

Math. Program., Ser. A (2008) 113:61–94
DOI 10.1007/s10107-006-0048-6

FULL LENGTH PAPER

Aggregation and discretization in multistage stochastic programming

Daniel Kuhn

Received: 5 January 2006 / Accepted: 12 September 2006 / Published online: 24 October 2006
© Springer-Verlag 2006

Abstract Multistage stochastic programs have applications in many areas and support policy makers in finding rational decisions that hedge against unforeseen negative events. In order to ensure computational tractability, continuous-state stochastic programs are usually discretized; and frequently, the curse of dimensionality dictates that decision stages must be aggregated. In this article we construct two discrete, stage-aggregated stochastic programs which provide upper and lower bounds on the optimal value of the original problem. The approximate problems involve finitely many decisions and constraints, thus principally allowing for numerical solution.

Keywords Stochastic programming · Approximation · Bounds · Aggregation · Discretization

Mathematics Subject Classification (2000) 90C15 · 90C25 · 49M29

1 Introduction

Many technical or economic decision problems depend on uncertain parameters, whose values are known only up to a probability distribution. Typically, these values are revealed sequentially in time, and (recourse) decisions are taken at each instant when new data is observed. As a future decision may depend on the realization of some random parameter, which is unknown from today's perspective, it is itself a random object.

D. Kuhn
Institute for Operations Research and Computational Finance (ior/cf-HSG),
University of St. Gallen, Bodanstrasse 6, 9000 St. Gallen, Switzerland
e-mail: daniel.kuhn@unisg.ch

By convention, a sequence of consecutive decisions is referred to as a strategy.¹ In financial applications, one usually attempts to find a strategy which satisfies certain physical or regulatory constraints and minimizes expected cost (or maximizes expected profit). Sometimes, the objective is adjusted to account for risk, in which case one would try to minimize some risk measure or maximize expected ‘utility’. Without loss of generality, we will focus on minimization problems in this article. Moreover, we will assume that the objective criterion is to minimize expected cost.

Any decision problem with the above-mentioned properties is conveniently formulated as a multistage stochastic program (MSP) (4,28,36). Unless the underlying random parameters are discretely distributed, stochastic programs represent abstract optimization problems over infinite-dimensional function spaces, which are extremely difficult to solve. Analytical solutions are not available except for very simple models of minor practical relevance. Numerical solutions, on the other hand, require suitable approximations.

Stochastic programming literature reports on a wide variety of approximation schemes, most of which are based on discretization of the underlying probability space. However, even discrete stochastic programs tend to be computationally demanding. An analysis of the computational complexity of Monte Carlo discretized MSPs is provided in (40,41). In general, problem size grows *polynomially* with the number of discretization points per random parameter and *exponentially* with the number of random parameters per stage. Moreover, it grows *exponentially* with the number of stages. Even if there is only one uncertain parameter per stage, and even if each of these parameters is approximated by a Bernoulli random variable, a discrete stochastic program may not have more than 15 stages in order to allow for (exact) numerical solution. Unfortunately, many real-life decision problems involve hourly, daily, or weekly decisions over a planning horizon of several years.² In these cases, decision stages must be aggregated in an appropriate way. Operations research practitioners usually employ simple heuristics, although the corresponding approximation errors are (sometimes even qualitatively) unknown. Theorists, on the other hand, have paid little attention to stage-aggregation so far. Instead, a substantial portion of research focusses on discretization and scenario generation, thereby simplifying the *spatial* structure (probability space) of a given MSP while preserving its *temporal* structure (time periods). The present article is aimed at alleviating this deficiency. It elaborates an approximation scheme which integrates stage-aggregation and discretization to ensure computational tractability, while providing deterministic error bounds.

Let us briefly summarize some approximation schemes which have received considerable attention in stochastic programming literature. We distinguish scenario generation techniques and methods based on aggregation.

