of the individual asset positions. Thus, we define $\bar{\boldsymbol{w}}_{i,h}^{+} = \mathrm{E}(\boldsymbol{w}_{i,h}^{+})$ and $\bar{\boldsymbol{w}}_{i,h}^{*} = \mathrm{E}(\boldsymbol{w}_{i,h}^{*})$. By using the

[2]For notational convenience, we will sometimes use an additional time point $t_0 := 0$.

notation introduced so far, the objective function of the time-discretized version of problem $\mathcal{P}_{\mathrm{c}}$ can be represented as

$$\sum_{h=1}^{H} \int_{I_h} \mathrm{Var} \left(\sum_{n=1}^{N} (\boldsymbol{w}_{n,h}^{+} - \boldsymbol{w}_{n,h}^{*}) \frac{\boldsymbol{p}_n(t)}{\boldsymbol{p}_n(t_h)} \right) \alpha(\mathrm{d}t) \,, \tag{3.1}$$

where $I_h := [t_h, t_{h+1})$ for $h < H$ and $I_H := [t_H, t_{H+1}]$ denote the time intervals between consecutive rebalancing dates. This expression can be further simplified by using the constants

$$\Delta_{mn,h} := \int_{I_h} \mathrm{E} \left(\frac{\boldsymbol{p}_m(t)}{\boldsymbol{p}_m(t_h)} \right) \mathrm{E} \left(\frac{\boldsymbol{p}_n(t)}{\boldsymbol{p}_n(t_h)} \right) \alpha(\mathrm{d}t) = \int_{I_h} e^{(\mu_m + \mu_n)(t - t_h)} \, \alpha(\mathrm{d}t)$$

and

$$\begin{aligned} \Lambda_{mn,h} &:= \int_{I_h} \mathrm{Cov} \left(\frac{\boldsymbol{p}_m(t)}{\boldsymbol{p}_m(t_h)}, \frac{\boldsymbol{p}_n(t)}{\boldsymbol{p}_n(t_h)} \right) \alpha(\mathrm{d}t) \\ &= \int_{I_h} e^{(\mu_m + \mu_n)(t - t_h)} \, (e^{\sigma_{mn}(t - t_h)} - 1) \, \alpha(\mathrm{d}t) \,. \end{aligned}$$

After some elementary manipulations, the objective function (3.1) takes the form

$$\mathrm{E} \Bigg(\sum_{h=1}^{H} \sum_{m,n=1}^{N} (\boldsymbol{w}_{m,h}^{+} - \boldsymbol{w}_{m,h}^{*})(\Lambda_{mn,h} + \Delta_{mn,h})(\boldsymbol{w}_{n,h}^{+} - \boldsymbol{w}_{n,h}^{*}) \tag{3.2}$$

$$- (\bar{\boldsymbol{w}}_{m,h}^{+} - \bar{\boldsymbol{w}}_{m,h}^{*}) \Delta_{mn,h} (\bar{\boldsymbol{w}}_{n,h}^{+} - \bar{\boldsymbol{w}}_{n,h}^{*}) \Bigg) \,.$$

Next, we reformulate the constraints of the time-discretized version of problem $\mathcal{P}_{\mathrm{c}}$. Our reformulation invokes vectors of price relatives

$$\boldsymbol{\xi}_h := (\boldsymbol{\xi}_{1,h}, \ldots, \boldsymbol{\xi}_{N,h}) \,, \quad \text{where} \quad \boldsymbol{\xi}_{n,h} := \boldsymbol{p}_n(t_h) / \boldsymbol{p}_n(t_{h-1}) \,.$$

Price relatives relate to the time periods between successive rebalancing dates. In our setting, they are serially independent. We define $\mathcal{F}^h := \sigma(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_h)$ to be the σ-algebra induced by the first h price relatives. All constraints discussed below are assumed to hold almost surely with respect to the probability measure P. A

set of static constraints applies for each $h = 1, \ldots, H$.

$$\begin{aligned}
&\boldsymbol{w}^+_{n,h} = \boldsymbol{w}^-_{n,h} + \boldsymbol{b}_{n,h} - \boldsymbol{s}_{n,h} \\
&(1 + c_b) \textstyle\sum_{n=1}^N \boldsymbol{b}_{n,h} = (1 - c_s) \sum_{n=1}^N \boldsymbol{s}_{n,h} \\
&\bar{\boldsymbol{w}}^+_{n,h} = \mathrm{E}(\boldsymbol{w}^+_{n,h})\,,\ \bar{\boldsymbol{w}}^*_{n,h} = \mathrm{E}(\boldsymbol{w}^*_{n,h}) \qquad (3.3a) \\
&\boldsymbol{w}^+_{n,h} \geq 0\,,\ \boldsymbol{b}_{n,h} \geq 0\,,\ \boldsymbol{s}_{n,h} \geq 0 \\
&\boldsymbol{w}^-_{n,h}\,,\ \boldsymbol{w}^+_{n,h}\,,\ \bar{\boldsymbol{w}}^+_{n,h}\,,\ \bar{\boldsymbol{w}}^*_{n,h}\,,\ \boldsymbol{b}_{n,h}\,,\ \boldsymbol{s}_{n,h} \text{ are } \mathcal{F}^h\text{-measurable}
\end{aligned}$$

The following dynamic constraints couple neighboring decision stages and therefore only hold for $h = 1, \ldots, H - 1$.

$$\boldsymbol{w}^-_{n,h+1} = \boldsymbol{w}^+_{n,h}\, \boldsymbol{\xi}_{n,h+1}\,, \quad \boldsymbol{w}^*_{n,h+1} = \boldsymbol{w}^*_{n,h}\, \boldsymbol{\xi}_{n,h+1} \qquad (3.3b)$$

The terminal wealth constraint represents a single static constraint associated with decision stage H.

$$\sum_{n=1}^N \bar{\boldsymbol{w}}^+_{n,H} \geq e^{\varrho T} \sum_{n=1}^N \bar{\boldsymbol{w}}^*_{n,H} \qquad (3.3c)$$

In summary, we denote by $\mathcal{P}$ the problem of minimizing the objective function (3.2) over all investment strategies subject to (3.3). Any strategy feasible in $\mathcal{P}$ corresponds to a strategy feasible in $\mathcal{P}_\mathrm{c}$ with the same objective value. Therefore, the approximate problem $\mathcal{P}$ provides an upper bound on the optimal value of the original problem $\mathcal{P}_\mathrm{c}$. Even though $\mathcal{P}$ has a finite number of decision stages, it involves a continuum of scenarios and thus remains computationally intractable.

4 Multistage stochastic programs

Discrete-time stochastic optimization problems involving a large or infinite number of scenarios are usually addressed by approximating the underlying random data by a discrete process that involves only a modest number of scenarios. This amounts to discretizing the state space of the data generating process. The stochastic program associated with the discrete approximate process has a finite

number of variables and constraints, thus principally allowing for numerical solution.

Stochastic programming research puts much effort in determining whether the optimal values of the approximate and original problems are close to each other and whether the optimal solution of the approximate problem can be related in any way to some near-optimal solution of the original problem. In this work we employ a discretization method which provides deterministic error bounds on the optimal value of the original problem. Moreover, we outline how the solution of the approximate problem can be transformed to a policy that is implementable in reality and achieves an objective value between those bounds. As our method applies to a broad problem class, we explain it by referring to an abstract cost minimization problem under uncertainty. This generalized stochastic program should incorporate expected value constraints, which are also present in the portfolio optimization problem of Section 3. As usual, we assume that decisions are selected at different time points (or stages) indexed $h = 1, \ldots, H$ when new information about the underlying random parameters becomes available. To keep our exposition reasonably simple, we now introduce some notational conventions applying to all discrete-time stochastic processes considered below.

