Barycentric Bounds in Stochastic Programming: Theory and Application

Karl Frauendorfer[a], Daniel Kuhn[b], and Michael Schürle[a]

[a]University of St. Gallen, St. Gallen, Switzerland.
[b]Imperial College of Science, Technology and Medicine, London, UK.

June 16, 2009

Abstract. The design and analysis of efficient approximation schemes is of fundamental importance in stochastic programming research. Bounding approximations are particularly popular for providing strict error bounds that can be made small by using partitioning techniques. In this article we develop a powerful bounding method for linear multistage stochastic programs with a generalized nonconvex dependence on the random parameters. Thereby, we establish bounds on the recourse functions as well as compact bounding sets for the optimal decisions. We further demonstrate that our bounding methods facilitate the reliable solution of important real-life decision problems. To this end, we solve a stochastic optimization model for the management of non-maturing accounts and compare the bounds on maximum profit obtained with different partitioning strategies.

Key words. stochastic programming, barycentric approximation scheme, bounds, scenario tree

1 Introduction

Multistage stochastic programs arise whenever sequential decisions have to be taken under incomplete information about some decision-relevant parameters. The general setup is as follows: a decision maker observes some random parameter and selects a decision based on this first observation. Next, a new random parameter is observed, in response to which a recourse decision is taken. The second decision may depend on both previous observations (and on the first decision, which adds no extra information, however, since it is merely a function of the first observation). This basic scheme of alternating observations and decisions is continued until the end of the planning horizon. Typically, decisions cannot be chosen freely but are subject to constraints of physical or regulatory nature. Moreover, rationality requires decision makers to select utility maximizing actions. When formulating a stochastic optimization model, thus, a suitable objective criterion must be specified. In order to simplify prose, we will henceforth assume that the objective is to maximize expected profit.

A stochastic optimization problem is said to be linear if the objective and constraint functions are linear affine in the decision variables, which will always

1

be assumed in the remainder of this article. Linear two- and multistage stochastic programs were first studied by Dantzig [9] and Beale [3] in 1955. Their immense popularity originates from a rich variety of applications in the fields of management, engineering, finance, and logistics, etc. If the underlying random variables are finitely supported, the extensive form [5] of a given stochastic program represents a large-scale linear program, which can in principal be solved by using Dantzig's simplex algorithm [10]. However, if the supports of the underlying random parameters are very large or (un)countably infinite, a stochastic program will in most cases allow only for an approximate solution.

The present article addresses an approximation scheme based on discretization. Concretely speaking, we will establish two auxiliary stochastic programs which emerge from the original problem by replacing the true distribution of the random parameters by two finitely supported approximate distributions. After a suitable transformation offsetting possible nonconvexities in the random parameters, the optimal values of the auxiliary stochastic programs can be shown to provide bounds for the optimal value of the original problem. Moreover, repeated solution of the auxiliary problems under different parameter settings will allow us to construct compact bounding sets for the true optimal decisions. Further emphasis will be put on convergence issues.

Our approximation scheme belongs to the class of bounding methods, which have a long tradition in stochastic optimization; see, e.g., [5, Chap. 9] for a textbook introduction. In developing so-called bounding probability measures, one typically exploits structural properties of the recourse functions (profit-to-go functions) associated with the underlying stochastic program. Frequently, bounding measures may be interpreted as solutions of generalized moment problems [7, 19, 20, 25, 41]. This idea is credited to Dupačová [16] and has fruitful applications also in models where the underlying probability distribution is only known in limited manner [13, 14, 15].

When the recourse functions are convex in the random parameters, a lower bound is found via Jensen's inequality [40], while an upper bound arises from the Edmundson-Madansky inequality.[1] Edmundson [21] treats the univariate case, whereas Madansky [44, 45] and Frauendorfer [24] consider the multivariate setting, given that the components of the random parameters are independent and dependent, respectively. Elaborate extensions are due to Gassmann and Ziemba [35] and Birge and Wets [6, 7]. Note that these bounds are based only on first-order moment information. One way of tightening them is by inclusion of higher order information in the construction of bounding measures. Edirisinghe [17] was the first to develop second order lower bounds, which were later extended by Dokov and Morton [12]. Birge and Dulá [4], Dupačová [16], and Kall [42] propose second order upper bounds. Higher order upper bounds are suggested in [11]. For more information on higher order bounds, see also the chapter in

---

[1]In the presence of concave recourse functions upper and lower bounds switch roles.

this book by Edirisinghe []. All bounds discussed so far can be improved to an arbitrary degree of precision by applying them on increasingly small subsets of the underlying domain. This so-called *partitioning* technique is documented in [6, 30, 39].

Similar bounding methods are also available for problems whose recourse functions are convex-concave saddle functions. First order bounds closely related to those of Jensen and Edmundson-Madansky are proposed by Frauendorfer [25] and Edirisinghe and Ziemba [19, 20]. Second order bounds on saddle functions are due to Edirisinghe [17], and generalized bounds for specific problems with nonconvex recourse functions are developed by Kuhn [43]. Furthermore, suitable partitioning schemes are discussed in [18, 25, 43].

The present article reviews some cornerstone results related to bounding techniques in stochastic linear programming and presents an illustrative application to finance. Our theoretical exposition follows the lines of [25, 26, 27, 43]. Section 2 formally introduces the class of optimization problems to be studied, while Sect. 3 elaborates a set of regularity conditions ensuring well-definedness of our mathematical models. In Sect. 4 we sketch the construction of discrete approximate distributions for the random parameters; these so-called *barycentric measures* facilitate the calculation of bounds on the expectations of subdifferentiable saddle functions. The theoretical main results are presented in Sect. 5. Most importantly, we will demonstrate the construction of tight bounds on the recourse functions and establish compact bounding sets for the optimal decisions of linear stochastic programs. Furthermore, we will report on convergence results. In order to illustrate the practical implementation of the theoretical concepts, Sect. 6 discusses the problem of reinvesting money from non-maturing accounts (for example savings products), which is a major concern for financial institutions. This particular decision problem is formulated as a linear multistage stochastic program and can be analyzed with our bounding methods. Numerical experiments show that the bounds on the optimal objective value can be made tight with affordable computational effort. In a series of test calculations, we investigate the convergence of the bounds for different families of scenario trees.

2 Problem Formulation

In this article we consider a class of constrained profit maximization problems under uncertainty. We assume that decisions may be selected at several time points $t = 1, \ldots, T$. At the outset, we must establish a probabilistic model for the underlying uncertainty. All random objects are defined on an abstract probability space (Ω, Σ, P). Adopting the standard terminology of probability theory, we will refer to Ω as the *sample space*. Furthermore, we use the following definition of a stochastic process.

Definition 2.1 (Stochastic Process). *We say that $\boldsymbol{\xi}$ is a stochastic process with state space Ξ if $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_T)$ and $\Xi = \times_{t=1}^{T} \Xi_t$ such that each random vector $\boldsymbol{\xi}_t$ maps (Ω, Σ) to the Borel space $(\Xi_t, \mathcal{B}(\Xi_t))$ and each Ξ_t is a convex closed subset of some finite-dimensional Euclidean space. Moreover, we define combined random vectors $\boldsymbol{\xi}^t := (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_t)$ valued in $\Xi^t := \times_{\tau=1}^{t} \Xi_\tau$ for all $t = 1, \ldots, T$.[2]*

As a notational convention, throughout this article, random objects will be represented in boldface, while their realizations will be denoted by the same symbols in normal face. We will frequently encounter stochastic processes with compact state spaces. Then, the corresponding random vectors are bounded, thus having finite moments of all orders.

Let $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ be two stochastic processes in the sense of definition 2.1 with state spaces Θ and Ξ, respectively. These processes are assumed to describe the random data which is revealed sequentially in the decision-making process. Thus, the information $\mathcal{F}^t$ available at stage t amounts to the σ-algebra generated by the random variables observed in stages 1 through t, i.e., $\mathcal{F}^t = \sigma(\boldsymbol{\eta}^t, \boldsymbol{\xi}^t)$. With these conventions, the space of non-anticipative *decision processes*[3] is defined as

$$\mathcal{N} = \{\boldsymbol{x} \text{ is a stochastic process with state space } X \text{ such}$$
$$\text{that } \boldsymbol{x}_t \in \mathcal{L}^\infty(\Omega, \mathcal{F}^t, P; X_t) \text{ for each } t = 1, \ldots, T\}.$$

Throughout this article we will assume that $X = \mathbb{R}^n$ and $X_t = \mathbb{R}^{n_t}$. Then, consistency requires that $n = n_1 + \cdots + n_T$. In order to simplify notation, we introduce $\boldsymbol{x}_0$ as a fictitious fixed decision of the past. The *static version* of a linear multistage stochastic program can now be formulated as

$$\underset{\boldsymbol{x} \in \mathcal{N}}{\text{maximize}} \, E\left[\sum_{t=1}^{T} \langle c_t^*(\boldsymbol{\eta}^t), \boldsymbol{x}_t \rangle \right] \tag{2.1}$$
$$\text{s.t.} \quad W_t(\boldsymbol{\xi}^t)\boldsymbol{x}_t + T_t(\boldsymbol{\xi}^t)\boldsymbol{x}_{t-1} \sim_t h_t(\boldsymbol{\xi}^t) \quad P\text{-a.s.} \quad t = 1, \ldots, T.$$

The objective and constraint functions are linear in the decision variables. However, linearity in the stochastic parameters is not required. Note that the process $\boldsymbol{\eta}$ exclusively appears in the objective function, whereas $\boldsymbol{\xi}$ only affects the dynamic constraints (but both $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ influence the set $\mathcal{N}$ accounting for the so-called non-anticipativity constraints). As argued in [25, Sect. 5], a distinction between $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ is always possible. However, some components of $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ may represent the same random parameter and must therefore be modelled as perfectly correlated random variables.[4]

---

[2]Sometimes, notation is simplified by further introducing a deterministic dummy random variable $\boldsymbol{\xi}^0$ of the past.

[3]Decision processes will also be referred to as *strategies* or *policies*.

[4]Random variables appearing both in the objective function and the constraints of the stochastic program must be duplicated. The first copy of such a random variable is appended to $\boldsymbol{\eta}$, while the second copy is appended to $\boldsymbol{\xi}$. Note that the two copies will be discretized differently.

We choose to work explicitly with inequality (less-or-equal, greater-or-equal) and equality constraints. Thus, for every stage we introduce an r_t-dimensional 'vector' $\sim_t$ of binary relations, each of whose entries is either '$\leq$' or '$\geq$' for inequalities, or '$=$' for equalities. Of course, the stochastic program (2.1) can be brought to the standard form with only 'less-or-equal' constraints if the equality constraints are replaced by two opposing inequality constraints, and the 'greater-or-equal' constraints are multiplied by -1. Notice that the constraints in (2.1) only couple neighboring decision stages. This can always be enforced, i.e., dependencies across more than one decision stage can systematically be eliminated by introducing additional decision variables. For every $t = 1, \ldots, T$, the $r_t \times n_t$ matrix W_t is termed *recourse matrix* and may generally depend on ξ^t. The $r_t \times n_{t-1}$ matrix T_t is referred to as *technology matrix* in literature. Obviously, the technology matrix determines the intertemporal coupling and may also depend on ξ^t. Moreover, the *right hand side (rhs) vector* h_t and the vector of *objective function coefficients* c_t^* may depend on ξ^t and η^t, respectively.