¹ In the remainder of this article, strategies will also be referred to as *decision processes*, *policies*, or *decision rules*.

² Prototypical examples include the management of an investment portfolio or the operation of a hydrothermal power system.

A survey and evaluation of popular scenario generation techniques is provided in (29). Sampling-based methods (4, Sect. 10) exhibit useful convergence properties and provide a *probabilistic* estimate of the approximation error. However, the number of samples required to achieve a satisfactory level of precision is usually high. Moment matching methods (see e.g. (25)) are reported to perform well in practice, although they can fail to replicate the original distribution as the number of scenarios goes to infinity. Of particular interest for our purposes are the so-called bounding methods (4, Sect. 9), which provide *deterministic* bounds on the true objective value. Typically, the approximation error (i.e., the difference of upper and lower bounds) can be made small by using partitioning techniques. Some bounding methods are applicable only if the recourse functions (cost-to-go functions) of the underlying stochastic program are convex (5, 17, 21, 27, 32, 43). Other variants, such as Frauendorfer's *barycentric approximation scheme*, apply if the recourse functions represent convex-concave saddle functions (12, 13, 18–20). A generalization to specific problems with nonconvex recourse functions is discussed in (31). Closely related to the approach presented here is the sequential scenario generation algorithm by Casey and Sen (6), which provides a converging sequence of lower bounds on the objective value. This algorithm approximates the primal (dual) decisions by piecewise affine (constant) functions of the random parameters and relies on convexity of the recourse functions. The concept of probability metrics is at the heart of several modern scenario generation methods of increasing popularity. *Optimal discretization* (35), for instance, synthesizes a tree-structured approximation to a given stochastic process which minimizes the Wasserstein distance (transportation metric). *Scenario reduction* (11, 23, 37), on the other hand, starts from a discrete process involving 'all' (or a vast number of) possible scenarios. Then, the algorithm determines a scenario subset of prescribed cardinality and a new process based on this set that is closest to the original process with respect to some probability metric. Last but not least, we should mention the versatile class of internal sampling methods. Instead of using an a priori discretization, these methods for solving stochastic programs generate, delete, or modify scenarios within the solution procedure (8, 10, 14, 24, 26).

Aggregation methods are capable of approximating a stochastic program with many (possibly an infinite number of) stages by simpler problems with only few periods. However, as opposed to scenario generation, aggregation has rarely been addressed in stochastic programming literature. First results are due to Birge (2, 3), who constructs bounds for linear MSPs with random right hand sides in terms of the solutions of aggregated deterministic two-stage problems. Wright (44) suggests a generalized aggregation scheme in an elegant measure-theoretic setting that applies to arbitrary linear MSPs. He obtains upper and lower bounds on the optimal objective value by aggregating decisions or constraints, respectively. However, only fully aggregated problems (with decisions *and* constraints subject to aggregation) are valid candidates for direct numerical solution; such problems are shown to provide bounds if the randomness appears exclusively either on the right hand side or in the objective. We will use some of the methodology of (44) in later sections. Other aggregation schemes

have been developed to deal with infinite horizon problems (15, 16, 30). These methods critically rely on the assumption that prospective costs are discounted and thus – beyond some future date – have a negligible effect on the objective.

The main contribution of the present article is to elaborate an integrated stage-aggregation and space-discretization scheme that applies to convex MSPs with randomness in the objective *and* the constraints. We will construct two discrete stochastic programs with a reduced number of stages that provide upper and lower bounds on the optimal value of the original problem. Unlike the partially aggregated MSPs in (44), these approximate problems involve a finite number of decisions and constraints, thus allowing for direct numerical solution. Note that the discretization part is very closely related to the barycentric approximation scheme (19, 20). However, our derivation offers new insights as it invokes no dynamic programming recursions and highlights the importance of specific conditional independence relationships between the involved random parameters. By using widely parallel arguments in developing space-discretization and stage-aggregation methods, we work towards unification of different approximation schemes.