Definition 4.1. *We say that $\boldsymbol{\zeta}$ is a discrete-time stochastic process with state space Z if $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_H)$ and $Z = \times_{h=1}^{H} Z_h$ such that each random vector $\boldsymbol{\zeta}_h$ maps (Ω, Σ) to the Borel space $(Z_h, \mathcal{B}(Z_h))$ and each Z_h is a convex closed subset of some finite-dimensional Euclidean space. Moreover, we define combined random vectors $\boldsymbol{\zeta}^h := (\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_h)$ valued in $Z^h := \times_{i=1}^{h} Z_i$ for all $h = 1, \ldots, H$.*

As an example we mention the process $\boldsymbol{\xi} := (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_H)$, where $\boldsymbol{\xi}_h$ represents the vector of price relatives introduced in Section 3. As stock prices never drop below zero, $\boldsymbol{\xi}$ constitutes a discrete-time stochastic process in the sense of Definition 4.1, whose state space is given by the nonnegative orthant of $\mathbb{R}^{NH}$. Furthermore, we can introduce a decision process $\boldsymbol{x}$ associated with problem $\mathcal{P}$

in Section 3. It is defined through

$$\boldsymbol{x}_h := \mathrm{vec}((\boldsymbol{w}^-_{n,h}, \boldsymbol{w}^+_{n,h}, \bar{\boldsymbol{w}}^+_{n,h}, \boldsymbol{w}^*_{n,h}, \bar{\boldsymbol{w}}^*_{n,h}, \boldsymbol{b}_{n,h}, \boldsymbol{s}_{n,h}) \,|\, n = 1, \ldots, N)\,,$$

where the operator 'vec' returns the concatenation of its arguments. Thus, $\boldsymbol{x}_h$ has dimension $n_h := 7N$. As the constraints (3.3) preclude negative $\boldsymbol{x}_h$, the state space of the process $\boldsymbol{x}$ can be chosen to be the nonnegative orthant of $\mathbb{R}^{7NH}$.

In order to present the most general form of our approximation method, we assume that the multistage stochastic program under consideration is driven by two exogenous stochastic processes $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ with state spaces Θ and Ξ, respectively. We assume that $\boldsymbol{\eta}$ impacts the objective function of $\mathcal{P}$, whereas $\boldsymbol{\xi}$ appears in the constraints.[3] For notational convenience, we also introduce the combined data process $\boldsymbol{\zeta} := (\boldsymbol{\eta}, \boldsymbol{\xi})$ with state space $Z := \Theta \times \Xi$. Furthermore, we let $\mathcal{F}^h := \sigma(\boldsymbol{\zeta}^h)$ be the information that is available at stage h by observing the data process. We will frequently use the shorthand notation $\mathcal{F} := \mathcal{F}^H$ and denote by $\mathbb{F} := \{\mathcal{F}^h\}_{h=1}^H$ the filtration generated by $\boldsymbol{\zeta}$. With these conventions, we can formulate an abstract multistage stochastic program with expected value constraints as follows.

$$\begin{aligned} &\underset{\boldsymbol{x} \in X(\mathbb{F})}{\text{minimize}}\ \mathrm{E}\,(c(\boldsymbol{x}, \boldsymbol{\eta})) \\ &\text{s.t.}\quad \mathrm{E}(f_h(\boldsymbol{x}, \boldsymbol{\xi}) \,|\, \mathcal{F}^h) \leq 0 \quad P\text{-a.s. } \forall h = 1, \ldots, H \end{aligned} \tag{$\mathcal{P}$}$$

Minimization is over a convex set of stochastic processes $\boldsymbol{x}$, all of which share the same state space $X \subset \times_{h=1}^H \mathbb{R}^{n_h}$. These processes are commonly called strategies, policies, or decision processes. The set of admissible strategies is defined as

$$X(\mathbb{F}) := \{\boldsymbol{x} \in \times_{h=1}^H \mathcal{L}^\infty(\Omega, \mathcal{F}^h, P; \mathbb{R}^{n_h}) \,|\, \boldsymbol{x}(\omega) \in X \text{ for } P\text{-a.e. } \omega \in \Omega\}\,.$$

Note that all strategies in $X(\mathbb{F})$ are adapted to $\mathbb{F}$, that is, they are *non-anticipative* with respect to the underlying data process [22]. The objective criterion and the

[3] The process of price relatives shows up only in the constraints of the portfolio problem of Section 3. In agreement with the current convention, it is therefore denoted by $\boldsymbol{\xi}$. There is no (or only a fictitious deterministic) process $\boldsymbol{\eta}$ that affects the objective function of that problem.

(explicit) constraints in $\mathcal{P}$ are determined through a real-valued cost function $c : X \times \Theta \to \mathbb{R}$ and a vector-valued constraint function $f_h : X \times \Xi \to \mathbb{R}^{m_h}$ for each stage index $h = 1, \ldots, H$. We emphasize that problem $\mathcal{P}$ accommodates expected value constraints. The corresponding constraint functions are required to be nonpositive in expectation (instead of almost everywhere), where expectation is conditional on the stagewise information sets. In order to elaborate our space discretization scheme, we impose the following regularity conditions:

(C1) c is convex in x, concave in η, and continuous;

(C2) f_h is representable as $f_h(x, \xi) = \tilde{f}_h((1, x) \otimes (1, \xi))$, where $\tilde{f}_h$ is convex, continuous, and constant in $x_i \otimes \xi_j$ for all $1 \leq j \leq i \leq H$, $h = 1, \ldots, H$;

(C3) X is convex and compact;

(C4) $\boldsymbol{\zeta}$ constitutes a serially independent process with compact state space Z.

Here, the operator '$\otimes$' stands for the usual dyadic product of vectors. The assumptions (C1) and (C2) allow for linear multistage stochastic programs which exhibit randomness in the objective function coefficients, the right hand side vectors, and the technology matrices (but not in the recourse matrices). As reasonable numerical solutions are always bounded, assumption (C3) is nonrestrictive. The compactness requirement in (C4) can always be enforced by truncating certain extreme scenarios of the data process $\boldsymbol{\zeta}$ that have a negligible effect on the solution of the stochastic program $\mathcal{P}$. Moreover, the serial independence requirement can often be circumvented by rewriting the original data process as a Rosenblatt transformation [23] of a serially independent noise process. In this case, any serial dependencies can be absorbed in the definition of the cost and constraint functions.

For the further discussion, we introduce the set

$$Y(\mathbb{F}) := \{\boldsymbol{y} \in \times_{h=1}^{H} \mathcal{L}^1(\Omega, \mathcal{F}^h, P; \mathbb{R}^{m_h}) \,|\, \boldsymbol{y} \geq 0 \ P\text{-a.s.}\},$$

which comprises all nonnegative integrable *dual* decision processes adapted to the filtration $\mathbb{F}$. By using $Y(\mathbb{F})$, the stochastic optimization problem $\mathcal{P}$ *with* explicit constraints can be recast as a min-max problem *without* explicit constraints.

Lemma 4.2 (Wright [25, § 4])**.** *Under the conditions (C1)–(C3) we have*

$$\inf \mathcal{P} = \inf_{\boldsymbol{x} \in X(\mathbb{F})} \sup_{\boldsymbol{y} \in Y(\mathbb{F})} \mathrm{E} \left(c(\boldsymbol{x}, \boldsymbol{\eta}) + \sum_{h=1}^{H} \boldsymbol{y}_h \cdot f_h(\boldsymbol{x}, \boldsymbol{\xi}) \right) . \tag{4.1}$$

The min-max formulation (4.1) of problem $\mathcal{P}$ will be very helpful for the development of our space discretization scheme in Section 5.

5 Space discretization

Recall that we assume $\boldsymbol{\zeta}$ to be a serially independent process with compact state space Z. Besides that we make no assumption concerning its distribution. We now construct another process $\boldsymbol{\zeta}^u = (\boldsymbol{\eta}^u, \boldsymbol{\xi}^u)$ such that the component processes $\boldsymbol{\eta}^u$ and $\boldsymbol{\xi}^u$ have state spaces Θ and Ξ, respectively. Thus, $\boldsymbol{\zeta}^u$ has the same state space $Z = \Theta \times \Xi$ as the original process $\boldsymbol{\zeta}$. Furthermore, $\boldsymbol{\zeta}^u$ has a discrete distribution and approximates $\boldsymbol{\zeta}$ in a sense that will be explained below.