For both theoretical and practical purposes it is sometimes more comfortable to work with the *dynamic version* of a given multistage stochastic program which relies on a backward recursion scheme. Setting $\Phi_{T+1} \equiv 0$, the optimal value functions or *recourse functions* of stages $t = T, \ldots, 1$ are defined as

$$\Phi_t(x^{t-1}, \eta^t, \xi^t) = \max_{x_t \in \mathbb{R}^{n_t}} \langle c_t^*(\eta^t), x_t \rangle + (E_t \Phi_{t+1})(x^t, \eta^t, \xi^t) \qquad (2.2)$$
$$\text{s.t.} \quad W_t(\xi^t)x_t + T_t(\xi^t)x_{t-1} \sim_t h_t(\xi^t).$$

Thereby, the expected recourse functions or *expectation functionals* are given by

$$(E_t \Phi_{t+1})(x^t, \eta^t, \xi^t) = \int \Phi_{t+1}(x^t, \eta^{t+1}, \xi^{t+1}) \, dP_{t+1}(\eta_{t+1}, \xi_{t+1} | \eta^t, \xi^t), \qquad (2.3)$$

and P_{t+1} denotes a regular conditional probability distribution of the random parameters observed at stage $t+1$ given the realizations of the random parameters observed at stages 1 through t. In real-life applications it is usually assumed that $\boldsymbol{\eta}_1$ and $\boldsymbol{\xi}_1$ are deterministic, i.e., they represent degenerate random vectors with a Dirac distribution (as was assumed for the fictitious past decision $\boldsymbol{x}_0$). Then, the first stage recourse function Φ_1 needs be evaluated only at one point, where it coincides with the optimal value of the underlying stochastic program.

To facilitate the formulation of precise statements about dynamic stochastic programs, it proves useful to introduce specific subsets of parameter space. For each $t = 1, \ldots, T$ we define

$$Z^t = \{(x^{t-1}, \eta^t, \xi^t) \in X^{t-1} \times \Theta^t \times \Xi^t \, | \, \forall s = 1, \ldots, t-1 :$$
$$W_s(\xi^s)x_s + T_s(\xi^s)x_{s-1} \sim_s h_s(\xi^s) \, \}. \qquad (2.4)$$

Points in the complement of Z^t correspond to impossible realizations of the random parameters or to forbidden decisions. Thus, only the restriction of Φ_t to

Z^t has physical meaning, and we will refer to Z^t as the *natural domain* of the recourse function of stage t.

Without suitable regularity conditions, it is not clear whether the static and dynamic versions of a given multistage stochastic program are well-defined and solvable (as is suggested by the use of the 'max'-operators). In particular, measurability of the integrands in (2.1) and (2.3) must be ensured. To this end, the next section will elaborate and discuss a set of suitable regularity conditions which guarantee well-definedness, solvability, and equivalence of the static and dynamic versions of a linear multistage stochastic program.

3   Regularity Conditions

In order to establish a set of concise regularity conditions which ensure that the linear stochastic program given by representation (2.1) or (2.2) is well-behaved, we have to introduce some terminology. A major concern is usually the functional dependence of the objective and rhs vectors on the stochastic parameters. Modelling c_t^* and h_t as differences of convex functions turns out to be particularly advantageous. Let us therefore state the following formal definition.

Definition 3.1. *Let Ξ be a convex subset of a finite-dimensional Euclidean space. A function $f : \Xi \to \mathbb{R}^r$ is called d.c. (abbreviation for difference of convex functions) if there are two componentwise convex functions κ^+ and κ^- such that*

$$f(\xi) = \kappa^+(\xi) - \kappa^-(\xi) \quad \forall \xi \in \Xi.$$

Notice that the decomposition of a d.c. function is never unique. In fact, by adding the same convex mapping to both κ^+ and κ^-, one obtains a valid alternative decomposition. Recently, d.c. functions have experienced considerable attention in the field of global optimization [22, 38, 46]. A survey of their properties is provided in [37]. Here, we recall only a few properties relevant in our context. First, the class of twice continuously differentiable functions is a linear subspace of the space of d.c. functions. This can easily be proved if Ξ is compact. However, the statement remains true for Ξ open or unbounded, as pointed out by Hartman [36]. Moreover, if Ξ is compact, the d.c. functions are dense in the space of continuous mappings from Ξ to $\mathbb{R}^r$ endowed with the topology of uniform convergence with respect to the Euclidean norm in $\mathbb{R}^r$. This follows directly from the Stone-Weierstrass theorem and the fact that all vector-valued polynomials are d.c. functions.

Well-definedness of the integrands in (2.1) and (2.3) not only depends on the properties of the objective and rhs vectors (and the constraint matrices), but also essentially on the properties of the stochastic data process.

Definition 3.2. *We say that the random data $(\boldsymbol{\eta}, \boldsymbol{\xi})$ follows a block-diagonal autoregressive process if it is driven by two serially independent stochastic processes*

ε^o and ε^r with state spaces $\mathcal{E}^o$ and $\mathcal{E}^r$, respectively. Moreover, for each t there are two matrices H_t^o and H_t^r with appropriate dimensions such that

$$\left[\begin{array}{c} \boldsymbol{\eta}_t \\ \boldsymbol{\xi}_t \end{array} \right] = \left[\begin{array}{cc} H_t^o & 0 \\ 0 & H_t^r \end{array} \right] \left[\begin{array}{c} \boldsymbol{\eta}^{t-1} \\ \boldsymbol{\xi}^{t-1} \end{array} \right] + \left[\begin{array}{c} \varepsilon_t^o \\ \varepsilon_t^r \end{array} \right], \quad t = 1, \ldots, T.$$

Block-diagonal autoregressive processes exhibit a linear dependence on history. Moreover, the AR coefficient matrices of all stages are block-diagonal. Note that definition 3.2 allows the noise processes ε^o and ε^r to be correlated. In the extreme case where the same random processes affect the objective and the constraint functions, we are obliged to set $\varepsilon^o \equiv \varepsilon^r$ (thereby doubling the dimension of the relevant state space, which is computationally expensive but not necessarily prohibitive).

In the remainder of this article we will study linear multistage stochastic programs of the form (2.1) which satisfy the following regularity conditions:

(C1) the marginal spaces Θ_t and Ξ_t are compact regular simplices, $t = 1, \ldots, T$;

(C2) the vector of objective function coefficients c_t^* is continuous and d.c. on a convex neighborhood of Θ^t, $t = 1, \ldots, T$;

(C3) the rhs vector h_t is continuous and d.c. on a convex neighborhood of Ξ^t, the matrices W_t and T_t are independent of the random parameters, and the recession cone $\{x_t | W_t x_t \sim_t 0\}$ is given by $\{0\}$, $t = 1, \ldots, T$;

(C4) the random data $(\boldsymbol{\eta}, \boldsymbol{\xi})$ follows a block-diagonal autoregressive process;

(C5) at any reference point in Z^t the parametric maximization problem (2.2) has a feasible point where the gradients of the active constraint functions are linearly independent, $t = 1, \ldots, T$.

The first condition requires the state spaces of the random parameters to be compact, which implies the data processes to have finite moments of all orders. Consequently, (C1) constitutes a restrictive condition. Assumption (C2), however, is nonrestrictive since the set of d.c. functions is dense in the set of continuous functions. Therefore, we have much flexibility in modelling the functional form of c_t^*. Analogously, condition (C3) offers considerable flexibility in choosing the rhs vector h_t, but it is restrictive in that it requires the recourse and technology matrices to be nonrandom and the feasible sets of the parametric maximization problems (2.2) to be compact. Assumption (C4) is certainly restrictive as it only allows for linear dependencies between random parameters of different stages. But this is not a serious deficiency. In fact, it is frequently possible to absorb all nonlinearities in the definition of the functions c_t^* and h_t. However, requiring these functions to be linear affine, as is frequently done, and, at the same time,

assuming the random parameters to follow a block-diagonal autoregressive process, would severely limit the scope of our methodology. Assumption (C5) has the character of a generalized Slater condition and implies non-anticipativity of the constraint multifunction in the sense of Rockafellar and Wets [47]. This condition is nonrestrictive, since any constraint multifunction can be made non-anticipative by explicitly introducing the so-called *induced constraints*, see, e.g., [48, 51]. Notice that, as opposed to popular strict feasibility conditions, our assumption (C5) is not in conflict with the presence of equality constraints.

Some basic consequences of the above regularity conditions are summarized in the following theorem.

Theorem 3.3. *Under the assumptions (C1)–(C5) the static and dynamic versions of a linear multistage stochastic program are both well-defined and solvable, and the optimal values coincide. Moreover, the recourse functions are finite and continuous on their natural domains.*

Proof. The assumptions (C1)–(C5) imply the weaker conditions used by Rockafellar and Wets [47]. Thus, the claim follows from [47, Theorem 1]. Alternatively, see [43, Proposition 2.5 and Theorem 2.6]. □

4 Barycentric Probability Measures

The solution of stochastic programs poses severe difficulties, especially in the multistage case. If the underlying probability measure is absolutely continuous with respect to Lebesgue measure, the static version of a stochastic program represents an optimization problem over an infinite-dimensional function space. Then, analytical solutions are available only for simple models of questionable practical relevance. Analytical treatment of the dynamic version of a stochastic program is no less challenging. Instead of a single optimization problem over a function space one faces a sequence of nested optimization problems over finite-dimensional Euclidean spaces. Evaluation of the expectation functionals is particularly involved: it requires multivariate integration of a function which is only known implicitly as the result of a subordinate parametric optimization problem.

Numerical solutions are usually based on discretization of the underlying probability space. The standard approach is to solve the stochastic program with respect to a finitely supported auxiliary measure instead of the original measure. Thereby, one effectively approximates the original stochastic program by an auxiliary optimization problem over a finite-dimensional space, which is numerically tractable. The auxiliary probability measure should approximate the original measure in a specific sense, i.e., it should be designed so as to guarantee that the optimal value and the solution set of the auxiliary problem are close to the optimal value and the solution set of the original stochastic program, respectively. Thereby, distance of optimal values is measured with respect to the Euclidean

metric on the real line, while distance of the solution sets is measured, e.g., with respect to the Pompeiu-Hausdorff metric. In this sense, proximity of the auxiliary and the original probability measures depends on the underlying stochastic program.

The selection of an appropriate discrete probability measure is referred to as *scenario tree construction* and represents a primary challenge in the field of stochastic programming. Our scenario tree construction method is based on the following procedure, which stems from [27, Sect. 3]. As usual, we let P_t be a regular conditional probability distribution of the stochastic parameters observed at stage t given the history of the stochastic parameters observed at stages 1 through $t-1$. Suppose that for each history of realizations $(\eta^{t-1}, \xi^{t-1}) \in \Theta^{t-1} \times \Xi^{t-1}$ a discrete measure $P_t^d(\cdot|\eta^{t-1}, \xi^{t-1})$ approximates the true conditional probability measure $P_t(\cdot|\eta^{t-1}, \xi^{t-1})$ on $\mathcal{B}(\Theta_t \times \Xi_t)$. Furthermore, assume that the assignment

$$P_t^d : \begin{cases} \mathcal{B}(\Theta_t \times \Xi_t) \times \Theta^{t-1} \times \Xi^{t-1} \to [0,1] \\ (B, \eta^{t-1}, \xi^{t-1}) \mapsto P_t^d(B|\eta^{t-1}, \xi^{t-1}) \end{cases}$$

characterizes a *transition probability*, i.e., it satisfies the following conditions:

(i) $P_t^d(\cdot|\eta^{t-1}, \xi^{t-1})$ is a probability measure on $\mathcal{B}(\Theta_t \times \Xi_t)$ for any fixed outcome history $(\eta^{t-1}, \xi^{t-1}) \in \Theta^{t-1} \times \Xi^{t-1}$;

(ii) $P_t^d(B|\cdot)$ is a Borel measurable function on $\Theta^{t-1} \times \Xi^{t-1}$ for every fixed Borel subset B of $\Theta_t \times \Xi_t$.

By the product measure theorem [1, Sect. 2.6], the transition probabilities P_t^d of all stages can be nested to form a unique probability measure P^d on the measurable space $(\Theta \times \Xi, \mathcal{B}(\Theta \times \Xi))$. Under suitable conditions, the discrete measure P^d then approximates the original measure P with respect to the underlying stochastic optimization problem. The above reasoning implies that we should first focus on the discretization of the conditional distributions $P_t(\cdot|\eta^{t-1}, \xi^{t-1})$. In a second step, the transition probabilities must be combined to form a discrete scenario tree.