Adopting a Lagrangian perspective, Sect. 2 formally introduces the class of stochastic programs under consideration. Given two discrete stochastic processes which relate appropriately to the original problem data, a simple bounding approximation scheme is proposed in Sect. 3. A posteriori, Sect. 4 argues that the postulated discrete processes can indeed be constructed under mild regularity conditions. Imposing a Markovian framework, Sect. 5 develops bounds based on stage-aggregation, and Sect. 6 evaluates their compatibility with the discretization scheme of Sect. 3. Finally, Sect. 7 concludes, while Appendix A reviews some basic properties of the conditional independence relation.

2 Problem formulation

Consider a constrained cost minimization problem under uncertainty, and assume that decisions may be selected at different time points $t = 1, \dots, T$. First, we set up a probabilistic model for the underlying uncertainty. All random objects are defined on a complete 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 1 (Stochastic Process) *We say that ξ is a stochastic process with state space Z if $\xi = (\xi_1, \dots, \xi_T)$ and $Z = \times_{t=1}^T Z_t$ such that each random vector ξ_t maps (Ω, Σ) to the Borel space $(Z_t, \mathcal{B}(Z_t))$ and each Z_t is a convex closed subset of some finite-dimensional Euclidean space. Moreover, we define combined random vectors $\xi^t := (\xi_1, \dots, \xi_t)$ valued in $Z^t := \times_{\tau=1}^t Z_\tau$ for all $t = 1, \dots, T$.³*

³ Sometimes, notation is simplified by further introducing a dummy random variable ξ^0 taking the constant value 0.

All stochastic processes introduced below are assumed to be of this kind. As a notational convention, random objects will always be represented in boldface, while their realizations will be denoted by the same symbols in normal face.

Let $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ be two stochastic processes with state spaces Θ and Ξ , respectively. Assume that $\boldsymbol{\eta}$ impacts the objective function of the decision problem, whereas $\boldsymbol{\xi}$ influences the constraints. For the sake of transparent notation, we introduce the combined data process $\boldsymbol{\zeta} := (\boldsymbol{\eta}, \boldsymbol{\xi})$ with state space $Z := \Theta \times \Xi$. The information \mathcal{F}^t available at time t by keeping track of the data process is given by the induced σ -algebra corresponding to the random vectors observed by that time, i.e., $\mathcal{F}^t := \sigma(\boldsymbol{\zeta}^t)$. Frequently, it is assumed that $\boldsymbol{\zeta}_1$ is a degenerate random vector such that \mathcal{F}^1 reduces to the trivial σ -algebra $\{\Omega, \emptyset\}$. Moreover, we use the convention $\mathcal{F} := \mathcal{F}^T$, and we define $\mathbb{F} := \{\mathcal{F}^t\}_{t=1}^T$ as the filtration induced by the data process.

Let \boldsymbol{x} and \boldsymbol{y} denote two additional stochastic processes with state spaces X and Y , respectively. In the remainder, \boldsymbol{x} will be called the *primal* decision process associated with the optimization problem at hand. Similarly, we will refer to \boldsymbol{y} as the *dual* decision process. Unlike the data processes, which are given exogenously, the decision processes are a priori unspecified and will be determined endogenously in the optimization procedure. Therefore, we must agree on suitable function spaces from which the primal and dual decision processes may be chosen. Let $\mathbb{G} = \{\mathcal{G}^t\}_{t=1}^T$ be any filtration on the probability space (Ω, Σ, P) . Then, for each $t = 1, \dots, T$ we define

$$\begin{aligned} X_t(\mathbb{G}) &:= \mathcal{L}^\infty(\Omega, \mathcal{G}^t, P; X_t), & X^t(\mathbb{G}) &:= \times_{\tau=1}^t X_\tau(\mathbb{G}), \\ Y_t(\mathbb{G}) &:= \mathcal{L}^1(\Omega, \mathcal{G}^t, P; Y_t), & Y^t(\mathbb{G}) &:= \times_{\tau=1}^t Y_\tau(\mathbb{G}). \end{aligned}$$