The construction of $\boldsymbol{\zeta}^u$ goes along the lines of [17, § 4]. To keep this article self-contained, we briefly sketch the basic procedure. Essentially, $\boldsymbol{\zeta}^u$ is constructed by specifying its conditional distribution given $\boldsymbol{\zeta}$. Then, the joint distribution of $\boldsymbol{\zeta}$ and $\boldsymbol{\zeta}^u$ is uniquely determined by the product measure theorem [1, Theorem 2.6.2]. As pointed out in [17, § 4], there always exists a rich enough sample space (Ω, Σ, P) on which both $\boldsymbol{\zeta}$ and $\boldsymbol{\zeta}^u$ are defined.[4]

The distribution of $\boldsymbol{\zeta}^u$ conditional on $\boldsymbol{\zeta} = \zeta$ is obtained by using the product measure theorem to combine the conditional distributions $\{P_h^u\}_{h=1}^H$, where P_h^u stands for the distribution of $\boldsymbol{\zeta}_h^u$ conditional on $\boldsymbol{\zeta} = \zeta$ and $\boldsymbol{\zeta}^{u,h-1} = \zeta^{u,h-1}$. Thus,

[4] We are free to set $\Omega = Z \times Z$ and $\Sigma = \mathcal{B}(Z \times Z)$, while P may be identified with the joint distribution of $\boldsymbol{\zeta}$ and $\boldsymbol{\zeta}^u$.

construction of the approximate process $\boldsymbol{\zeta}^u$ boils down to specifying P_h^u for each h. In order to so, we select a disjoint set partition $\mathcal{Z}_h(\zeta^{u,h-1})$ of Z_h which may depend on the realizations of the approximate process up to time $h-1$ and satisfies

$$Z_h = \bigcup_{W \in \mathcal{Z}_t(\zeta^{u,h-1})} W\,, \quad W \cup W' = \emptyset \quad \text{for} \quad W \neq W' \in \mathcal{Z}_h(\zeta^{u,h-1})\,.$$

Moreover, we require each $W \in \mathcal{Z}_h(\zeta^{u,h-1})$ to be representable as $W = U \times V$ where $U \subset \Theta_t$ and $V \subset \Xi_t$ are bounded nondegenerate (not necessarily closed) simplices. For a bounded nondegenerate simplex V, an extreme point $e \in \operatorname{ext}\operatorname{cl} V$, and a vector $v \in \operatorname{cl} V$ we denote by $P_V(e|v)$ the convex weight of v with respect to e. By definition, we thus have

$$\sum_{e \in \operatorname{ext}\operatorname{cl} V} P(e|v) = 1 \quad \text{and} \quad \sum_{e \in \operatorname{ext}\operatorname{cl} V} eP(e|v) = v\,.$$

Moreover, we introduce a combined vector $\zeta_h^V(e) = (\eta_h^V(e), \xi_h^V(e))$ defined through

$$\xi_h^V(e) := e \quad \text{and} \quad \eta_h^V(e) := \frac{\mathrm{E}(\boldsymbol{\eta}_h\, P_V(e|\boldsymbol{\xi}_h)\, 1_W(\boldsymbol{\zeta}_h))}{\mathrm{E}(1_W(\boldsymbol{\zeta}_h))}\,.$$

Without much loss of generality we may assume that the denominator in the rightmost expression is always nonzero. Using the notation introduced so far, we can specify the conditional distribution P_h^u. For $\zeta \in Z$ and $\zeta^{u,h-1} \in Z^{h-1}$ we set

$$P_h^u(\cdot|\zeta, \zeta^{u,h-1}) \;:=\; \sum_{\substack{W \in \mathcal{Z}_h(\zeta^{u,h-1}) \\ W = U \times V}} \sum_{e \in \operatorname{ext}\operatorname{cl} V} P_V(e|\xi_h)\, 1_W(\zeta_h)\, \delta_{\zeta_h^V(e)}(\cdot)\,, \qquad (5.1)$$

where $\delta_{\zeta_h^V(e)}$ denotes the Dirac measure concentrated at $\zeta_h^V(e)$. The approximate process $\boldsymbol{\zeta}^u$ obtained by combining the P_h^u in the appropriate way has several intriguing properties. First, by averaging (5.1) over all ζ_h, one verifies that the marginal distribution of $\boldsymbol{\zeta}^u$ constitutes a *barycentric scenario tree* in the sense of Frauendorfer, see e.g. [7, 9]. As opposed to the traditional construction of barycentric scenario trees, the present approach allows us to view $\boldsymbol{\zeta}^u$ and $\boldsymbol{\zeta}$ as

correlated processes on a common probability space. In the reminder of this section we will elaborate some of the advantages of this new perspective.

From now on we denote by $\mathbb{F}^u$ the filtration generated by $\boldsymbol{\zeta}^u$, i.e., $\mathbb{F}^u := \{\mathcal{F}^{u,h}\}_{h=1}^H$ where $\mathcal{F}^{u,h} := \sigma(\boldsymbol{\zeta}^{u,h})$, and we use the convention $\mathcal{F}^u := \mathcal{F}^{u,H}$.

Lemma 5.1 (Kuhn [17, § 4])**.** *The following relations hold for suitable versions of the conditional expectations, respectively.*

$$\mathrm{E}(\boldsymbol{x}|\mathcal{F}) \in X(\mathbb{F}) \text{ for all } \boldsymbol{x} \in X(\mathbb{F}^u) \tag{5.2a}$$

$$\mathrm{E}(\boldsymbol{y}|\mathcal{F}^u) \in Y(\mathbb{F}^u) \text{ for all } \boldsymbol{y} \in Y(\mathbb{F}) \tag{5.2b}$$

$$\mathrm{E}(\boldsymbol{\xi}^u|\mathcal{F}) = \boldsymbol{\xi} \tag{5.2c}$$

$$\mathrm{E}(\boldsymbol{\eta}|\mathcal{F}^u) = \boldsymbol{\eta}^u \tag{5.2d}$$

The relations 5.2 are crucial for our main result on the approximation of discrete-time stochastic programs; see Theorem 5.5 below. A rigorous proof of Lemma 5.1 is provided in [17, § 4]. Here, we only give some intuition. The inclusions (5.2a) and (5.2b) are related to the recursive construction of the $\{\boldsymbol{\zeta}_h^u\}_{h=1}^T$. Relation (5.2c) follows immediately from the definition of P_h^u. In fact, we have

$$\begin{aligned} \mathrm{E}(\boldsymbol{\xi}_h^u|\boldsymbol{\zeta}, \boldsymbol{\zeta}^{u,h-1}) &= \int_{Z_h} \xi_h^u \, P_h^u(\mathrm{d}\zeta_h^u|\boldsymbol{\zeta}, \boldsymbol{\zeta}^{u,h-1}) \\ &= \sum_{\substack{W \in \mathcal{Z}_h(\zeta^{u,h-1}) \\ W = U \times V}} 1_W(\boldsymbol{\zeta}_h) \sum_{e \in \operatorname{ext}\operatorname{cl} V} e \, P_V(e|\boldsymbol{\xi}_h) \;=\; \boldsymbol{\xi}_h \quad P\text{-a.s.} \end{aligned} \tag{5.3}$$

The law of iterated conditional expectations then allows us to conclude that $\mathrm{E}(\boldsymbol{\xi}_h^u|\boldsymbol{\zeta}) = \boldsymbol{\xi}_h$ for all h, which is equivalent to (5.2c). The last relation (5.2d) is proved in a similar manner. We first use Bayes' theorem to evaluate the conditional expectation of $\boldsymbol{\eta}_h$ given $\{\boldsymbol{\zeta}_i\}_{i \neq h}$ and $\boldsymbol{\zeta}^{u,h}$ and then employ the law of iterated conditional expectations.

It should be emphasized that there is considerable flexibility in the construction of $\boldsymbol{\zeta}^u$ since there are many different ways to specify the partitions $\mathcal{Z}_h(\zeta^{u,h-1})$. If all of these partitions contain only sets of diameter less than ϵ, then the supremum distance of $\boldsymbol{\zeta}$ and $\boldsymbol{\zeta}^u$ is at most ϵ, that is, $\|\boldsymbol{\zeta} - \boldsymbol{\zeta}^u\|_\infty \leq \epsilon$.