In any efficient scenario tree construction method, the choice of discrete transition probabilities should account for the structural properties of the underlying stochastic program. Our approach exploits distinct convexity properties of the recourse functions (2.2), which follow from the prevailing regularity conditions. In fact, the assumptions (C1)–(C5) imply that the recourse functions are concave in the decisions and d.c. in the stochastic variables on a neighborhood of their natural domains [43, Chap. 5]. By adding suitable correction terms, which may depend on the random parameters but not on the decision variables, the recourse functions can be transformed to saddle functions being convex in η and

9

jointly concave in x and ξ.[5] Thus, evaluation of the expectation functionals is intimately related to calculating the expected value of a saddle function on a compact domain.

Approximating the conditional distributions of the random parameters in a multistage stochastic program is basically equivalent to approximating the unconditional distribution of the random parameters in a one-stage problem. Therefore, we may temporarily omit time indices and suppress any dependencies on the outcome and decision history. Our task now reduces to approximating the joint distribution P of two random vectors $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ by a discrete probability measure P^d. Recall also that $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are valued in compact simplices Θ and Ξ, respectively. Having in mind the remarks of the previous paragraph, we try to establish a definite relation between the expectations of some (a priori unknown) saddle function $\Phi(\eta, \xi)$ with respect to the complementary measures P and P^d, assuming only that Φ is convex in η, concave in ξ, and subdifferentiable. This problem has been extensively studied in the monograph [25]; see also the related work in [17, 18, 19, 20]. A promising approach is via moment problems, as will be outlined below. Let $\mathcal{P}(m)$ be the set of all Borel probability measures on $\Theta \times \Xi$ which have the same first-order and second-order cross-moments m as the original measure P, i.e.,

$$
m = \int_{\Theta \times \Xi} \begin{bmatrix} 1 \\ \eta \end{bmatrix} \begin{bmatrix} 1 \\ \xi \end{bmatrix}^\top dP(\eta, \xi).
$$

Next, choose a measure P^l from the minimizer set of

$$
\inf_{Q \in \mathcal{P}(m)} \int_{\Theta \times \Xi} \Phi(\eta, \xi) \, dQ(\eta, \xi). \tag{4.1}
$$

Note that P^l exists since Φ is subdifferentiable and saddle-shaped [25, Chap. 3]. Furthermore, choose a measure P^u from the maximizer set of the symmetric problem

$$
\sup_{Q \in \mathcal{P}(m)} \int_{\Theta \times \Xi} \Phi(\eta, \xi) \, dQ(\eta, \xi), \tag{4.2}
$$

which is solvable for the same reasons as (4.1). By construction, we find the following chain of inequalities for the expected values of the saddle function Φ with respect to the three measures under consideration.

$$
\int_{\Theta \times \Xi} \Phi(\eta, \xi) \, dP^l(\eta, \xi) \leq \int_{\Theta \times \Xi} \Phi(\eta, \xi) \, dP(\eta, \xi) \leq \int_{\Theta \times \Xi} \Phi(\eta, \xi) \, dP^u(\eta, \xi) \tag{4.3}
$$

---

[5]If the objective function coefficients and the rhs vectors are linear affine functions of the stochastic parameters, then the recourse functions exhibit a saddle shape themselves, and no correction terms are needed. This situation is investigated in [26, 27].

An elegant duality argument shows that P^l and P^u can be chosen to be discrete, while depending solely on the matrix m of cross-moments and the geometry of the state space $\Theta \times \Xi$.[6] This implies that for fixed distributions P, P^l, and P^u, the estimate (4.3) is universally valid for all subdifferentiable saddle functions, and not just for the specific function Φ. Universality is a very useful feature, since the recourse functions of a given multistage stochastic program are a priori unknown. Note that universality may be lost if we naively attempt to match higher order moments in the semi-infinite linear programs (4.1) and (4.2). However, universality can be restored if one changes the shape of the underlying domain of the random variables, see [17, 18].

Loosely speaking, the discrete measure P^l concentrates probability mass at the barycenter of Θ and at the extreme points of Ξ, while P^u concentrates probability mass at the barycenter of Ξ and at the extreme points of Θ. Therefore, we will refer to P^l and P^u as lower and upper *barycentric probability measures*. Although we have established a definite relation between the original and the barycentric probability measures, it is not clear whether the inequalities in (4.3) are tight. If not, the probability measure P can be partitioned into smaller pieces, i.e., it may be represented as a convex combination of specific probability measures with smaller supports. Then, the barycentric measures are constructed for each component separately, and their convex combinations provide improved estimates in (4.3). By successively increasing the number of components in a partition, one can construct two sequences of refined barycentric probability measures $\{P_J^l\}_{J \in \mathbb{N}_0}$ and $\{P_J^u\}_{J \in \mathbb{N}_0}$, where the integer J indices the current partition and will be called the *refinement parameter*. It can be shown that both sequences converge weakly to the original measure P if the supports of all components become uniformly small for large J. This will always be assumed in the reminder of this article; for details see [27, 43].

Let us now return to the multistage case. The recipe of the previous paragraph can be used to approximate the conditional probability measure $P_t(\cdot|\eta^{t-1}, \xi^{t-1})$ by lower and upper barycentric measures $P_{J,t}^l(\cdot|\eta^{t-1}, \xi^{t-1})$ and $P_{J,t}^u(\cdot|\eta^{t-1}, \xi^{t-1})$, which may depend on the refinement parameter, the stage index, and the outcome history. We assume that all lower (upper) barycentric measures have the same number of atoms for fixed values of J and t. This is no major restriction since any atom can be viewed as a group of collapsed single atoms. If the probability masses and the coordinates of the discretization points are measurable functions of the outcome history (η^{t-1}, ξ^{t-1}), then $P_{J,t}^l$ and $P_{J,t}^u$ satisfy the defining conditions of a transition probability. By the product measure theorem [1, Sect. 2.6], the barycentric transition probabilities $\{P_{J,t}^l\}_{t=1}^T$ ($\{P_{J,t}^u\}_{t=1}^T$) can be combined to form a unique barycentric probability measure P_J^l (P_J^u) on the joint state space of the stochastic processes $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$. It can be shown that the inequalities (4.3)

---

[6]Analytical formulae for the masses and coordinates of the atoms of P^l and P^u are provided in [25]; cf. also [26, 27, 43].

still hold in the multistage case for any subdifferentiable saddle function Φ being convex in $\eta = (\eta_1, \ldots, \eta_T)$ and concave in $\xi = (\xi_1, \ldots, \xi_T)$. Moreover, the overall barycentric measures converge weakly to the true probability measure P as the refinement parameter J tends to infinity [43, Proposition 4.4].

For each fixed $J \in \mathbb{N}_0$, the barycentric transition probabilities can be used to establish two sequences of *auxiliary recourse functions*, which will be shown to approximate the true recourse functions (2.2). Set $\Phi^l_{J,T+1} \equiv \Phi^u_{J,T+1} \equiv 0$, and define

$$\Phi^l_{J,t}(x^{t-1}, \eta^t, \xi^t) = \max_{x_t \in \mathbb{R}^{n_t}} \langle c^*_t(\eta^t), x_t \rangle + (E^l_{J,t}\Phi^l_{J,t+1})(x^t, \eta^t, \xi^t)$$
$$\text{s.t.} \quad W_t(\xi^t)x_t + T_t(\xi^t)x_{t-1} \sim_t h_t(\xi^t) \tag{4.4a}$$

$$\Phi^u_{J,t}(x^{t-1}, \eta^t, \xi^t) = \max_{x_t \in \mathbb{R}^{n_t}} \langle c^*_t(\eta^t), x_t \rangle + (E^u_{J,t}\Phi^u_{J,t+1})(x^t, \eta^t, \xi^t)$$
$$\text{s.t.} \quad W_t(\xi^t)x_t + T_t(\xi^t)x_{t-1} \sim_t h_t(\xi^t) \tag{4.4b}$$

in a backward manner for $t = T, \ldots, 1$. Thereby, the (auxiliary) expectation functionals are constructed in the obvious way with the help of the transition probabilities $P^l_{J,t+1}$ and $P^u_{J,t+1}$.

$$(E^l_{J,t}\Phi^l_{J,t+1})(x^t, \eta^t, \xi^t) = \int \Phi^l_{J,t+1}(x^t, \eta^{t+1}, \xi^{t+1}) \, dP^l_{J,t+1}(\eta_{t+1}, \xi_{t+1} | \eta^t, \xi^t)$$

$$(E^u_{J,t}\Phi^l_{J,t+1})(x^t, \eta^t, \xi^t) = \int \Phi^u_{J,t+1}(x^t, \eta^{t+1}, \xi^{t+1}) \, dP^u_{J,t+1}(\eta_{t+1}, \xi_{t+1} | \eta^t, \xi^t)$$

Note that the set Z^t defined in Sect. 3 can be interpreted as the natural domain of both $\Phi^l_{J,t}$ and $\Phi^u_{J,t}$. Moreover, unlike the true recourse function Φ_t, the auxiliary recourse functions are numerically computable. Their calculation either relies on the solution of a finite-dimensional linear program, which can, e.g., be solved by using the classical simplex algorithm [10], or on some specialized decomposition schemes [5, Sect. 7.1].

5   Bounds for Stochastic Programs

In this section we will argue that — after a suitable transformation — the auxiliary recourse functions provide bounds on the true recourse functions and can be used to construct bounding sets for the optimal decisions. Furthermore, we will discuss convergence issues.

As usual, assume the regularity conditions (C1)–(C5) to hold. If, beyond that, the objective coefficients and the rhs vectors are linear functions of the random parameters, then the recourse functions are subdifferentiable and saddle-shaped [26, Sect. 2], while the auxiliary recourse functions can be shown to bracket the true recourse functions on their natural domains. Mathematically speaking, this translates to

$$\Phi^l_{J,t} \leq \Phi_t \leq \Phi^u_{J,t} \quad \text{on } Z^t, \, t = 1, \ldots, T, \, J \in \mathbb{N}. \tag{5.1}$$

12

Frauendorfer proved the inequalities (5.1) by backward induction with respect to t [26, Theorem 4.1 and Lemma 4.2], using subdifferentiability and the saddle structure of the recourse functions and the universal bounding property (4.3) of the barycentric transition probabilities. If linearity of the objective and rhs vectors is abandoned, as is necessary in certain applications, the relation (5.1) fails to hold in general. Therefore, we will present a stronger result in this article, which remains applicable if c_t^* and h_t are generic d.c. functions. In fact, we will argue that bounds on Φ_t can generally be expressed in terms of the auxiliary recourse functions shifted by specific correction terms, which have an intuitive structure and are numerically accessible.