By definition, $X(\mathbb{G}) := X^T(\mathbb{G})$ contains all essentially bounded \mathbb{G} -adapted primal strategies valued in X , whereas $Y(\mathbb{G}) := Y^T(\mathbb{G})$ contains all integrable \mathbb{G} -adapted dual strategies valued in Y . In stochastic programming, one always postulates that decisions be adapted to the filtration generated by the data process $\boldsymbol{\zeta}$. Thus, \boldsymbol{x} and \boldsymbol{y} are usually selected from the convex sets $X(\mathbb{F})$ and $Y(\mathbb{F})$, respectively. This is an abstract formulation of the standard requirement that decisions be chosen non-anticipatively with respect to the underlying data process, see e.g. (38).

For each stage $t = 1, \dots, T$ there is a cost function $c_t : X^t \times \Theta_t \rightarrow \mathbb{R}$ and a constraint function $f_t : X^t \times \Xi_t \rightarrow \mathbb{R}^{m_t}$ which are Borel measurable and bounded. This minimal requirement will be further tightened, below, to ensure applicability of the approximation schemes presented in Sects. 3 and 5. A general (nonlinear) multistage stochastic program (MSP) can now be formulated as

$$\begin{aligned} &\underset{\boldsymbol{x} \in X(\mathbb{F})}{\text{minimize}} \quad E \left(\sum_{t=1}^T c_t(\boldsymbol{x}^t, \boldsymbol{\eta}_t) \right) && (\mathcal{P}) \\ &\text{s.t.} \quad f_t(\boldsymbol{x}^t, \boldsymbol{\xi}_t) \leq 0 \quad P\text{-a.s.} \quad \forall t = 1, \dots, T. \end{aligned}$$

The objective criterion is to minimize the expectation of total cost, i.e., the sum of the stagewise cost functions. Decisions are subject to the stagewise constraints, which are assumed to hold almost surely with respect to the probability measure P . Moreover, as mentioned above, decisions must be selected non-anticipatively, i.e., they must be adapted to the filtration \mathbb{F} induced by the data process. Note that the dual decision process comes into play later when we establish a Lagrangian reformulation of problem \mathcal{P} . By our assumptions on the cost and constraint functions, the minimization problem \mathcal{P} is in fact well-defined.⁴ For the further analysis, we will impose the following regularity conditions ($t = 1, \dots, T$):

- (C1) c_t is convex in x^t , concave in η_t , and continuous on $X^t \times \Theta_t$;
- (C2) f_t is additively separable, $f_t = g_t + h_t$, where the mappings $g_t : X^t \rightarrow \mathbb{R}^{m_t}$ and $h_t : \Xi^t \rightarrow \mathbb{R}^{m_t}$ are componentwise convex and continuous;
- (C3) X_t is a convex compact subset of \mathbb{R}^{m_t} , and Y_t is the closed nonnegative orthant of \mathbb{R}^{m_t} .

The first step towards a flexible approximation scheme consists in a reformulation and generalization of problem \mathcal{P} . To this end, assume that $\mathbb{G} = \{\mathcal{G}^t\}_{t=1}^T$ and $\mathbb{H} = \{\mathcal{H}^t\}_{t=1}^T$ are arbitrary filtrations on the probability space (Ω, Σ, P) . As inspired by Wright (44), we can now define a family of optimization problems, which depend parametrically on the two filtrations \mathbb{G} and \mathbb{H} as well as on the data processes η and ξ .

$$\begin{aligned} & \underset{x \in X(\mathbb{G})}{\text{minimize}} \quad E \left(\sum_{t=1}^T c_t(x^t, \eta_t) \right) && (\mathcal{P}(\mathbb{G}, \mathbb{H}; \eta, \xi)) \\ & \text{s.t.} \quad E(f_t(x^t, \xi_t) \mid \mathcal{H}^t) \leq 0 \quad P\text{-a.s.} \quad \forall t = 1, \dots, T. \end{aligned}$$