It proves useful to introduce another discrete process $\boldsymbol{\zeta}^l = (\boldsymbol{\eta}^l, \boldsymbol{\xi}^l)$ such that $\boldsymbol{\eta}^l$ and $\boldsymbol{\xi}^l$ are valued in Θ and Ξ, respectively. Again, $\boldsymbol{\zeta}^l$ is supposed to approximate the original data process $\boldsymbol{\zeta}$. We construct $\boldsymbol{\zeta}^l$ in exactly the same way as $\boldsymbol{\zeta}^u$, but the roles of $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are interchanged. Again, the construction is very flexible, relying on a family of disjoint set partitions $\mathcal{Z}_h(\zeta^{l,h-1})$ that depend on the history of the approximate process. By choosing these partitions appropriately, for every given tolerance $\epsilon > 0$ we can construct a discrete processes $\boldsymbol{\zeta}^l$ with $\|\boldsymbol{\zeta} - \boldsymbol{\zeta}^l\|_\infty \leq \epsilon$.

The induced filtration $\mathbb{F}^l$ is constructed as usual, i.e., $\mathbb{F}^l := \{\mathcal{F}^{l,h}\}_{h=1}^H$ where $\mathcal{F}^{l,h} := \sigma(\boldsymbol{\zeta}^{l,h})$, and we use the convention $\mathcal{F}^l := \mathcal{F}^{l,H}$. By permutation symmetry, the following result is an immediate consequence of Lemma 5.1.

Corollary 5.2. *The following relations hold for suitable versions of the conditional expectations, respectively.*

$$\mathrm{E}(\boldsymbol{x}|\mathcal{F}^l) \in X(\mathbb{F}^l) \text{ for all } \boldsymbol{x} \in X(\mathbb{F}) \tag{5.4a}$$

$$\mathrm{E}(\boldsymbol{y}|\mathcal{F}) \in Y(\mathbb{F}) \text{ for all } \boldsymbol{y} \in Y(\mathbb{F}^l) \tag{5.4b}$$

$$\mathrm{E}(\boldsymbol{\xi}|\mathcal{F}^l) = \boldsymbol{\xi}^l \tag{5.4c}$$

$$\mathrm{E}(\boldsymbol{\eta}^l|\mathcal{F}) = \boldsymbol{\eta} \tag{5.4d}$$

If we replace the true data process $\boldsymbol{\zeta}$ by $\boldsymbol{\zeta}^u$ and the true filtration $\mathbb{F}$ by $\mathbb{F}^u$ in $\mathcal{P}$, then we obtain an approximate optimization problem denoted $\mathcal{P}^u$. Another approximate problem $\mathcal{P}^l$ is obtained by substituting $\boldsymbol{\zeta}^l$ for $\boldsymbol{\zeta}$ and $\mathbb{F}^l$ for $\mathbb{F}$. Note that replacing the filtrations has a primal and a dual effect, that is, after substitution the primal decisions as well as the constraints are adapted to the approximate filtration. Notice that Lemma 4.2 remains valid for problems $\mathcal{P}^u$ and $\mathcal{P}^l$ with their corresponding data processes and filtrations. In the remainder of this section we will prove that the optimal values of $\mathcal{P}^l$ and $\mathcal{P}^u$ bracket the optimal value of the original problem $\mathcal{P}$. To this end, we first establish two technical lemmas.

Lemma 5.3. *The following relations hold for suitable versions of the conditional expectations and for all* $1 \le i < j \le H$.

(i) $\mathrm{E}(\boldsymbol{x}_i \otimes \boldsymbol{\xi}_j^u | \mathcal{F}) = \mathrm{E}(\boldsymbol{x}_i | \mathcal{F}) \otimes \boldsymbol{\xi}_j$ *for all* $\boldsymbol{x} \in X(\mathbb{F}^u)$

(ii) $\mathrm{E}(\boldsymbol{x}_i \otimes \boldsymbol{\xi}_j | \mathcal{F}^l) = \mathrm{E}(\boldsymbol{x}_i | \mathcal{F}^l) \otimes \boldsymbol{\xi}_j^l$ *for all* $\boldsymbol{x} \in X(\mathbb{F})$

Proof. Select arbitrary time indices i and j such that $1 \le i < j \le H$, and let $\boldsymbol{x}$ be an element of $X(\mathbb{F}^u)$. Then, we find

$$\begin{aligned}
\mathrm{E}(\boldsymbol{x}_i \otimes \boldsymbol{\xi}_j^u | \mathcal{F}) &= \mathrm{E}(\mathrm{E}(\boldsymbol{x}_i \otimes \boldsymbol{\xi}_j^u | \mathcal{F} \wedge \mathcal{F}^{u,j-1}) | \mathcal{F}) \\
&= \mathrm{E}(\boldsymbol{x}_i \otimes \mathrm{E}(\boldsymbol{\xi}_j^u | \mathcal{F} \wedge \mathcal{F}^{u,j-1}) | \mathcal{F}) \\
&= \mathrm{E}(\boldsymbol{x}_i \otimes \boldsymbol{\xi}_j | \mathcal{F}) \\
&= \mathrm{E}(\boldsymbol{x}_i | \mathcal{F}) \otimes \boldsymbol{\xi}_j \,,
\end{aligned}$$

where the third equality follows from (5.3). The above argument proves assertion (i). Next, let $\boldsymbol{x}$ be an element of $X(\mathbb{F})$. By assumption, $\boldsymbol{x}_i$ is a function of $\boldsymbol{\xi}^i$, which is independent of $\boldsymbol{\xi}_j$. Assertion (ii) immediately follows from (5.4c) and independence of $\boldsymbol{x}_i$ and $\boldsymbol{\xi}_j$. □

Lemma 5.4. *The following relations hold for suitable versions of the conditional expectations and for all* $1 \le h \le H$.

(i) $\mathrm{E}(f_h(\boldsymbol{x}, \boldsymbol{\xi}^u) | \mathcal{F}) \ge f_h(\mathrm{E}(\boldsymbol{x} | \mathcal{F}), \boldsymbol{\xi})$ *for all* $\boldsymbol{x} \in X(\mathbb{F}^u)$

(ii) $\mathrm{E}(f_h(\boldsymbol{x}, \boldsymbol{\xi}) | \mathcal{F}^l) \ge f_h(\mathrm{E}(\boldsymbol{x} | \mathcal{F}^l), \boldsymbol{\xi}^l)$ *for all* $\boldsymbol{x} \in X(\mathbb{F})$

Proof. Select $\boldsymbol{x} \in X(\mathbb{F}^u)$. By using condition (C2) and the conditional Jensen inequality we find

$$\begin{aligned}
\mathrm{E}(f_h(\boldsymbol{x}, \boldsymbol{\xi}^u) | \mathcal{F}) &= \mathrm{E}(\tilde{f}_h((1, \boldsymbol{x}) \otimes (1, \boldsymbol{\xi}^u)) | \mathcal{F}) \\
&\ge \tilde{f}_h(\mathrm{E}((1, \boldsymbol{x}) \otimes (1, \boldsymbol{\xi}^u) | \mathcal{F})) \\
&= \tilde{f}_h(\mathrm{E}((1, \boldsymbol{x}) | \mathcal{F}) \otimes (1, \boldsymbol{\xi})) \\
&= f_h(\mathrm{E}(\boldsymbol{x} | \mathcal{F}), \boldsymbol{\xi}) \,.
\end{aligned}$$

Note that the equality in the third line follows from Lemma 5.3(i) and the independence of $\tilde{f}_h$ in $x_i \otimes \xi_j$ for all $i \geq j$. This establishes (i). Assertion (ii) is proved in a similar manner. □

Armed with the above preliminary results we are now ready to state the main theorem of this section.