In order to characterize these correction terms, we have to introduce some additional notation. Recall first that, by assumption (C2), the vector-valued function c_t^* is d.c. Thus, there are two convex mappings κ_t^{*+} and κ_t^{*-} on a closed convex neighborhood of Θ^t such that

$$c_t^* = \kappa_t^{*+} - \kappa_t^{*-}, \quad t = 1, \ldots, T. \tag{5.2}$$

Similarly, condition (C3) stipulates that h_t is d.c. We may thus suppose that there are two convex mappings κ_t^+ and κ_t^- on a closed convex neighborhood of Ξ^t with

$$h_t = \kappa_t^+ - \kappa_t^-, \quad t = 1, \ldots, T. \tag{5.3}$$

Next, let us return to the dynamic version (2.2) of the given multistage stochastic program. It should be emphasized again that this representation explicitly deals with inequality (less-or-equal, greater-or-equal) and equality constraints. Let

$$X_{\mathrm{opt},t}(x^{t-1}, \eta^t, \xi^t) \subset \mathbb{R}^{n_t} \quad \text{and} \quad D_{\mathrm{opt},t}^*(x^{t-1}, \eta^t, \xi^t) \subset \mathbb{R}^{r_t}$$

be the primal and dual solution sets associated with the parametric optimization problem (2.2), respectively. The dual solutions correspond to the Lagrange multipliers associated with the explicit constraints in (2.2). We will interpret $X_{\mathrm{opt},t}$ and $D_{\mathrm{opt},t}^*$ as multifunctions on the underlying parameter space. Under the given regularity conditions it can be proved that both $X_{\mathrm{opt},t}$ and $D_{\mathrm{opt},t}^*$ are non-empty-valued, bounded, and Berge upper semicontinuous on a neighborhood of the natural domain Z^t; see [43, Chap. 5]. Thus, there are nonnegative finite bounding vectors $X_t^+, X_t^- \in \mathbb{R}^{n_t}$ and $D_t^{*+}, D_t^{*-} \in \mathbb{R}^{r_t}$ such that

$$-X_t^- \le x_t \le X_t^+ \quad \text{uniformly for all} \quad x_t \in X_{\mathrm{opt},t}(Z^t) \tag{5.4a}$$

and

$$-D_t^{*-} \le d_t^* \le D_t^{*+} \quad \text{uniformly for all} \quad d_t^* \in D_{\mathrm{opt},t}^*(Z^t). \tag{5.4b}$$

Due to the assumption about the recession cone in (C3), the primal bounding vectors are usually easy to find. As for the dual bounding vectors, recall that

Lagrange multipliers are nonnegative for less-or-equal constraints and nonpositive for greater-or-equal constraints; no a priori statement about the sign of Lagrange multipliers is available for equality constraints. Thus, given that $\sim_t$ only contains less-or-equal relations, we may choose $D_t^{*-} = 0$. Conversely, if $\sim_t$ is exclusively made up of greater-or-equal relations, we will set $D_t^{*+} = 0$. These basic rules persist on a componentwise level if $\sim_t$ represents a heterogeneous mixture of inequality relations. Apart from that, there are no universal a priori guidelines for how to determine the dual bounding vectors. To find them, however, a good understanding of the decision problem at hand and a physical interpretation of the involved Lagrange multipliers are usually sufficient.

By means of the d.c. components of the objective and rhs vectors on one hand and the bounding vectors for the primal and dual solutions on the other hand, it is possible to define appropriate *correction terms* α_t^o and α_t^r, respectively.

$$
\begin{aligned}
\alpha_t^\text{o}(\eta^t) &= +\langle \kappa_t^{*-}(\eta^t), X_t^+ \rangle + \langle \kappa_t^{*+}(\eta^t), X_t^- \rangle & \text{(5.5a)} \\
\alpha_t^\text{r}(\xi^t) &= -\langle D_t^{*+}, \kappa_t^+(\xi^t) \rangle - \langle D_t^{*-}, \kappa_t^-(\xi^t) \rangle & \text{(5.5b)}
\end{aligned}
$$

These definitions reflect the intrinsic primal-dual symmetry of linear (stochastic) programs. Obviously, the correction term α_t^o associated with the nonconvexities in the objective function has the same general structure as the correction term α_t^r corresponding to the nonconvexities in the constraints. Concretely speaking, (5.5a) pairs the bounding vectors of the primal solutions with the d.c. components of the objective function coefficients, whereas (5.5b) pairs the bounding vectors of the dual solutions with the d.c. components of the rhs vector. Next, for all $t = 1, \ldots, T$ we define a combined *correction term* as

$$
\alpha_t(\eta^t, \xi^t) = \alpha_t^\text{o}(\eta^t) + \alpha_t^\text{r}(\xi^t). \tag{5.6}
$$

By construction, α_t is a continuous saddle function on a neighborhood of $\Theta^t \times \Xi^t$ being convex in η^t and concave in ξ^t. Let us now introduce three sequences of additional functions. Set $A_{T+1} \equiv A_{J,T+1}^l \equiv A_{J,T+1}^u \equiv 0$ and define for $t = T, \ldots, 1$

$$
\begin{aligned}
A_t(\eta^t, \xi^t) &= \alpha_t(\eta^t, \xi^t) + \int A_{t+1}(\eta^{t+1}, \xi^{t+1}) \ dP_{t+1}(\eta_{t+1}, \xi_{t+1}|\eta^t, \xi^t), \\
A_{J,t}^l(\eta^t, \xi^t) &= \alpha_t(\eta^t, \xi^t) + \int A_{J,t+1}^l(\eta^{t+1}, \xi^{t+1}) \ dP_{J,t+1}^l(\eta_{t+1}, \xi_{t+1}|\eta^t, \xi^t), \quad \text{(5.7)} \\
A_{J,t}^u(\eta^t, \xi^t) &= \alpha_t(\eta^t, \xi^t) + \int A_{J,t+1}^u(\eta^{t+1}, \xi^{t+1}) \ dP_{J,t+1}^u(\eta_{t+1}, \xi_{t+1}|\eta^t, \xi^t).
\end{aligned}
$$

The functions (5.7) will be referred to as *conditional correction terms* below. It is important to notice that the conditional correction terms are computationally accessible. Evaluation of $A_{J,t}^l$ and $A_{J,t}^u$ requires calculation of a finite sum, while A_t can usually be evaluated by means of numerical integration techniques. Since

14

α_t is continuous for every t, A_t, $A^l_{J,t}$, and $A^u_{J,t}$ are continuous[7] and bounded on $\Theta^t \times \Xi^t$. Moreover, by the saddle structure of the correction terms and the properties of the barycentric measures, we find $A^l_{J,t} \leq A_t \leq A^u_{J,t}$ on $\Theta^t \times \Xi^t$. Using the above definitions, we are now ready to state our main result.

Theorem 5.1 (Bounds on the Recourse Functions). *Consider a linear stochastic program satisfying the regularity conditions (C1)–(C5), and define the conditional correction terms as in (5.7). Then, we find*

$$\Phi^l_{J,t} + A^l_{J,t} - A_t \leq \Phi_t \leq \Phi^u_{J,t} + A^u_{J,t} - A_t \quad on \ Z^t, \ t = 1, \ldots, T, \ J \in \mathbb{N}.$$

As the refinement parameter J tends to infinity, the conditional correction terms $A^l_{J,t}$ and $A^u_{J,t}$ converge to A_t uniformly on $\Theta^t \times \Xi^t$, while the auxiliary recourse functions $\Phi^l_{J,t}$ and $\Phi^u_{J,t}$ converge to Φ_t uniformly on Z^t for $t = 1, \ldots, T$.

Proof. The claim follows from Theorems 5.16 and 5.17 as well as the discussion at the end of Sect. 5.4 in [43]. □

This theorem tells us how to approximate the unknown original recourse functions by known quantities (i.e., quantities which can at least principally be evaluated, given sufficient computer power), and provides a nonprobabilistic error estimate as the difference of the two bounds. In particular, the theorem points out a possibility to construct numerically calculable and arbitrarily tight bounds on the optimal objective value. A decision-maker, however, is not only interested in an accurate estimate of maximal expected reward, but also in the corresponding optimal policy. Thus, we will demonstrate below that bounds on the recourse functions also entail *bounding sets* for the optimal decisions.

Consider again the stochastic program (2.1) subject to the regularity conditions (C1)–(C5). For notational convenience we introduce the extended-real-valued functional

$$\rho_t(x^t, \eta^t, \xi^t) = \begin{cases} \langle c^*_t(\eta^t), x_t \rangle & \text{for } W_t x_t + T_t x_{t-1} \sim_t h_t(\xi^t), \\ -\infty & \text{else,} \end{cases}$$

which penalizes infeasible decisions with an infinite loss and can thus be viewed as the effective profit earned at stage t. Moreover, define

$$F_t(x^t, \eta^t, \xi^t) = \rho_t(x^t, \eta^t, \xi^t) + (E_t \Phi_{t+1})(x^t, \eta^t, \xi^t),$$

$$F^l_{J,t}(x^t, \eta^t, \xi^t) = \rho_t(x^t, \eta^t, \xi^t) + (E^l_{J,t} \Phi_{t+1})(x^t, \eta^t, \xi^t) + A^l_{J,t}(\eta^t, \xi^t) - A_t(\eta^t, \xi^t),$$

$$F^u_{J,t}(x^t, \eta^t, \xi^t) = \rho_t(x^t, \eta^t, \xi^t) + (E^u_{J,t} \Phi_{t+1})(x^t, \eta^t, \xi^t) + A^u_{J,t}(\eta^t, \xi^t) - A_t(\eta^t, \xi^t).$$

Given sufficient CPU speed and storage capacity, the functionals $F^l_{J,t}$ and $F^u_{J,t}$ can be numerically evaluated, whereas, in general, F_t remains computationally

---

[7]Continuity of A_t follows inductively form the dominated convergence theorem.

15

untractable. In the sequel, we will interpret these extended-real-valued mappings as functions of x_t, while the arguments x^{t-1}, η^t, and ξ^t are interpreted as parameters. Theorem 5.1 implies that $F_{J,t}^l \leq F_t \leq F_{J,t}^u$ for all (feasible and infeasible) decisions x_t as well as for all parameters in Z^t. As a consequence, the optimal stage t decisions are necessarily contained in the polyhedron[8]

$$C_{J,t}(x^{t-1}, \eta^t, \xi^t) = \left\{ x_t \;\middle|\; F_{J,t}^u(x^t, \eta^t, \xi^t) \geq \max_{x_t'} F_{J,t}^l(x_t', x^{t-1}, \eta^t, \xi^t) \right\}. \qquad (5.8)$$

This set depends parametrically on the outcome and decision history, thus defining a multifunction $C_{J,t}$. Notice that the evaluation of $C_{J,t}$ at any fixed point, though possibly time consuming, is numerically feasible as it is only based on knowledge of the functionals $F_{J,t}^l$ and $F_{J,t}^u$. The most important properties of the multifunction $C_{J,t}$, all of which can be deduced from Theorem 5.1, are summarized in the following statement.

Theorem 5.2 (Bounding Sets for the Optimal Decisions). *Consider a linear multistage stochastic program subject to the regularity conditions (C1)–(C5). Then, the multifunction $C_{J,t}$ is compact-convex-valued, and we find*

$$X_{\mathrm{opt},t} \subset C_{J,t} \quad on \; Z^t, \; t = 1, \ldots, T, \; J \in \mathbb{N}.$$

As the refinement parameter J tends to infinity, $C_{J,t}$ converges to $X_{\mathrm{opt},t}$ pointwise on Z^t in the sense of set convergence.[9]

Proof. The claim is an immediate consequence of [43, Theorem 5.22]. $\qquad \square$

Let $X_{\mathrm{opt},J,t}^l(x^{t-1}, \eta^t, \xi^t)$ and $X_{\mathrm{opt},J,t}^u(x^{t-1}, \eta^t, \xi^t)$ be the solution sets of the auxiliary parametric optimization problems (4.4a) and (4.4b), respectively. As usual, we may interpret $X_{\mathrm{opt},J,t}^l$ and $X_{\mathrm{opt},J,t}^u$ as multifunctions valued in $\mathbb{R}^{n_t}$. Using this new terminology, we may state the following corollary to Theorem 5.2.

Corollary 5.3. *Let $x_{J,t}^\star$ be a selector of $X_{\mathrm{opt},J,t}^l \cup X_{\mathrm{opt},J,t}^u$ on Z^t, and assume that $x_{J,t}^\star$ converges pointwise to $x_t^\star$ as the refinement parameter J tends to infinity. Then, $x_t^\star$ is a selector of $X_{\mathrm{opt},t}$ on Z^t for $t = 1, \ldots, T$.*

Proof. The assertion follows immediately from the inclusion

$$X_{\mathrm{opt},J,t}^l \cup X_{\mathrm{opt},J,t}^u \subset C_{J,t}, \quad t = 1, \ldots, T, \; J \in \mathbb{N},$$

and the fact that $\{C_{J,t}\}_{J \in \mathbb{N}}$ converges pointwise to $X_{\mathrm{opt},t}$ on Z^t. $\qquad \square$

---

[8]Notice that F_t^l and F_t^u are concave polyhedral functions of x_t for all fixed parameter values since we are dealing with *linear* stochastic programs and since the barycentric measures have *finite* supports.