Note that problem \mathcal{P} can be identified with problem $\mathcal{P}(\mathbb{F}, \mathbb{F}; \eta, \xi)$ since \mathbb{F} stands for the filtration induced by the process $\zeta = (\eta, \xi)$. The development of suitable approximation schemes will be based on a Lagrangian reformulation of the stochastic program $\mathcal{P}(\mathbb{G}, \mathbb{H}; \eta, \xi)$. The *Lagrangian*⁵ $L : X \times Y \times \Theta \times \Xi \rightarrow \mathbb{R}$ associated with the problem data is defined through

$$L(x, y; \eta, \xi) := \sum_{t=1}^T c_t(x^t, \eta_t) + y_t \cdot f_t(x^t, \xi_t).$$

By the basic regularity conditions, L is convex in (x, ξ) , concave in (y, η) , and continuous on its domain. Let us now establish a useful reformulation of the stochastic program $\mathcal{P}(\mathbb{G}, \mathbb{H}; \eta, \xi)$ in terms of the corresponding Lagrangian.

⁴ However, \mathcal{P} neither needs to be solvable nor feasible.

⁵ Rigorously speaking, L should be termed the *Lagrangian density*. However, for the sake of transparent terminology, it will simply be referred to as the *Lagrangian* in this article.

Proposition 1 *Under the conditions (C1), (C2), and (C3) we have*

$$\inf \mathcal{P}(\mathbb{G}, \mathbb{H}; \boldsymbol{\eta}, \boldsymbol{\xi}) = \inf_{\mathbf{x} \in X(\mathbb{G})} \sup_{\mathbf{y} \in Y(\mathbb{H})} E[L(\mathbf{x}, \mathbf{y}; \boldsymbol{\eta}, \boldsymbol{\xi})].$$

Proof Extend Wright’s argument (44, Sect. 4) to the nonlinear convex case. \square

3 Bounds based on space-discretization

Assume that there are stochastic processes $\boldsymbol{\eta}^u$ and $\boldsymbol{\xi}^u$ with state spaces Θ and Ξ , respectively. Thus, $\boldsymbol{\eta}^u$ takes values in the same space as the data process $\boldsymbol{\eta}$ introduced in Sect. 2. Similarly, $\boldsymbol{\xi}^u$ takes values in the same space as $\boldsymbol{\xi}$. As in the case of the original data processes, notation can be simplified if we introduce a combined process $\boldsymbol{\zeta}^u = (\boldsymbol{\eta}^u, \boldsymbol{\xi}^u)$ with state space Z . One can think of $\boldsymbol{\zeta}^u$ as an approximation of $\boldsymbol{\zeta}$. For the further argumentation, we need the filtration \mathbb{F}^u induced by the process $\boldsymbol{\zeta}^u$, i.e., $\mathbb{F}^u := \{\mathcal{F}^{u,t}\}_{t=1}^T$ where $\mathcal{F}^{u,t} := \sigma(\boldsymbol{\zeta}^{u,t})$, and we use the convention $\mathcal{F}^u := \mathcal{F}^{u,T}$. In the remainder of this section, we assume the following conditions to hold for suitable versions of the conditional expectations, respectively.