Theorem 5.5. *Assume that the problems $\mathcal{P}^l$ and $\mathcal{P}^u$ are solvable with finite optimal values. If $\boldsymbol{x}^u$ solves $\mathcal{P}^u$, then $\hat{\boldsymbol{x}} := \mathrm{E}(\boldsymbol{x}^u|\mathcal{F})$ is feasible in $\mathcal{P}$ and*

$$\inf \mathcal{P}^l \leq \inf \mathcal{P} \leq \mathrm{E}(c(\hat{\boldsymbol{x}}, \boldsymbol{\eta})) \leq \inf \mathcal{P}^u.$$

Proof. Our argumentation relies on ideas from the proofs of Theorems 1 and 2 in [17] as well as Theorem 5.1 in [18]. A preliminary calculation yields

$$\begin{aligned}
\inf \mathcal{P} &\geq \inf_{\boldsymbol{x} \in X(\mathbb{F})} \sup_{\boldsymbol{y} \in Y(\mathbb{F}^l)} \mathrm{E}\left(c(\boldsymbol{x}, \mathrm{E}(\boldsymbol{\eta}^l|\mathcal{F})) + \sum_{h=1}^{H} \mathrm{E}(\boldsymbol{y}_h|\mathcal{F}) \cdot f_h(\boldsymbol{x}, \boldsymbol{\xi}) \right) \\
&\geq \inf_{\boldsymbol{x} \in X(\mathbb{F})} \sup_{\boldsymbol{y} \in Y(\mathbb{F}^l)} \mathrm{E}\left(c(\boldsymbol{x}, \boldsymbol{\eta}^l) + \sum_{h=1}^{H} \boldsymbol{y}_h \cdot f_h(\boldsymbol{x}, \boldsymbol{\xi}) \right) \\
&= \inf_{\boldsymbol{x} \in X(\mathbb{F})} \sup_{\boldsymbol{y} \in Y(\mathbb{F}^l)} \mathrm{E}\left(\mathrm{E}(c(\boldsymbol{x}, \boldsymbol{\eta}^l)|\mathcal{F}^l) + \sum_{h=1}^{H} \boldsymbol{y}_h \cdot \mathrm{E}(f_h(\boldsymbol{x}, \boldsymbol{\xi})|\mathcal{F}^l) \right).
\end{aligned}$$

The first inequality follows from Lemma 4.2 as well as the relations (5.4b) and (5.4d). Moreover, the second inequality uses the conditional Jensen inequality, which applies since the cost function is concave in its second argument, while $\boldsymbol{x}$ and $\boldsymbol{\xi}$ are $\mathcal{F}$-measurable. From the above chain of inequalities follows

$$\begin{aligned}
\inf \mathcal{P} &\geq \inf_{\boldsymbol{x} \in X(\mathbb{F})} \sup_{\boldsymbol{y} \in Y(\mathbb{F}^l)} \mathrm{E}\left(c(\mathrm{E}(\boldsymbol{x}|\mathcal{F}^l), \boldsymbol{\eta}^l) + \sum_{h=1}^{H} \boldsymbol{y}_h \cdot f_h(\mathrm{E}(\boldsymbol{x}|\mathcal{F}^l), \boldsymbol{\xi}^l) \right) \\
&\geq \inf_{\boldsymbol{x} \in X(\mathbb{F}^l)} \sup_{\boldsymbol{y} \in Y(\mathbb{F}^l)} \mathrm{E}\left(c(\boldsymbol{x}, \boldsymbol{\eta}^l) + \sum_{h=1}^{H} \boldsymbol{y}_h \cdot f_h(\boldsymbol{x}, \boldsymbol{\xi}^l) \right).
\end{aligned}$$

Here, the first inequality holds by Lemma 5.4 (ii) and the conditional Jensen inequality, which applies since the cost function is convex in its first argument,

while $\boldsymbol{y}$ and $\boldsymbol{\eta}^l$ are $\mathcal{F}^l$-measurable. Note that the second inequality is due to assumptions (5.4a). In conclusion, we thus have shown $\inf \mathcal{P}^l \leq \inf \mathcal{P}$.

Next, we use the fact that $\boldsymbol{x}^u$ is an element of $X(\mathbb{F}^u)$, implying via (5.2a) that the conditional expectation $\mathrm{E}(\boldsymbol{x}^u|\mathcal{F})$ is an element of $X(\mathbb{F})$.

$$\begin{aligned}
\inf \mathcal{P} &\leq \sup_{\boldsymbol{y} \in Y(\mathbb{F})} \mathrm{E}\left(c(\mathrm{E}(\boldsymbol{x}^u|\mathcal{F}), \boldsymbol{\eta}) + \sum_{h=1}^{H} \boldsymbol{y}_h \cdot f_h(\mathrm{E}(\boldsymbol{x}^u|\mathcal{F}), \boldsymbol{\xi}) \right) \\
&\leq \sup_{\boldsymbol{y} \in Y(\mathbb{F})} \mathrm{E}\left(\mathrm{E}(c(\boldsymbol{x}^u, \boldsymbol{\eta})|\mathcal{F}) + \sum_{h=1}^{H} \boldsymbol{y}_h \cdot \mathrm{E}(f_h(\boldsymbol{x}^u, \boldsymbol{\xi}^u)|\mathcal{F}) \right) \qquad (5.5) \\
&= \sup_{\boldsymbol{y} \in Y(\mathbb{F})} \mathrm{E}\left(c(\boldsymbol{x}^u, \boldsymbol{\eta}) + \sum_{h=1}^{H} \boldsymbol{y}_h \cdot f_h(\boldsymbol{x}^u, \boldsymbol{\xi}^u) \right).
\end{aligned}$$

The second inequality in (5.5) uses Lemma 5.4 (i) and the conditional Jensen inequality, while the equality relies on the law of iterated conditional expectations. Another application of the conditional Jensen inequality yields

$$\begin{aligned}
\inf \mathcal{P} &\leq \sup_{\boldsymbol{y} \in Y(\mathbb{F})} \mathrm{E}\left(c(\boldsymbol{x}^u, \mathrm{E}(\boldsymbol{\eta}|\mathcal{F}^u)) + \sum_{h=1}^{H} \mathrm{E}(\boldsymbol{y}_h|\mathcal{F}^u) \cdot f_h(\boldsymbol{x}^u, \boldsymbol{\xi}^u) \right) \\
&\leq \sup_{\boldsymbol{y} \in Y(\mathbb{F}^u)} \mathrm{E}\left(c(\boldsymbol{x}^u, \boldsymbol{\eta}^u) + \sum_{h=1}^{H} \boldsymbol{y}_h \cdot f_h(\boldsymbol{x}^u, \boldsymbol{\xi}^u) \right) \\
&= \inf_{\boldsymbol{x} \in X(\mathbb{F}^u)} \sup_{\boldsymbol{y} \in Y(\mathbb{F}^u)} \mathrm{E}\left(c(\boldsymbol{x}, \boldsymbol{\eta}^u) + \sum_{h=1}^{H} \boldsymbol{y}_h \cdot f_h(\boldsymbol{x}, \boldsymbol{\xi}^u) \right).
\end{aligned}$$

Here, the second inequality holds by the assumptions (5.2b) and (5.2d), entailing a relaxation of the dual feasible set. The last line of the above expression corresponds to $\inf \mathcal{P}^u$, which is finite by assumption. This implies that the supremum over $Y(\mathbb{F})$ in the first line of (5.5) is also finite, and $\boldsymbol{y} = \boldsymbol{0}$ is an optimal solution. Thus, $\hat{\boldsymbol{x}} = \mathrm{E}(\boldsymbol{x}^u|\mathcal{F})$ is feasible in $\mathcal{P}$, and the corresponding objective value $\mathrm{E}(c(\hat{\boldsymbol{x}}, \boldsymbol{\eta}))$ satisfies the postulated inequalities. □

As the marginal distributions of $\boldsymbol{\zeta}^l$ and $\boldsymbol{\zeta}^u$ are discrete, the approximate problems $\mathcal{P}^l$ and $\mathcal{P}^u$ constitute finite-dimensional convex programs with a finite number of constraints. Thus, they principally allow for numerical solution.

By Theorem 5.5, the optimal values of the approximate problems provide an a priori estimate of the minimal cost that is principally achievable. However, the optimal strategies corresponding to $\mathcal{P}^l$ and $\mathcal{P}^u$ are only given for discrete scenarios and are therefore not implementable almost surely. It is a priori unclear how the optimal solutions of the approximate problems can be used to determine a near-optimal strategy for the original problem. Theorem 5.5 provides a particularly satisfactory answer to this question by proposing a policy $\hat{\boldsymbol{x}}$ which is implementable in every possible scenario and whose expected cost is bracketed by $\inf \mathcal{P}^l$ and $\inf \mathcal{P}^u$. Note that $\mathrm{E}(c(\hat{\boldsymbol{x}}, \boldsymbol{\eta}))$ represents an a posteriori estimate of the minimal cost which is achievable in reality. It can conveniently be calculated by Monte Carlo simulation. In fact, since $\boldsymbol{x}^u$ is finitely supported, evaluation of $\hat{\boldsymbol{x}}$ for an arbitrary realization of $\boldsymbol{\zeta}$ reduces to the evaluation of a finite sum and poses no computational challenges. Generically, $\mathrm{E}(c(\hat{\boldsymbol{x}}, \boldsymbol{\eta}))$ constitutes a better upper bound on the true objective value than $\inf \mathcal{P}^u$.