[9]For a survey of the theory of set convergence see [49, Chap. 4].

Corollary 5.3 asserts that if J tends to infinity, all accumulation points of solutions to the auxiliary stochastic programs (4.4) solve the true recourse problem (2.2). In [25, Sect. 18] this qualitative convergence result is derived for the two-stage case by using the concept of epi-convergence due to Attouch and Wets [2]. The approach presented here, however, is more quantitative in nature as the bounding sets (5.8) additionally provide non-probabilistic error estimates for the true solutions.

6   Application in Financial Risk Management

As an illustration of the functionality and performance of bounding methods, we now present a practical example from financial risk management. It is a simplified version of a model that was developed for a major Swiss bank (for details, see [23, 33, 34]) to solve the following problem: A significant portion of a typical bank's balance sheet consists of liability positions with no contractual maturity like savings deposits. Their characteristic feature is that bank clients may freely add or withdraw investments anytime at no penalty. On the other hand, the bank is allowed to adjust the rate paid on these investments at all times as a matter of policy. As a consequence, the total volume of such a position may fluctuate heavily as clients react to changes in the offered deposit rate or the relative attractiveness of alternative investment opportunities. The uncertainty of the future volume complicates the task of the financial managers who have to reinvest the money on the market or internally for the funding of credits: The generation of sufficient income requires an allocation of larger amounts in long-term instruments which, on the other hand, increases the risk that a substantial portion of the savings volume is withdrawn, and the bank runs into liquidity problems (if a drop in volume cannot be compensated by maturing investments). Given the large volumes of such positions, it is obvious that the composition of the reinvested portfolio has a high impact on the bank's risk profile and, thus, a careful analysis of the problem is required.

6.1   Formulation as Multistage Stochastic Program

Assume that investments can be made in fixed-income securities whose maturities are given by the set $\mathcal{D} = \{1, \ldots, D\}$ (D is the longest available maturity). Let $r(\eta^t; d)$ denote the interest rate at time t for a given history of risk factors η^t. An investment of \$1 in maturity $d \in \mathcal{D}$ at time $t = 1, \ldots, T$ generates a (discounted) income of

$$\varphi_t^d(\eta^t) = \sum_{i=1}^{\Delta} a_t^d(t+i)\big(r(\eta^t; d) - s^d\big) \tag{6.1}$$

over the lifetime of the instrument. Since the objective is the maximization of the (expected) discounted interest income during the planning horizon, payments arising after time $T + 1$ are truncated by setting $\Delta := \min\{T - t + 1, d\}$. The

(deterministic) function $a_t^d(\tau)$ calculates the discount factor at time τ from the *initial* yield curve or is zero if an instrument with maturity d in which the model invests in t does not pay a coupon in τ. Transaction costs in terms of a bid spread are given by s^d. If a decrease in the total volume cannot be compensated by the sum of maturing tranches, the resulting gap must be refinanced in the shortest maturity. Then, a penalty p is charged in addition to the one-period market rate that represents the cost for liquidity risk, i.e., the corresponding coefficient in the objective function is $\vartheta_t(\eta^t) = a_t^1(t+1)(r(\eta^t;1) + p)$. The volume $v_t(\xi^t)$ in the account at time t depends on the past realizations of some process $\boldsymbol{\xi}$. The specific form of the processes $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ will be introduced in the sequel.

At each point in time $t = 1, \ldots, T$, decisions on the amounts $x_{\mathrm{inv},t}^d$, $d \in \mathcal{D}$, have to be made according to which the sum of maturing tranches, corrected by a change in the total volume of the position, is reinvested in the available instruments. Any investment $x_{\mathrm{inv},t}^d$ increases the total volume $x_{\mathrm{pos},t}^d$ of the positions with maturity d. An amount refinanced in the shortest maturity as compensation for large drops in volume is represented by $x_{\mathrm{ref},t}$. Thus, the complete decision vector at stage t is defined as

$$x_t = (x_{\mathrm{inv},t}^1, \ldots, x_{\mathrm{inv},t}^D, x_{\mathrm{pos},t}^1, \ldots, x_{\mathrm{pos},t}^D, x_{\mathrm{ref},t}) \in \mathbb{R}^{2D+1}.$$

As pointed out in Sect. 2, x_t is interpreted as the time t realization of a stochastic process $\boldsymbol{x}$ characterizing the entire scenario-dependent decision strategy, and its subprocesses $\boldsymbol{x}_{\mathrm{inv}}^1, \ldots, \boldsymbol{x}_{\mathrm{inv}}^D, \boldsymbol{x}_{\mathrm{pos}}^1, \ldots, \boldsymbol{x}_{\mathrm{pos}}^D$ and $\boldsymbol{x}_{\mathrm{ref}}$ are defined in the obvious way. With the objective to maximize the expected income on the reinvested portfolio minus the costs for liquidity risk, our decision problem can be formally represented as the multistage stochastic program

$$\underset{\boldsymbol{x} \in \mathcal{N}}{\mathrm{maximize}} \, E\left[\sum_{t=1}^{T} \left(\sum_{d \in \mathcal{D}} \varphi_t^d(\boldsymbol{\eta}^t) \, \boldsymbol{x}_{\mathrm{inv},t}^d - \vartheta_t(\boldsymbol{\eta}^t) \, \boldsymbol{x}_{\mathrm{ref},t} \right) \right]$$

subject to:

$$
\left.
\begin{aligned}
\boldsymbol{x}_{\mathrm{pos},t}^1 - \boldsymbol{x}_{\mathrm{pos},t-1}^2 - \boldsymbol{x}_{\mathrm{inv},t}^1 + \boldsymbol{x}_{\mathrm{ref},t} &= 0 \\
\boldsymbol{x}_{\mathrm{pos},t}^d - \boldsymbol{x}_{\mathrm{pos},t-1}^{d+1} - \boldsymbol{x}_{\mathrm{inv},t}^d &= 0 \qquad d = 2, \ldots, D \\
\sum_{d \in \mathcal{D}} \boldsymbol{x}_{\mathrm{pos},t}^d &= v_t(\boldsymbol{\xi}^t) \\
0 \le \boldsymbol{x}_{\mathrm{inv},t}^d \le \ell^d & \qquad d = 1, \ldots, D \\
0 \le \boldsymbol{x}_{\mathrm{ref},t} \le \ell^1 &
\end{aligned}
\right\}
\begin{aligned}
& (6.2) \\
\\
& P\text{-a.s.} \\
& t = 1, \ldots, T.
\end{aligned}
$$

Herein, $\boldsymbol{x}_{\mathrm{pos},0}^d$ are degenerate deterministic random variables and denote positions with maturity d held in the initial portfolio. Upper limits ℓ^d on the transaction volumes reflect liquidity restrictions in the Swiss market for certain maturities; in case of non-standard maturities that are not traded, these limits are set to zero. We also assume here that the initial values $\boldsymbol{\eta}_1$ and $\boldsymbol{\xi}_1$ of the underlying stochastic

18

processes, which affect the coefficients in the first stage, can be observed in the market and are thus deterministic.

6.2 Risk Factor Models

It now remains to specify the stochastic processes that drive the evolution of interest rates and volume. Empirical studies imply that two factors explain most of the volatility of the term structure, and that these factors can be chosen as the level $\boldsymbol{\eta}_l$ of the yield curve, e.g., in terms of the rate for an infinite maturity, and the spread $\boldsymbol{\eta}_s$ between the instantaneous short rate and the level factor (see for example [32]). For the ease of exposition, we apply a simplified discrete-time term structure model, in which the stochastic changes of the factors are given by

$$
\begin{aligned}
\boldsymbol{\eta}_{s,t} - \boldsymbol{\eta}_{s,t-1} &= a_s(\theta_s - \boldsymbol{\eta}_{s,t-1})\,\Delta t + \hat{\boldsymbol{\varepsilon}}^{\mathrm{o}}_{s,t} \\
\boldsymbol{\eta}_{l,t} - \boldsymbol{\eta}_{l,t-1} &= a_l(\theta_l - \boldsymbol{\eta}_{l,t-1})\,\Delta t + \hat{\boldsymbol{\varepsilon}}^{\mathrm{o}}_{l,t}
\end{aligned}
\qquad t = 1,\dots,T. \tag{6.3}
$$

This process specification incorporates the 'mean reversion' property, i.e., there is a drift term that forces the process $\boldsymbol{\eta}_i$, $i \in \{s,l\}$, from its current value towards the long-term mean θ_i at a speed controlled by a_i. This reflects the empirical observation that interest rates fluctuate within a certain range. The disturbances $\hat{\boldsymbol{\varepsilon}}^{\mathrm{o}}_{s,t}$ and $\hat{\boldsymbol{\varepsilon}}^{\mathrm{o}}_{l,t}$ are conditionally independent of the past given $\boldsymbol{\eta}_{l,t}$. We further assume that they are conditionally normally distributed given $\boldsymbol{\eta}_{l,t}$, i.e.,

$$
\hat{\boldsymbol{\varepsilon}}^{\mathrm{o}}_{s,t}, \hat{\boldsymbol{\varepsilon}}^{\mathrm{o}}_{l,t} \mid \boldsymbol{\eta}_{l,t} \sim \mathcal{N}(0, \Sigma^{\mathrm{o}}_t) \quad \text{where} \quad \Sigma^{\mathrm{o}}_t = \begin{pmatrix} \sigma_s^2 & \varrho\sigma_s\sigma_l\boldsymbol{\eta}^{\gamma}_{l,t} \\ \varrho\sigma_s\sigma_l\boldsymbol{\eta}^{\gamma}_{l,t} & \sigma_l^2\boldsymbol{\eta}^{2\gamma}_{l,t} \end{pmatrix} \Delta t. \tag{6.4}
$$

Parameter values $\gamma > 0$ are used to reflect a possible heteroscedasticity that may be found in historical interest rate data. For $\gamma = 0$, in contrast, heteroscedasticity is lost, and the disturbances $\hat{\boldsymbol{\varepsilon}}^{\mathrm{o}}_{s,t}$ and $\hat{\boldsymbol{\varepsilon}}^{\mathrm{o}}_{l,t}$ become serially independent. Note that all parameters in the above model are annualized.

A very simple approach to obtain the interest rates of the maturities relevant for investment is to model the yield curve at time t by some exponential function:

$$
r(\eta_s^t, \eta_l^t; d) = (\eta_{s,t} + \beta_1 d)\,e^{-\beta_2 d} + \eta_{l,t}. \tag{6.5}
$$

Note that the constants β_1, β_2 control the shape of the yield curve while the rates $r(\eta_s^t, \eta_l^t; d)$, $d \in \mathcal{D}$, themselves are linear in the factors. The parameters of (6.3) and (6.4) can easily be estimated when the factors are approximated by observed interest rates, e.g., the five-year rate for $\boldsymbol{\eta}_l$ and the difference between the one-month rate and the latter for $\boldsymbol{\eta}_s$. Then, estimates of β_1 and β_2 in the yield curve function (6.5) are derived in a second step by minimizing the differences between the rates implied by the model and those of a historical sample.

Remark: In contrast to the simplified example considered here, the more general formulation of our optimization model, which is used for 'real-world applications', contains also variables for short sales (negative investments). This

extended model uses a term structure model which precludes arbitrage opportunities in combination with very low bid-ask spreads. If arbitrage opportunities occurred in the scenarios, the optimization model would try to exploit them by refinancing at the cheapest and investing at the highest interest rate, which would lead to unrealistic investment decisions. However, we avoid a detailed discussion of this case here for the ease of exposition.