$$E(\mathbf{x}|\mathcal{F}) \in X(\mathbb{F}) \quad \text{for all } \mathbf{x} \in X(\mathbb{F}^u), \tag{3.1a}$$

$$E(\mathbf{y}|\mathcal{F}^u) \in Y(\mathbb{F}^u) \quad \text{for all } \mathbf{y} \in Y(\mathbb{F}), \tag{3.1b}$$

$$E(\boldsymbol{\xi}^u|\mathcal{F}) = \boldsymbol{\xi}, \tag{3.1c}$$

$$E(\boldsymbol{\eta}|\mathcal{F}^u) = \boldsymbol{\eta}^u. \tag{3.1d}$$

Next, assume that there is another process $\boldsymbol{\zeta}^l = (\boldsymbol{\eta}^l, \boldsymbol{\xi}^l)$ such that $\boldsymbol{\eta}^l$ and $\boldsymbol{\xi}^l$ are valued in Θ and Ξ , respectively. Again, $\boldsymbol{\zeta}^l$ is meant to approximate the data process $\boldsymbol{\zeta}$. The induced filtration \mathbb{F}^l is constructed as usual, i.e., $\mathbb{F}^l := \{\mathcal{F}^{l,t}\}_{t=1}^T$ where $\mathcal{F}^{l,t} := \sigma(\boldsymbol{\zeta}^{l,t})$, and we use the convention $\mathcal{F}^l := \mathcal{F}^{l,T}$. From now on, the following conditions are assumed to hold for suitable versions of the conditional expectations, respectively.

$$E(\mathbf{x}|\mathcal{F}^l) \in X(\mathbb{F}^l) \quad \text{for all } \mathbf{x} \in X(\mathbb{F}), \tag{3.2a}$$

$$E(\mathbf{y}|\mathcal{F}) \in Y(\mathbb{F}) \quad \text{for all } \mathbf{y} \in Y(\mathbb{F}^l), \tag{3.2b}$$

$$E(\boldsymbol{\xi}|\mathcal{F}^l) = \boldsymbol{\xi}^l, \tag{3.2c}$$

$$E(\boldsymbol{\eta}^l|\mathcal{F}) = \boldsymbol{\eta}. \tag{3.2d}$$

From a computational point of view, it is desired that $\boldsymbol{\zeta}^u$ and $\boldsymbol{\zeta}^l$ have discrete distributions. Sect. 4 will provide a constructive proof for the existence of discrete processes, which satisfy the above conditions, and the flexibility in their construction will thoroughly be investigated. In this section, however, we will argue that such processes (if they exist) can be used to construct bounds on

the optimal value of any stochastic program, which complies with the basic regularity conditions. The following theorems make this statement precise.

Theorem 1 *Assume the conditions (C1), (C2), and (C3) hold. If the data process ζ and its approximation ζ^u satisfy (3.1), then*

$$\inf \mathcal{P}(\mathbb{F}^u, \mathbb{F}^u; \eta^u, \xi^u) \geq \inf \mathcal{P}.$$

Proof The proof is based on restriction of the primal feasible set, repeated application of the conditional Jensen inequality, and relaxation of the dual feasible set. Concretely speaking, we find

$$\begin{aligned} \inf \mathcal{P} &\leq \inf_{\mathbf{x} \in X(\mathbb{F}^u)} \sup_{\mathbf{y} \in Y(\mathbb{F})} E[L(E[\mathbf{x}|\mathcal{F}], \mathbf{y}; \eta, E[\xi^u|\mathcal{F}])] \\ &\leq \inf_{\mathbf{x} \in X(\mathbb{F}^u)} \sup_{\mathbf{y} \in Y(\mathbb{F})} E[L(\mathbf{x}, \mathbf{y}; \eta, \xi^u)] \\ &\leq \inf_{\mathbf{x} \in X(\mathbb{F}^u)} \sup_{\mathbf{y} \in Y(\mathbb{F})} E[L(\mathbf{x}, E[\mathbf{y}|\mathcal{F}^u]; E[\eta|\mathcal{F}^u], \xi^u)] \\ &\leq \inf_{\mathbf{x} \in X(\mathbb{F}^u)} \sup_{\mathbf{y} \in Y(\mathbb{F}^u)} E[L(\mathbf{x}, \mathbf{y}; \eta^u, \xi^u)]. \end{aligned}$$