Note that there is considerable flexibility in constructing the discrete approximate processes. In fact, we are entirely free in choosing the partitions $\mathcal{Z}_h(\zeta^{u,h-1})$ and $\mathcal{Z}_h(\zeta^{l,h-1})$ for $h = 1, \ldots, H$, and each choice corresponds to a discrete process $\boldsymbol{\zeta}^u$ and $\boldsymbol{\zeta}^l$, respectively, which provides an upper or lower bound on the optimal objective value. By making the diameters of the polytopes in all partitions uniformly small, we can construct discrete stochastic processes that approximate the original process arbitrarily well with respect to the $\mathcal{L}^\infty$-norm. Thus, for each $J \in \mathbb{N}$ we can introduce two discrete processes $\boldsymbol{\zeta}^u_J = (\boldsymbol{\eta}^u_J, \boldsymbol{\xi}^u_J)$ and $\boldsymbol{\zeta}^l_J = (\boldsymbol{\eta}^l_J, \boldsymbol{\xi}^l_J)$ of the above kind such that

$$\lim_{J\to\infty} \|\boldsymbol{\zeta}^u_J - \boldsymbol{\zeta}\|_\infty = \lim_{J\to\infty} \|\boldsymbol{\zeta}^l_J - \boldsymbol{\zeta}\|_\infty = 0\,.$$

In the remainder, let $\mathcal{P}^u_J$ and $\mathcal{P}^l_J$ be the approximate stochastic programs associated with the discrete processes $\boldsymbol{\zeta}^u_J$ and $\boldsymbol{\zeta}^l_J$, respectively. If we assume that the original stochastic program $\mathcal{P}$ satisfies an unrestrictive strict feasibility condition,

then for $J \to \infty$ we have

$$\inf \mathcal{P}_J^u \to \inf \mathcal{P} \quad \text{and} \quad \inf \mathcal{P}_J^l = \inf \mathcal{P}\,. \tag{5.6}$$

The proof of this convergence result follows the lines of [18, § 6] and is omitted for brevity of exposition. For each $J \in \mathbb{N}$ for which $\mathcal{P}_J^u$ and $\mathcal{P}_J^l$ are solvable, we can further introduce a policy $\hat{\boldsymbol{x}}_J$ as in Theorem 5.5. Combining (5.6) and Theorem 5.5, we conclude that $\hat{\boldsymbol{x}}_J$ is feasible in $\mathcal{P}$ and that for $J \to \infty$ we have

$$\mathrm{E}(c(\hat{\boldsymbol{x}}_J, \boldsymbol{\eta})) \to \inf \mathcal{P}\,.$$

In spite of these satisfying theoretical results, the approximation error $E_J := \inf \mathcal{P}_J^u - \inf \mathcal{P}_J^l$ cannot be made arbitrarily small in practice due to limited computer power. It is a major challenge to select the set partitions $\mathcal{Z}_h(\zeta^{u,h-1})$ and $\mathcal{Z}_h(\zeta^{l,h-1})$ recursively for $h = 1, \ldots, H$ in such a way that the approximation error becomes small for manageable problem sizes. Ideally we would like to proceed as follows: find the least number of discretization points given an acceptable level of error. Of course, such a scheme is difficult to implement since computing the error for a given configuration of discretization points involves the solution of two large-scale multistage stochastic programs, which is extremely time consuming. For this reason, we use a heuristic method based on a stochastic search algorithm that can find a good solution without requiring a large number of function evaluations. The heuristic we use is known under the name 'threshold accepting', see [3]. The basic idea is to perform a random walk on all the possible tree configurations of a given complexity. For our purposes, the complexity of a tree is defined as the number of its nodes. The advantage of threshold accepting is that it requires little tuning of its parameters in order for it to compute an acceptable solution. Our algorithm can be described as follows:

Step 1 Set $J = 1$, and let $\mathcal{Z}_J^u := \{\mathcal{Z}_{J,h}^u\}_{h=1}^H$ and $\mathcal{Z}_J^l := \{\mathcal{Z}_{J,h}^l\}_{h=1}^H$ be the stagewise set partitions corresponding to some scenario trees of the upper and lower

approximations, respectively.[5] Let E_J, be the error (difference between upper and lower bounds) associated with $\mathcal{Z}_J^u$ and $\mathcal{Z}_J^l$.

Step 2 If the error E_J is smaller than some predefined tolerance, then stop. Otherwise, perform small random perturbations (described in § 6) in the current trees to obtain $\widehat{\mathcal{Z}}^u$ and $\widehat{\mathcal{Z}}^l$.

Step 3 Solve the deterministic equivalent problems associated with $\widehat{\mathcal{Z}}^u$ and $\widehat{\mathcal{Z}}^l$, and define the error $\widehat{E}$ in the obvious way.

Step 4 Define $\Delta E := E_J - \widehat{E}$ and set

$$(\mathcal{Z}_{J+1}^u, \mathcal{Z}_{J+1}^l, E_{J+1}) := \begin{cases} (\widehat{\mathcal{Z}}^u, \widehat{\mathcal{Z}}^l, \widehat{E}) & \text{if } \Delta E > -\tau, \\ (\mathcal{Z}_J^u, \mathcal{Z}_J^l, E_J) & \text{otherwise.} \end{cases}$$

Step 5 If the best of all visited configurations does not improve for more that u iterations, then lower the threshold acceptance criterion τ. Finally, increase J by one unit and go to Step 2.

6 Case study

We now discuss a small case study using the discretization scheme proposed above. In Section 5 we described how the continuous refinement of the probability space will eventually lead to the solution of the time-discretized problem. In this section we discuss the practical implementation of such a scheme.

We assume that we are given a time-discretized version $\mathcal{P}$ of the original portfolio problem, which involves H decision stages. In the remainder we keep H fixed and focus on space discretization. In order to arrive at an implementable version of the problem, we start by covering the support of the random vector $\boldsymbol{\xi}_h$

[5] To keep our numerical experiments simple, we assume here that the disjoint set partitions are not path-dependent. This simplification will be relaxed in future work to fully exploit the flexibility of the bounding approximation scheme presented in this section.

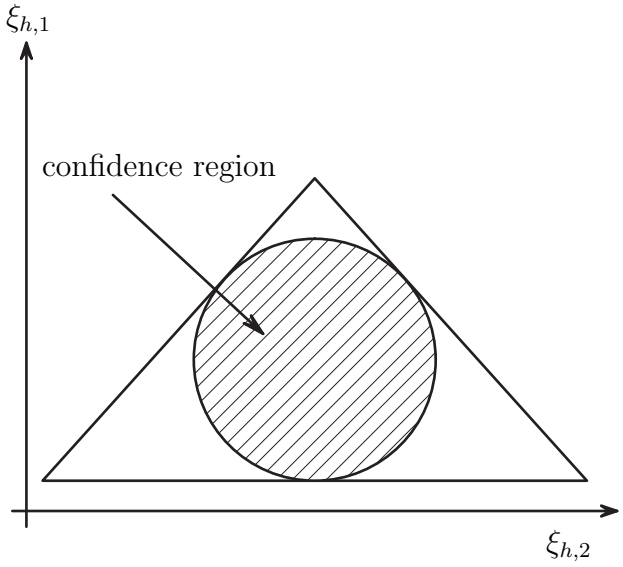

Figure 1: Simplicial covering of a confidence region with high probability mass.

by a compact simplex for each $h = 1, \ldots, H$. Clearly this procedure is problem dependent. A simple approach is to truncate the lognormal distribution of the price relatives outside some bounded confidence region which covers most of the probability mass.[6] Next, we enclose the confidence region of the price relatives in a suitable simplex. A graphical illustration for a problem with $N = 2$ assets is shown in Figure 1. The first space approximations consist of using either the expected values or the extreme points of the simplices as the realizations of the scenario tree, respectively. Solving the expected value problem will yield a lower bound. The problem involving the extreme scenarions (that is, the vertices of the simplices with probabilities estimated from (5.1)) results in an upper bound.