For obvious reasons, the volume of the non-maturing account cannot become negative since clients are not allowed to withdraw higher amounts than their previous investments. We therefore model the volume in t as an exponential function $v_t(\xi^t) = \exp(\xi_t)$ of a stochastic factor which itself follows a first-order autoregressive process:

$$\boldsymbol{\xi}_t = a_v + b\,\boldsymbol{\xi}_{t-1} + \hat{\boldsymbol{\varepsilon}}_t^{\mathrm{r}}, \qquad \hat{\varepsilon}_t^{\mathrm{r}} \sim N(0, \sigma_\varepsilon), \qquad t = 1, \ldots, T. \tag{6.6}$$

The random variables $\hat{\varepsilon}_t^{\mathrm{r}}$ are assumed to be serially independent but may be correlated with the noise factors $\hat{\varepsilon}_{s,t}^{\mathrm{o}}$ and $\hat{\varepsilon}_{l,t}^{\mathrm{o}}$ of the interest rate model. This can reflect a possible dependency between the volume and the yield curve, as is often observed for non-maturing accounts.

In order to facilitate comparison with our theoretical results of the previous sections, we define for each $t = 1, \ldots, T$ the AR(1) coefficient matrices

$$H_t^{\mathrm{o}} = \begin{pmatrix} 1 - a_s\,\Delta t & 0 \\ 0 & 1 - a_l\,\Delta t \end{pmatrix}, \qquad H_t^{\mathrm{r}} = b,$$

and introduce two random vectors

$$\boldsymbol{\varepsilon}_t^{\mathrm{o}} = \begin{pmatrix} a_s\,\theta_s\,\Delta t \\ a_l\,\theta_l\,\Delta t \end{pmatrix} + \begin{pmatrix} \hat{\varepsilon}_{s,t}^{\mathrm{o}} \\ \hat{\varepsilon}_{l,t}^{\mathrm{o}} \end{pmatrix}, \qquad \boldsymbol{\varepsilon}_t^{\mathrm{r}} = a_v + \hat{\varepsilon}_t^{\mathrm{r}}.$$

Using these conventions, the risk factor processes can be recast as

$$\boldsymbol{\eta}_t = H_t^{\mathrm{o}}\boldsymbol{\eta}_{t-1} + \boldsymbol{\varepsilon}_t^{\mathrm{o}} \quad \text{and} \quad \boldsymbol{\xi}_t = H_t^{\mathrm{r}}\boldsymbol{\xi}_{t-1} + \boldsymbol{\varepsilon}_t^{\mathrm{r}}. \tag{6.7}$$

6.3 Check of Regularity Conditions

Let us first assume that $\gamma = 0$ in (6.4) precluding heteroscedasticity in the interest rate model. Then, the risk factors $\boldsymbol{\eta}_t$ and $\boldsymbol{\xi}_t$ are normally distributed and have infinite support. In order to satisfy condition (C1), it is thus required that the distributions of $\boldsymbol{\varepsilon}_t^{\mathrm{o}}$ and $\boldsymbol{\varepsilon}_t^{\mathrm{r}}$ are truncated outside some regular simplices which are large enough to contain most of the probability mass. For simplicity, consider the case of a one-dimensional standard normal distribution (s.n.d.). A simplex that contains a given percentage p of the mass of the s.n.d. reduces to the interval $[-\delta, \delta]$, where $\delta = \Phi(1 - \frac{1-p}{2})$, and Φ denotes the standard normal distribution function. In the two-dimensional case, the corresponding simplex is an equilateral triangle with an inner circle of radius δ which represents a δ-confidence

region for the s.n.d. in $\mathbb{R}^2$. The vertices of this triangle may be chosen, e.g., as $u_0 = (-\sqrt{3}\delta, \delta), u_1 = (\sqrt{3}\delta, \delta)$ and $u_2 = (0, -2\delta)$. Simplicial coverages in higher dimensions are discussed in [29].

To determine the simplicial support of the stochastic factor ε_t^o influencing the objective, we transform the vertices of a simplex constructed for an uncorrelated two-dimensional s.n.d. according to $\mu_t^o + \Gamma_t u_i$, $i \in \{0, 1, 2\}$, where $\mu_t^o = (a_s \theta_s \Delta t, a_l \theta_l \Delta t)$ and Γ_t is the Cholesky transformation of the covariance matrix Σ_t^o in (6.4). Analogously, the simplicial support of the risk factor ε_t^r, that controls the coefficients on the rhs, is given by the interval $[a_v - \delta\sigma_\varepsilon, a_v + \delta\sigma_\varepsilon]$. For the calculations presented below, we assume $\delta = 2$ so that any outcome within a range of at least two standard deviations around the expectation will be taken into account.

Consequently, more extremal events than $a_v \pm \delta\sigma_\varepsilon$ are ignored. This is consistent with the fact that in reality the decision maker would accept that with a certain (sufficiently small) probability a drop in volume cannot always be compensated by maturing instruments. Otherwise, he or she had to invest the complete amount into the shortest maturity and, thus, give away potential return which is not consistent with the usual practice of investing balance sheet items without contractual maturity. In this spirit, a specific choice of δ reflects the decision maker's tolerance towards liquidity risk.

Condition (C2) is satisfied since the objective function coefficients are linear in η according to (6.1) and (6.5). The rhs functions of the constraints in the optimization problem (6.2) are constant or convex in ξ and, a fortiori, d.c. Moreover, all constraint matrices are deterministic, and the feasible sets of the dynamic version of (6.2) are uniformly compact in each stage due to the bounds on the transaction volumes. This implies that condition (C3) is satisfied. It can easily be seen from (6.7) that $(\boldsymbol{\eta}, \boldsymbol{\xi})$ follows a block-diagonal autoregressive process if $\gamma = 0$, i.e., the conditional covariance matrix of $\boldsymbol{\eta}_{t+1}$ is independent of the current level of the risk factor $\boldsymbol{\eta}_t$. Regularity condition (C4) is thus fulfilled.

Condition (C5) requires that the stochastic program has relatively complete recourse. Extreme fluctuations in the $\boldsymbol{\xi}$ process might in principle lead to a situation where the volume constraint cannot hold in combination with tight transaction limits. However, this situation does not occur for reasonable parameter values which imply that the support of $\boldsymbol{\xi}$ is relatively small (partly because of the inherent 'mean reversion property'; this might be different for nonstationary processes). Due to the possibility of borrowing money on a short-term basis, we can thus take condition (C5) for granted. Note also that a possible remedy to achieve relatively complete recourse might be the selection of a smaller parameter δ for the truncation of the original supports.

Since the right hand side of the third equation in (6.2) is nonlinear in ξ_t, we must determine correction terms of the form (5.6) for each stage in order to employ Theorem 5.1. To this end, we study the dynamic version of (6.2). Considering the stage t subproblem, we must find an upper bound $D_{\text{vol},t}^{*+}$ for the

dual variables associated with the constraint

$$\sum_{d \in \mathcal{D}} x_{\text{pos},t}^d = \exp(\xi^t) \tag{6.8}$$

uniformly over all outcome and decision histories in Z^t. This dual variable can be seen as the *shadow price* of the volume in the account. A raise in the volume v_t by \$1 will increase at least one of the portfolio positions $x_{\text{pos},t}^1, \ldots, x_{\text{pos},t}^D$ which can only be achieved by increasing also some of the investments $x_{\text{inv},t}^d$, $d \in \mathcal{D}$. Therefore, the maximum gain in t per additional unit of currency is

$$D_{\text{vol},t}^{*+} = \max\{\varphi_t^d(\eta^t) \mid d \in \mathcal{D}, \eta^t \in \Theta^t\}. \tag{6.9}$$

This quantity is finite since φ_t^d is linear in η^t and the domain Θ^t is bounded. Note that any increase in the volume at stage t will affect only the objective function coefficients at this stage because the additional money is invested immediately. We can thus determine the upper bounds in (6.9) from the maximum values of $\eta_{i,t}$ over Θ^t, $i \in \{s, l\}$, which are obtained by recursively using (6.3). Thereby, we assume that $\gamma = 0$ and $1 - a_i \Delta t \geq 0$.

$$\max \eta_{i,t} = a_i \theta_i \Delta t + (1 - a_i \Delta t) \max \eta_{i,t-1} + \max \hat{\varepsilon}_{i,t}^{\text{o}}, \quad i \in \{s, l\}, \quad t > 1$$

An obvious choice for the d.c. decomposition of the right hand side in (6.8) is $\kappa_t^+(\xi^t) = \exp(\xi_t)$ and $\kappa_t^-(\xi^t) = 0$. The correction term (5.5b) then becomes

$$\alpha_t^{\text{r}}(\xi^t) = -D_{\text{vol},t}^{*+} \exp(\xi^t), \qquad t = 1, \ldots, T.$$

Because $\alpha_t^{\text{o}}(\eta^t) = 0$ for all stages (no corrections of nonconvexities in the objective are required), this equals also the combined correction term $\alpha_t(\eta^t, \xi^t)$. Note that we do not have to estimate a lower bound $D_{\text{vol},t}^{*-}$ in (5.5b) due to $\kappa_t^- = 0$. Based on this information, we can now evaluate the conditional correction terms defined in (5.7). As mentioned earlier, $A_{J,t}^l$ and $A_{J,t}^u$ are calculated by direct summation, while A_t must be evaluated either analytically or via numerical integration techniques. Notice that the recourse functions as well as the conditional correction terms of the first stage have to be evaluated only at one point, since $\boldsymbol{x}_0$, $\boldsymbol{\eta}_1$, and $\boldsymbol{\xi}_1$ are deterministic. As a consequence, we will suppress the arguments x_0, η_1, and ξ_1 below. In the simplified example presented here, we calculate A_1 approximately with respect to the unrestricted normal distribution and ignore the truncation of the state space. To this end, ξ_t is rewritten for each $t > 1$ as a linear combination of the noise terms, i.e.,

$$\xi_t = H_{1,t}^{\text{r}} \xi_1 + \sum_{\tau=1}^t H_{\tau+1,t}^{\text{r}} \varepsilon_\tau^{\text{r}}, \quad \text{where} \quad H_{\tau,t}^{\text{r}} = \begin{cases} \prod_{\tau'=\tau}^t H_{\tau'}^{\text{r}} & \text{for } \tau \leq t, \\ 1 & \text{otherwise.} \end{cases}$$

Then, the conditional correction term A_1 becomes

$$A_1 = -\sum_{t=1}^{T} D_{\text{vol},t}^{*+} E[\exp(\xi_t)]$$

$$\approx -\sum_{t=1}^{T} D_{\text{vol},t}^{*+} \exp\left[H_{1,t}^{\text{r}} \xi_1 + \sum_{\tau=1}^{t}\left(H_{\tau+1,t}^{\text{r}} a_v + \frac{1}{2}(H_{\tau+1,t}^{\text{r}} \sigma_\varepsilon)^2\right)\right].$$

According to Theorem 5.1, the bounds on the optimal value Φ_1 of the multistage stochastic program (6.2) consist in the optimal values $\Phi_{J,1}^l$, $\Phi_{J,1}^u$ of the discretized auxiliary stochastic programs (4.4a) and (4.4b) shifted by combinations of the conditional correction terms A_1, $A_{J,1}^l$ and $A_{J,1}^u$ for the current refinement parameter $J \in \mathbb{N}_0$.

Remark: We emphasized above that $\gamma > 0$ in (6.3) allows to reflect heteroskedasticity of interest rates. However, this choice is not consistent with definition 3.2. It is argued in [34] that the saddle property of the recourse functions is sometimes given also for $\gamma = 1$, since this specification still entails a linear dependency of $\boldsymbol{\eta}_t$ on its history of observations. In order to cope with other parameter values[10] $\gamma > 0$, one has to represent the risk factors as nonlinear combinations of some serially independent disturbances. Considering these disturbances as the fundamental data process, and packing all nonlinearities into the definition of the objective function coefficients, our bounding technique remains applicable. However, this approach involves a completely different set of correction terms and will not be further pursued in the present article.