The first inequality follows from (3.1a) and (3.1c). It entails restriction of the primal feasible set to those decisions which are representable as the conditional expectation (given \mathcal{F}) of some $\mathbf{x} \in X(\mathbb{F}^u)$. Next, we use the conditional version of Jensen's inequality for moving the conditional expectations out of the Lagrangian. This is allowed by convexity of the Lagrangian in the first and the fourth arguments, and since \mathbf{y} and η are \mathcal{F} -measurable. Repeated application of the conditional Jensen inequality justifies the third line. Here, we exploit concavity of the Lagrangian in the second and the third arguments together with the \mathcal{F}^u -measurability of \mathbf{x} and ξ^u . Finally, the fourth inequality holds by the assumptions (3.1b) and (3.1d). It entails relaxation of the dual feasible set from those decisions which are representable as the conditional expectation (given \mathcal{F}^u) of some $\mathbf{y} \in Y(\mathbb{F})$ to all decisions in $Y(\mathbb{F}^u)$. \square

Theorem 2 *Assume the conditions (C1), (C2), and (C3) hold. If the data process ζ and its approximation ζ^l satisfy (3.2), then*

$$\inf \mathcal{P}(\mathbb{F}^l, \mathbb{F}^l; \eta^l, \xi^l) \leq \inf \mathcal{P}.$$

Proof The proof is analogous to that of Theorem 1. Without a detailed description of the involved manipulations, we may therefore state the following chain of inequalities

$$\begin{aligned}
 \inf \mathcal{P} &\geq \inf_{\mathbf{x} \in X(\mathbb{F})} \sup_{\mathbf{y} \in Y(\mathbb{F}^l)} E \left[L(\mathbf{x}, E[\mathbf{y}|\mathcal{F}]; E[\boldsymbol{\eta}^l|\mathcal{F}], \boldsymbol{\xi}) \right] \\
 &\geq \inf_{\mathbf{x} \in X(\mathbb{F})} \sup_{\mathbf{y} \in Y(\mathbb{F}^l)} E \left[L(\mathbf{x}, \mathbf{y}; \boldsymbol{\eta}^l, \boldsymbol{\xi}) \right] \\
 &\geq \inf_{\mathbf{x} \in X(\mathbb{F})} \sup_{\mathbf{y} \in Y(\mathbb{F}^l)} E \left[L(E[\mathbf{x}|\mathcal{F}^l], \mathbf{y}; \boldsymbol{\eta}^l, E[\boldsymbol{\xi}|\mathcal{F}^l]) \right] \\
 &\geq \inf_{\mathbf{x} \in X(\mathbb{F}^l)} \sup_{\mathbf{y} \in Y(\mathbb{F}^l)} E \left[L(\mathbf{x}, \mathbf{y}; \boldsymbol{\eta}^l, \boldsymbol{\xi}^l) \right].
 \end{aligned}$$

The first inequality follows from (3.2b) and (3.2d), while the second and the third inequalities are due to the conditional Jensen inequality. Finally, the fourth inequality holds by the assumptions (3.2a) and (3.2c). □

If ζ^u represents a finitely supported discrete process, then any \mathbb{F}^u -adapted primal or dual strategy is finitely supported, as well. In this case, the extensive form (4) of $\mathcal{P}(\mathbb{F}^u, \mathbb{F}^u; \boldsymbol{\eta}^u, \boldsymbol{\xi}^u)$ involves only a finite number of variables and constraints, implying that it principally allows for numerical solution. One may argue in a similar way that the extensive form of $\mathcal{P}(\mathbb{F}^l, \mathbb{F}^l; \boldsymbol{\eta}^l, \boldsymbol{\xi}^l)$ is computationally tractable if ζ^l represents a finitely supported discrete process. These observations together with Theorems 1 and 2 show that we can (numerically) calculate upper and lower bounds on $\inf \mathcal{P}$ whenever it is possible to find discrete processes ζ^u and ζ^l subject to the assumptions (3.1) and (3.2), respectively. The next section presents a systematic approach towards constructing such approximate processes.