In order to improve the bounds, we have to refine the trivial coverings $\mathcal{Z}^u_{1,h} = \mathcal{Z}^l_{1,h} = \{Z_h\}$ for each $h = 1, \ldots, H$. One possibility to partition a simplex is by choosing an interior point and considering the resulting $N + 1$ sub-simplices. In our implementation we always choose the barycenter of the simplex.[7] For

[6]The truncated distribution must be re-normalized for consistency reasons. Note that the error introduced by truncation and renormalization is generally small and may be disregarded.

[7]Here, we work with the *classical* barycenter corresponding to a uniform mass distribution.

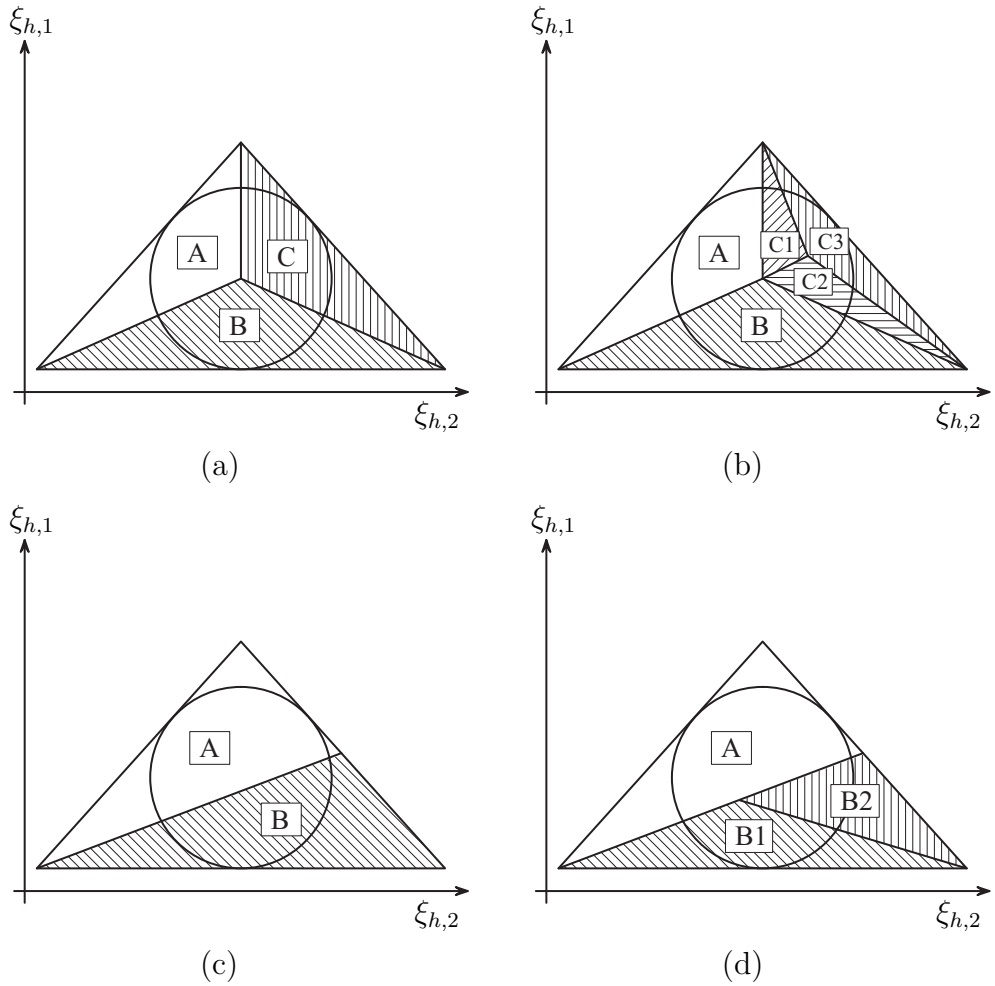

Figure 2: Refinement of simplicial partition.

example, when the simplex in Figure 1 is further refined, the three simplices A,B and C in Figure 2(a) are created. When simplex C is further refined, then three new simplices C1, C2 and C3 are created, see Figure 2(b). As an illustration, we consider a small problem with three assets, and a planning horizon comprising three time periods. The problem is discretized by using the scheme described in Section 5, and the resulting deterministic equivalent problems are solved by using the OOQP solver [11]. All the numerical experiments were performed on a Linux machine with a 2.0Ghz processor and 2GB of RAM. The convergence of the upper and lower bounds due to a naive refinement strategy is shown in

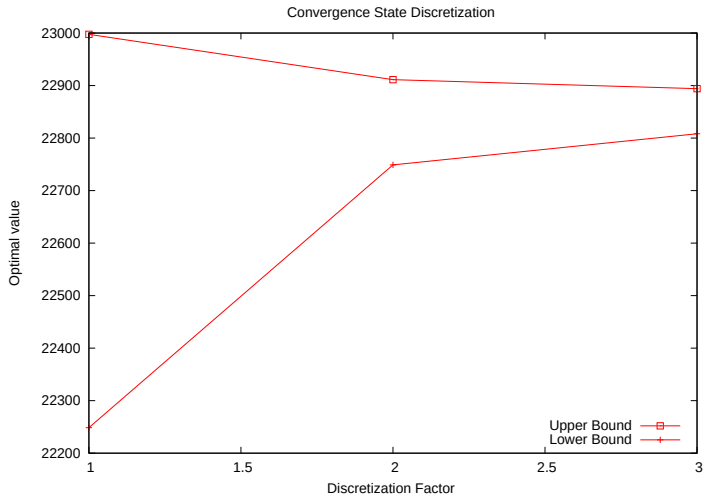

Figure 3: Convergence of bounds due to naive refinement strategy.

Disc. Factor	#Columns	#Rows	# Nodes
1	33	19	3
2	195	91	21
3	2463	1099	2733

Table 1: Problem statistics for lower bound problem, 3 time periods, and 3 assets.

Figure 3. In this example, the discretization factor refers to the depth of recursive simplicial subdivisions. For example, Figure 1 corresponds to the discretization factor $J = 1$, while Figure 2(a) corresponds to $J = 2$. If in Figure 2(b) we further subdivided all the sub-simplices (not just C), then the resulting covering would be assigned the discretization factor $J = 3$. Problems with higher discretization factor cannot be solved due to the fact that problem size increases rapidly with the branching factor of the underlying scenario tree. This *curse of dimensionality* is even more acute in problems with a large number of assets or a large number of time periods. Tables 1 and 2 show how the size of the problem increases as we increase the discretization factor.

As argued at the end of Section 5, we may expect to improve the approxima-

Disc. Factor	#Columns	#Rows	# Nodes
1	195	91	21
2	285	131	31
3	825	371	91

Table 2: Problem statistics for upper bound problem, 3 time periods, and 3 assets.

tion error if the brute force refinement strategy is replaced by a more sophisticated one based on a threshold accepting algorithm. Before we report our initial experience with this advanced approach, we discuss some practical issues relating to its implementation. We first explain how to generate small perturbations of a scenario tree configuration. For a given a configuration we use a Poisson process to randomly select the sub-simplices that will be merged with other sub-simplices. By sampling from another Poisson process we also determine the simplices that will be further partitioned. The use of Poisson processes ensures that certain discretization points will randomly "depart" and "arrive" in the scenario tree. However, the number of nodes of the tree may never exceed a certain budget. Our sampling is designed to make "arrivals" of new discretization points more likely in early decision stages, while "departures" of existing discretization points are more likely in late stages. This accounts for the general consensus in stochastic programming literature that scenario trees should have a higher branching factor in the initial periods [15]. In our implementation the threshold parameter τ is initially set to 100 and reduced by 10% whenever a reduction is required. The threshold is reduced if the scenario tree remains unchanged for three iterations. After five iterations with no change the algorithm terminates.

The main goal of our numerical studies is to find scenario trees that are numerically tractable yet have a small error. Our first numerical experiments show that a threshold accepting scheme can be beneficial. In Figure 4 we show how the algorithm reduces the error as it attempts to find an optimal discretization.