6.4 Numerical Solution

We solved a 4-stage problem with quarterly planning, investment opportunities in the maturities 3M, 6M, 1Y, 2Y, 3Y, 4Y and 5Y, and parameters estimated from real data. The barycentric measures were refined by successively partitioning the simplicial supports $\mathcal{E}_t^{\text{o}}$ and $\mathcal{E}_t^{\text{r}}$ of the disturbances $\boldsymbol{\varepsilon}_t^{\text{o}}$ and $\boldsymbol{\varepsilon}_t^{\text{r}}$, respectively. Starting with an initial trivial partition $J = 0$ that consists of the product $\mathcal{E}_t^{\text{o}} \times \mathcal{E}_t^{\text{r}}$ for all stages $t = 1, \ldots, T$ we obtained (uncorrected) objective function values of $\Phi_{0,1}^l = 1011.5718$ and $\Phi_{0,1}^u = 1221.7353$ (recall that the first index represents the refinement parameter J). Taking into account the conditional correction terms $A_{0,1}^l - A_1$ and $A_{0,1}^u - A_1$, the bounds become 1011.5712 and 1221.7353, respectively. This illustrates that the magnitude of the corrections is relatively small, a result which should not be generalized to other (more complex) problems where corrections apply also to the objective, or the bounding vectors in (5.4a) and (5.4b) turn out to be very large. On the other hand, the difference between the two bounds itself is relatively large here. Furthermore, the corresponding

---

[10]For instance, the well-known Cox, Ingersoll and Ross model [8] involves a 'square root process' with $\gamma = 0.5$.

first-stage decisions $x^l_{\text{opt},0,1}$ and $x^u_{\text{opt},0,1}$ do not coincide. The solution of the lower bounding problem recommends to invest the whole amount available at $t = 1$ in the longest maturity, while the suggestion of the upper bounding problem is to invest in 3M, and thus the decision maker does not obtain a unique solution.

As outlined in Sect. 4, the accuracy of the approximation can be improved by increasing the number of components in the existing partitions. This can be achieved by splitting a $\mathcal{E}^o_t$- or a $\mathcal{E}^r_t$-simplex (or both) at some stage $t = 1, \ldots, T$. Here we restrict the procedure to splits of the $\mathcal{E}^o_t$-simplices only since numerical experiments have shown that this leads to the largest improvements during the first refinement steps. For our numerical example, we implemented the following refinement procedure:

(1) Fix an initial partition with N simplices, and set $J = 0$.
(2) Let $\mathcal{E}^{o(i)}_{J,t}$, $i = 1, \ldots, N + J$, be simplices in the current partition that cover the support of the distribution of ε^o_t, $t = 1, \ldots, T$. Solve the corresponding lower and upper bounding problems.
(3) If the difference between the objective function values after correction (or some other measure of accuracy) is sufficiently low, then terminate.
(4) Otherwise, among $\mathcal{E}^{o(1)}_{J,t}, \ldots, \mathcal{E}^{o(J+N)}_{J,t}$ determine the simplex with the longest edge and split the simplex at the midpoint of this edge. Replace it in the existing partition by the resulting (sub-) simplices.
(5) Set $J := J + 1$ and goto step (2).

It can be helpful for the determination of the splitting edge in step (4) to weight the lengths of all edges by the probabilities of the corresponding simplices. This is motivated by the fact that the partition of a subcell with low probability mass will not improve the accuracy. Note that the index J is equivalent to the number of refinements.

In this way, for $N = 1$ and with one refinement (i.e., $J = 1$) we obtain objective function values of $\Phi^l_{1,1} = 1027.9749$ and $\Phi^u_{1,1} = 1136.5517$ that become 1027.9743 and 1136.5520, respectively, after correction. The relative difference between the upper and lower bound is only half as large as for the initial calculation with $J = 0$. More importantly, the decision vectors $x^l_{\text{opt},1,1}$ and $x^u_{\text{opt},1,1}$ now coincide, i.e., the first-stage decision of the upper bounding problem switched also to the investment in the longest maturity. Thus, the decision maker obtains a unique solution at a sufficiently high accuracy, which was our original intention.

Further refinement steps (i.e., $J = 2, 3, \ldots$) improve the accuracy only marginally, while the corresponding auxiliary optimization problems, that approximate the original multistage stochastic program, become intractably large: This can immediately be attributed to the fact that, as outlined in Sect. 4, the approximation is based on the barycenters and vertices of the simplices in the partition. The size of the auxiliary problems thus may become too large as J increases. Numerical experiments have shown that the underlying value functions exhibit a
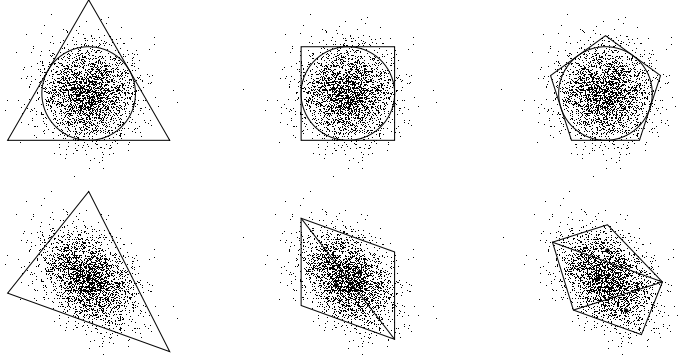
Figure 1: Coverages for the uncorrelated and correlated case with barycenters

N	J	$\Phi^l_{J,1}$	corr.	LB	$\Phi^u_{J,1}$	corr.	UB	acc.
	0	1011.5718	-5.45	1011.5712	1221.7353	2.89	1221.7356	18.82%
	1	1027.9749	-5.95	1027.9743	1136.5517	2.89	1136.5520	10.03%
1	2	1029.1618	-6.43	1029.1612	1128.2428	2.89	1128.2431	9.19%
	3	1029.2132	-6.60	1029.2125	1127.8974	2.89	1127.8976	9.15%
	4	1029.2132	-6.64	1029.2126	1127.8856	2.89	1127.8858	9.15%

N	J	$\Phi^l_{J,1}$	corr.	LB	$\Phi^u_{J,1}$	corr.	UB	acc.
	0	1034.2863	-5.41	1034.2858	1067.7401	2.17	1067.7403	3.18%
	1	1037.0787	-5.89	1037.0781	1057.9300	2.17	1057.9302	1.99%
2	2	1037.2189	-6.19	1037.2183	1057.7964	2.17	1057.7966	1.96%
	3	1037.3000	-6.31	1037.2994	1057.7764	2.17	1057.7766	1.95%
	4	1037.3063	-6.38	1037.3057	1057.7740	2.17	1057.7742	1.95%

N	J	$\Phi^l_{J,1}$	corr.	LB	$\Phi^u_{J,1}$	corr.	UB	acc.
	0	1011.5534	-4.75	1011.5529	1126.2541	2.30	1126.2544	10.73%
	1	1027.5144	-5.24	1027.5138	1087.5954	2.30	1087.5956	5.68%
3	2	1028.9230	-5.65	1028.9225	1086.2739	2.30	1086.2741	5.42%
	3	1028.9583	-5.82	1028.9577	1086.1394	2.30	1086.1396	5.41%
	4	1028.9589	-5.89	1028.9583	1086.1340	2.30	1086.1342	5.41%

Table 1: Objective function values of the lower and upper auxiliary stochastic programs before and after correction for nonlinearities. N denotes the number of simplices in the initial partition that covers the support of ε^o_t in each stage $t > 1$, while the index J is equivalent to the number of refinements. The columns corr. contain the conditional correction terms $A^l_{J,1} - A_1$ for the lower and $A^u_{J,1} - A_1$ for the upper bounding problems multiplied by 10^4. LB and UB are the corresponding objective function values $\Phi^l_{J,1} + A^l_{J,1} - A_0$ and $\Phi^u_{J,1} + A^u_{J,1} - A_0$, respectively, after correction. The accuracy acc. of the approximation is defined as $\frac{UB-LB}{0.5(LB+UB)}$.

relatively high degree of convexity with respect to η_t. Loosely speaking, the vertices of a (single) initial simplex $\mathcal{E}_t^{\mathrm{o}}$ covering the support of the truncated normal distribution of $\varepsilon_t^{\mathrm{o}}$ have geometrically 'too extreme' coordinates. Since a split of a simplex generates only one new point (the midpoint of an edge of one simplex in the given partition) but the existing vertices remain, the influence of extreme outcomes in the initial discrete approximation decreases only slowly.

To achieve tighter bounds with less refinement steps, we start from the consideration that the circle which covers the truncated support of a two-dimensional standard normal distribution can better be approximated by a polygon with a higher number of vertices. As an example, we consider coverages by tetragons and pentagons that are themselves partitioned into two or three simplices. Their vertices are transformed according to the expectations and covariance matrices of the actual distributions analogously to the procedure described in Sect. 6.3. The shapes of the resulting partitions are illustrated in Fig. 1. Then, barycentric measures are derived for each simplex individually.

Results for our exemple investment problem are shown in Table 1 for initial partitions that consist of triangles ($N = 1$), tetragons ($N = 2$), and pentagons ($N = 3$). It can be seen from the last columns labelled *acc.* that the unrefined problems in the latter two cases already provide an accuracy which may be achieved otherwise only after some refinement steps. Moreover, the first-stage decisions coincide already for $J = 0$. A suitable selection of the initial partitions is therefore of utmost importance not only to achieve a sufficient accuracy of the approximation but also to keep the overall numerical efforts moderate since each additional refinement step requires the solution of a corresponding (large-scale) auxiliary optimization problem.

7 Conclusions

This article addresses the approximation of linear multistage stochastic programs. Thereby, the original optimization problem is approximated by a lower (upper) auxiliary stochastic program which arises by substituting the true distribution of the data process by a lower (upper) discrete barycentric measure. After a suitable transformation offsetting possible nonconvexities in the random parameters, the optimal value of the lower (upper) auxiliary stochastic program provides a lower (upper) bound on the optimal value of the original stochastic program. More generally, we are able to derive arbitrarily tight bounds on the recourse functions as well as arbitrarily tight bounding sets for the optimal decisions associated with the original optimization problem.

Applicability of the presented bounding methods relies on a set of regularity conditions requiring that the feasible sets and state spaces are compact, the constraint matrices are deterministic, the objective function coefficients and rhs vectors are representable as differences of convex functions, and the random data

(a) No refinements   (b) Refinement in A   (c) Refinement in B

(d) Split of $\mathcal{E}_t^{\mathrm{o}}$ or $\mathcal{E}_t^{\mathrm{r}}$ (e) Alternative edges   (f) Alternative points
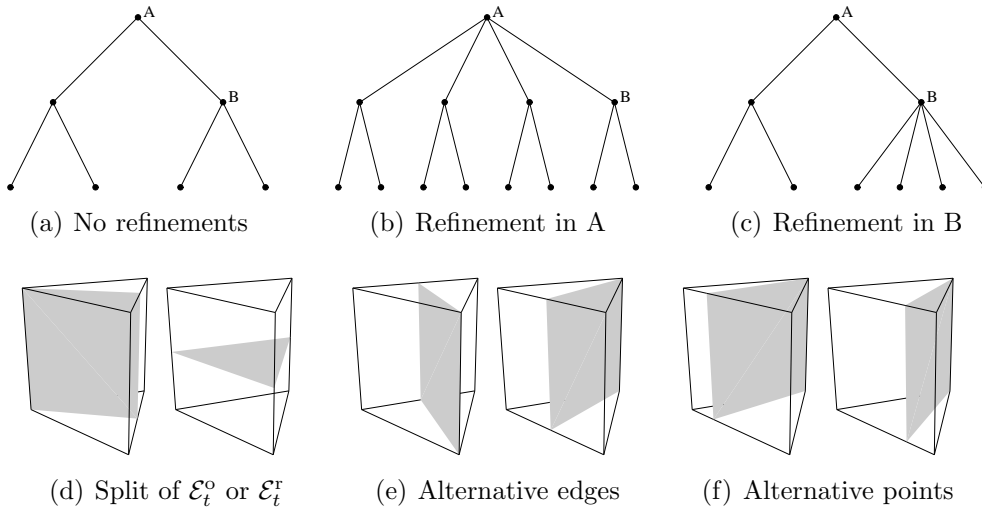
Figure 2: Possible refinements of a scenario tree and simplices

is governed by a block-diagonal autoregressive process. Moreover, relatively complete recourse is stipulated. Focussing on block-diagonal autoregressive processes is not as restrictive as it might seem, since any absolutely continuous process is representable as an (inverse) Rosenblatt transformation of a serially independent noise process [50]. Considering this noise process as the fundamental data process, and absorbing the nonlinear Rosenblatt transformation in the definition of the objective function coefficients and rhs vectors, one usually obtains a regular reformulation of the stochastic program satisfying all conditions (C1)–(C5).

The theoretical concepts are tested on a real-life decision problem calibrated to observed data. In this example, which is slightly simplified for didactic reasons, the (conditional) correction terms hardly affect the bounds on the optimal objective value. The reason for this is that the curvature of the exponential function is small on the support of the $\boldsymbol{\xi}$ process. Of course, one could think of problems with larger correction terms. It is possible, for instance, to construct models with highly nonconvex objective function coefficients or rhs vectors such that $\Phi_{J,t}^u$ occasionally drops below $\Phi_{J,t}^l$; especially if the barycentric measures are poorly refined. In this case, the conditional correction terms will substantially impact the bounds.

The results presented in Section 6 illustrate that the numerical efforts required for the solution of a successively refined multistage stochastic program increase significantly as the partition size $N + J$ grows. For the problem under consideration, the size of the corresponding extensive form with refinement parameter $J = 4$ is 125, 27 or 13 times larger than for the initial state with $J = 0$, depending on the coverage setting $N = 1, 2$ or 3, respectively. For this reason, we could solve problems with a relatively small number of stages ($T = 4$) only.

Despite the increasing problem size, the gain in accuracy soon becomes mar-

ginal as J advances. This must be attributed to the fact that we applied a rather crude refinement strategy in our example, where the simplices covering the supports of the (conditional) distributions at each stage were always split at their midpoints. A more sophisticated strategy would analyze the differences between the upper and lower bounds for each node of the scenario trees that are generated as discrete approximations of the underlying stochastic processes. Then, scenarios will only be added by a split of a simplex in the existing partition if the approximation error, which may be expressed by the difference between the bounds, is large for a certain node. On the other hand, an approximation error close to zero means that the approximation is already exact, and thus further refinements will not improve the accuracy.

For an efficient implementation of such refinement strategies, the following aspects must be considered:

(1) In which node should the scenario tree be refined (i.e., what is the threshold error above which an existing nodal partition is refined; this will have an immediate impact on the number of scenarios and the problem size, see Fig. 2 (a)–(c))?

(2) Does splitting $\mathcal{E}_t^{\mathrm{o}}$ or $\mathcal{E}_t^{\mathrm{r}}$ provide a higher accuracy, see Fig. 2 (d)?

(3) Which edge of the simplex should be split, see Fig. 2 (e), and

(4) where should this edge be split, see Fig. 2 (f)?

Adding new scenarios only where profitable will keep the increase in problem size relatively moderate, which in turn allows the solution of problems with a higher number of stages (see [18] or [31] for a detailed discussion of refinement techniques and their implementation). While each refinement involves the solution of successively increasing large-scale linear programs, a significant speedup can be achieved by an algorithmic technique described in [28] where the solution of the LP in refinement $J + 1$ is based on the solutions of the previous J steps. This allows to manage numerous refinement steps (up to several hundred) successfully with adequate time exposure.

References

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

[2] ATTOUCH, H., AND WETS, R.-B. Approximation and convergence in non-linear optimization. In *Nonlinear Programming 4* (1981), O. Mangasarian, R. Meyer, and S. Robinson, Eds., Academic Press, pp. 367–394.

[3] BEALE, E. On minimizing a convex function subject to linear inequalities. *Journal of the Royal Statistical Society 17B* (1955), 173–184.

[4] BIRGE, J., AND DULÁ, J. Bounding separable recourse functions with limited distribution information. *Ann. Oper. Res. 30* (1991), 277–298.

[5] BIRGE, J., AND LOUVEAUX, F. *Introduction to Stochastic Programming.* Springer-Verlag, New York, 1997.

[6] BIRGE, J., AND WETS, R.-B. Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse. *Math. Program. Study 27* (1986), 54–102.

[7] BIRGE, J., AND WETS, R.-B. Computing bounds for stochastic programming problems by means of a generalized moment problem. *Math. Oper. Res. 12* (1987), 149–162.

[8] COX, J. C., INGERSOLL, J. E., AND ROSS, S. A. A Theory of the Term Structure of Interest Rates. *Econometrica 53*, 2 (1985), 385–407.

[9] DANTZIG, G. Linear programming under uncertainty. *Manage. Sci. 1* (1955), 197–206.

[10] DANTZIG, G. B., AND THAPA, M. N. *Linear Programming 1: Introduction.* Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1997.

[11] DOKOV, S., AND MORTON, D. Higher-order upper bounds on the expectation of a convex function. The Stochastic Programming E-Print Series (SPEPS), January 2002.

[12] DOKOV, S. P., AND MORTON, D. P. Second-order lower bounds on the expectation of a convex function. *Math. Oper. Res. 30*, 3 (2005), 662–677.

[13] DUPAČOVÁ, J. Minimax stochastic programs with nonconvex nonseparable penalty functions. In *Progress in Operations Research* (1976), A. Prékopa, Ed., vol. 1 of *Colloquia Mathematica Societatis János Bolyai*, North-Holland Publishing, Amsterdam, pp. 303–316.

[14] DUPAČOVÁ, J. The minimax approach to stochastic programming and an illustrative application. *Stochastics 20* (1987), 73–88.

[15] DUPAČOVÁ, J. Applications of stochastic programming under incomplete information. *Comput. Appl. Math. 56* (1994), 113–125.

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

[17] Edirisinghe, N. New second-order bounds on the expectation of saddle functions with applications to stochastic linear programming. *Operations Research 44* (1996), 909–922.

[18] Edirisinghe, N., and You, G.-M. Second-order senario approximation and refinement in optimization under uncertainty. *Ann. Oper. Res. 64* (1996), 143–178.

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

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

[21] Edmundson, H. Bounds on the expectation of a convex function of a random variable. Tech. rep., The Rand Corporation Paper 982, Santa Monica, California, 1956.

[22] Floudas, C. *Deterministic Global Optimization.* Kluwer Academic Publishers, Dordrecht, 2000.

[23] Forrest, B., Frauendorfer, K., and Schürle, M. A stochastic optimization model for the investment of savings account deposits. In *Operations Research Proceedings 1997* (1998), P. Kischka, H.-W. Lorenz, U. Derigs, W. Domschke, P. Kleinschmidt, and R. Mhring, Eds., Springer, pp. 382–387.

[24] Frauendorfer, K. Solving SLP recourse problems with arbitrary multivariate distributions — the dependent case. *Math. Oper. Res. 13* (1988), 377–394.

[25] Frauendorfer, K. *Stochastic two-stage programming*, vol. 392 of *Lect. Notes Econ. Math. Syst.* Springer-Verlag, Berlin, 1992.

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

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

[28] Frauendorfer, K., and Haarbrücker, G. Solving sequences of refined multistage stochastic linear programs. *Ann. Oper. Res. 124* (2003), 133–163.

[29] FRAUENDORFER, K., AND HÄRTEL, F. On the goodness of discretizing diffusion processes for stochastic programming. Working paper, Institute of Operations Research, University of St. Gallen, St. Gallen, 1995.

[30] FRAUENDORFER, K., AND KALL, P. A solution method for SLP recourse problems with arbitrary multivariate distributions — the independent case. *Problems in Control and Information Theory 17* (1988), 177–205.

[31] FRAUENDORFER, K., AND MAROHN, C. Refinement issues in stochastic multistage linear programming. In *Stochastic Programming Methods and Technical Applications (Proceedings of the 3rd GAMM/IFIP Workshop 1996)*, vol. 458 of *Lect. Notes Econ. Math. Syst.* Springer-Verlag, 1998, pp. 305–328.

[32] FRAUENDORFER, K., AND SCHÜRLE, M. Term structure models in multistage stochastic programming: Estimation and approximation. *Ann. Oper. Res. 100* (2001), 189–209.

[33] FRAUENDORFER, K., AND SCHÜRLE, M. Management of non-maturing deposits by multistage stochastic programming. *Eur. J. Oper. Res. 151* (2003), 602–616.

[34] FRAUENDORFER, K., AND SCHÜRLE, M. Refinancing mortgages in Switzerland. In *Applications of Stochastic Programming* (Philadelphia, 2005), S. Wallace and W. Ziemba, Eds., MPS-SIAM Series in Optimization, pp. 445–469.

[35] GASSMANN, H., AND ZIEMBA, W. A tight upper bound for the expectation of a convex function of a multivariate random variable. *Math. Program. Study 27* (1986), 39–53.

[36] HARTMAN, P. On functions representable as a difference of convex functions. *Pac. J. Math. 9* (1959), 707–713.

[37] HIRIART-URRUTY, J.-B. Generalized differentiability, duality and optimization for problems dealing with differences of convex functions. In *Convexity and Duality in Optimization (Groningen, 1984)*, vol. 256 of *Lect. Notes Econ. Math. Syst.* Springer-Verlag, 1985, pp. 37–70.

[38] HORST, R., AND TUY, H. *Global Optimization: Deterministic Approaches*, 3rd ed. Springer Verlag, Berlin, 1996.

[39] HUANG, C., ZIEMBA, W., AND BEN-TAL, A. Bounds on the expectation of a convex function of a random variable: With applications to stochastic programming. *Operations Research 25* (1977), 315–325.

[40] Jensen, J. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica 30* (1906), 157–193.

[41] Kall, P. Stochastic programming with recourse: upper bounds and moment problems — a review. In *Advances in Mathematical Optimization*, J. Guddat, B. Bank, H. Hollatz, P. Kall, D. Klatte, B. Kummer, K. Lommatzch, K. Tammer, M. Vlach, and K. Zimmermann, Eds. Akademie-Verlag, 1988, pp. 86–103.

[42] Kall, P. An upper bound for SLP using first and total second moments. *Ann. Oper. Res. 30* (1991), 267–276.

[43] Kuhn, D. *Generalized Bounds for Convex Multistage Stochastic Programs*, vol. 548 of *Lect. Notes Econ. Math. Syst.* Springer-Verlag, Berlin, 2004.

[44] Madansky, A. Bounds on the expectation of a convex function of a multivariate random variable. *Annals of Mathematical Statistics 30* (1959), 743–746.

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

[46] Pardalos, P., Edwin Romeijn, H., and Tuy, H. Recent development and trends in global optimization. *Comput. Appl. Math. 124* (2000), 209–228.

[47] Rockafellar, R., and Wets, R.-B. Nonanticipativity and L^1-martingales in stochastic optimization problems. In *Stoch. Syst.: Model., Identif., Optim. II; Math. Program. Study 6* (1976), pp. 170–187.

[48] Rockafellar, R., and Wets, R.-B. Stochastic convex programming: Relatively complete recourse and induced feasibility. *SIAM J. Control Optimization 14* (1976), 574–589.

[49] Rockafellar, R., and Wets, R.-B. *Variational Analysis*, vol. 317 of *A Series of Comprehensive Studies in Mathematics.* Springer-Verlag, New York, 1998.

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

[51] Wets, R.-B. Induced constraints for stochastic optimization problems. In *Techniques of Optimization* (1972), A. Balakrishnan, Ed., Academic Press, pp. 433–443.