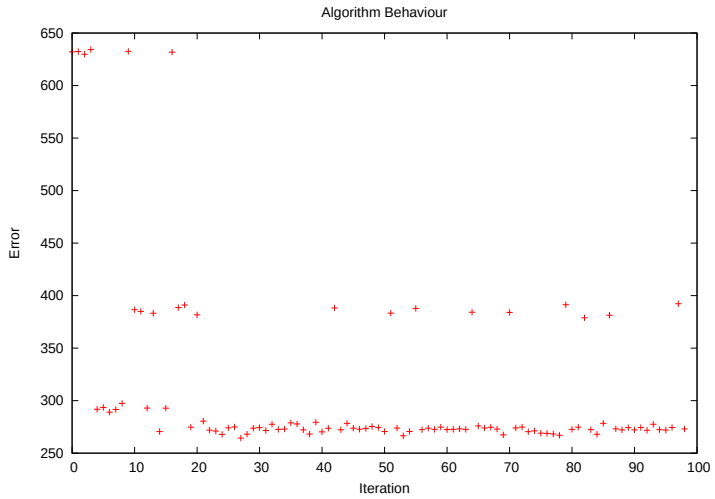

Figure 4: Error reduction by threshold accepting algorithm (barycenter splitting).

The error is reduced dramatically after only a few iterations. From Figure 4 one concludes that for the problem under consideration there exist at least two local minima. The algorithm appears to be able to jump from one cluster of local minima to the next. However, the approximation error never drops below 250 in the first 100 iterations.

The approximation error can be further reduced by using an alternative partitioning strategy. Instead of introducing a new point at the barycenter of an existing simplex, thereby subdividing it into $N + 1$ sub-simplices, we may introduce a new point in the middle of one of its edges.[8] In doing so, the initial simplex is split into two sub-simplices only, see Figures 2(c) and 2(d). "Edge splitting" has the advantage that the diameters of the sub-simplices can be made uniformly small, which is a prerequisite for the theoretical convergence result in Section 5. The "barycenter splitting" technique considered before fails to have this property since edges are never split. However, it produces $N + 1$ sub-simplices per additional vertex, which is superior to the mere two sub-simplices for edge splitting.

[8]We use the standard convention that an edge of a simplex is the line segment between two of its vertices.

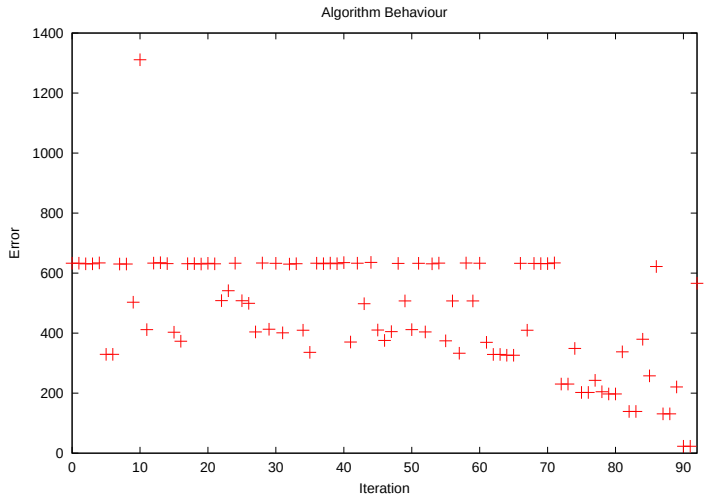

Figure 5: Error reduction by threshold accepting algorithm (edge splitting).

We tested our threshold accepting algorithm also in conjunction with the edge splitting strategy, while randomly selecting the edges to be cut in half. Figure 5 indicates that the approximation error decreases slower if edge splitting is used instead of barycenter splitting. However, the approximation error is virtually eliminated after 90 iterations although tree size does not increase after the tenth iteration.

Our preliminary results suggest that randomized algorithms can be used to find the best bounding scenario tree from a set of trees that do not exceed a certain size. In the future we will allow the simplicial partitions of the state spaces to be path dependent. This increased flexibility will be exploited to achieve a specified error tolerance with even smaller scenario trees and, in addition, to address more complex portfolio problems with many decision stages. Future research will also focus on keeping the time discretization flexible when searching for optimal scenario trees. However, much more work is required in order to find efficient and theoretically appealing methods for solving this problem.

Acknowledgments. Daniel Kuhn thanks the Swiss National Science Foundation for financial support. The authors also wish to acknowledge support from

EPSRC grants GR/T02560/01 and EP/C513584/1.

References

[1] Ash, R. *Real Analysis and Probability.* Probability and Mathematical Statistics. Academic Press, Berlin, 1972.

[2] Birge, J., and Louveaux, F. *Introduction to Stochastic Programming.* Springer-Verlag, New York, 1997.

[3] Dueck, G., and Scheuer, T. Threshold accepting: a general purpose optimization algorithm appearing superior to simulated annealing. *J. Comput. Phys. 90*, 1 (1990), 161–175.

[4] Dupačová (as Žáčková), J. On minimax solutions of stochastic linear programming problems. *Časopis pro Pěstování Matematiky 91* (1966), 423–429.

[5] Edirisinghe, N., and Ziemba, W. Bounding the expectation of a saddle function with application to stochastic programming. *Math. Oper. Res. 19* (1994), 314–340.

[6] Edirisinghe, N., and Ziemba, W. Bounds for two-stage stochastic programs with fixed recourse. *Math. Oper. Res. 19* (1994), 292–313.

[7] Frauendorfer, K. Multistage stochastic programming: Error analysis for the convex case. *Z. Oper. Res. 39*, 1 (1994), 93–122.

[8] Frauendorfer, K. The stochastic programming extension of the Markowitz model. *Int. J. Neural Mass-Parallel Comput. Inform. Syst. 5* (1995), 449–460.

[9] Frauendorfer, K. Barycentric scenario trees in convex multistage stochastic programming. *Math. Program. 75*, 2 (1996), 277–294.

[10] Frauendorfer, K., and Siede, H. Portfolio selection using multi-stage stochastic programming. *Central European J. Oper. Res. 7* (2000), 277–290.

[11] Gertz, E. M., and Wright, S. J. Object-oriented software for quadratic programming. *ACM Trans. Math. Software 29*, 1 (2003), 58–81.

[12] Gülpinar, N., and Rustem, B. Worst-case robust decisions for multi-period mean-variance portfolio optimization. *European J. Oper. Res.* (2006). In press.

[13] Gülpinar, N., Rustem, B., and Settergren, R. Multistage stochastic programming in computational finance. In *Computational Methods in Decision-Making, Economics and Finance* (2002), E. J. Kontoghiorghes, B. Rustem, and S. Siokos, Eds., vol. 74 of *Applied Optimization*, Kluwer Academic Publishers, pp. 33–45.

[14] Gülpinar, N., Rustem, B., and Settergren, R. Optimization and sinmulation approaches to scenario tree generation. *J. Econ. Dyn. Control 28*, 7 (2004), 1291–1315.

[15] Hochreiter, R., and Pflug, G. Financial scenario generation for stochastic multi-stage decision processes as facility location problems. *Ann. Oper. Res. 156*, 1 (2007), 257–272.

[16] Kall, P., and Wallace, S. *Stochastic Programming*. John Wiley & Sons, Chichester, 1994.

[17] Kuhn, D. Aggregation and discretization in multistage stochastic programming. *Math. Program.* (2007). Online First.

[18] Kuhn, D. Convergent bounds for stochastic programs with expected value constraints. *The Stochastic Programming E-Print Series (SPEPS)* (2007).

[19] Madansky, A. Inequalities for stochastic linear programming problems. *Manage. Sci. 6* (1960), 197–204.

[20] Markowitz, H. M. Portfolio selection. *Journal of Finance 7*, 1 (1952), 77–91.

[21] Merton, R. C. *Continuous Time Finance*. Basil Blackwell, Cambridge, MA, 1992.

[22] Rockafellar, R., and Wets, R.-B. The optimal recourse problem in discrete time: L^1-multipliers for inequality constraints. *SIAM J. Control Optimization 16* (1978), 16–36.

[23] Rosenblatt, M. Remarks on a multivariate transformation. *Ann. Math. Stat. 23*, 3 (1952), 470–472.

[24] Steinbach, M. C. Markowitz revisited: mean-variance models in financial portfolio analysis. *SIAM Rev. 43*, 1 (2001), 31–85.

[25] Wright, S. Primal-dual aggregation and disaggregation for stochastic linear programs. *Math. Oper. Res. 19*, 4 (1994), 893–908.

[26] Zhou, X., and Li, D. Continuous time mean-variance portfolio selection: A stochastic LQ framework. *Appl. Math. Optim. 42* (2000), 19–33.