4 Scenario generation

Given a stochastic process ζ , the construction of a discrete approximate process with finite support is referred to as *scenario generation*. In practice, the support of the approximate process should consist of few discretization points (i.e., scenarios) only. In addition, it is usually required that the distributions of ζ and its discrete approximation are close in a certain sense, e.g. with respect to some \mathcal{L}^p -norm. In this section, we focus on the construction of a discrete process ζ^u subject to the conditions (3.1). The construction of a discrete process ζ^l subject to the symmetric conditions (3.2) is completely analogous and will be omitted for brevity of exposition. Notice that the difficulty of finding a valid process ζ^u may depend on the properties of the underlying data process ζ . Sometimes, finding a suitable ζ^u may even be impossible. We will show that ζ^u can systematically be constructed whenever ζ has a compact state space and belongs to some class of autoregressive processes. Compactness of Z will be assumed throughout this section. For didactic reasons, we study the one-stage case first.

4.1 Basic one-stage case

In the one-stage case under consideration, the inclusions (3.1a) and (3.1b) are automatically satisfied; cf. the argument below. We may thus focus on validating the equalities (3.1c) and (3.1d) under the assumption that Θ and Ξ are compact polytopes. In this section, we find it more natural to work directly with the induced probability spaces rather than referring to the abstract sample space. Concretely speaking, we will mostly work with marginal and conditional distributions. Notational conventions are agreed on in the following definition.

Definition 2 *Let α and β be finite-dimensional random vectors on (Ω, Σ, P) taking values in some Borel sets A and B , respectively. The (marginal) distribution of β is denoted by P_β , while $P_{\beta|\alpha}$ stands for the (regular) conditional distribution of β given $\alpha = \alpha$. Thus, we have*

$$\left. \begin{aligned} P_\beta(B) &= P(\beta \in B) \\ P_{\beta|\alpha}(B|\alpha) &= P(\beta \in B|\alpha = \alpha) \end{aligned} \right\} \quad \forall B \in \mathcal{B}(B), \quad \alpha \in A.$$

Note that the data process ζ and its distribution P_ζ are a priori known. We will construct ζ^u by specifying the conditional distribution $P_{\zeta^u|\zeta}$. To be a regular conditional distribution, $P_{\zeta^u|\zeta}(B|\zeta)$ must be a probability measure on $\mathcal{B}(Z)$ for fixed $\zeta \in Z$ and a Borel measurable function on Z for fixed $B \in \mathcal{B}(Z)$. Then, the joint distribution of ζ and ζ^u is uniquely determined by the product measure theorem (1, Theorem 2.6.2), while the conditional distribution $P_{\zeta|\zeta^u}$ is obtainable via Bayes' theorem (39, Theorem 1.31). It should be pointed out that this construction also guarantees the existence of a rich enough sample space (Ω, Σ, P) on which both ζ and ζ^u are defined. In fact, we may simply identify Ω with $Z \times Z$, let Σ be the Borel field on $Z \times Z$, and let P be the joint distribution of ζ and ζ^u . In this argumentation, the right choice of (Ω, Σ, P) is known only after constructing ζ^u . However, there is no problem assuming that the sample space was chosen appropriately already at the outset.

Let us elaborate the above ideas in more detail. First, select a Borel measurable function $P_\Xi : \Xi \times Z \rightarrow [0, 1]$ such that

$$\sum_{e \in \text{ext } \Xi} P_\Xi(e|\zeta) = 1 \quad \text{and} \quad \sum_{e \in \text{ext } \Xi} e P_\Xi(e|\zeta) = \xi \quad \forall \zeta \in Z,$$

where $\text{ext } \Xi$ represents the set of extreme points of Ξ . $\text{ext } \Xi$ is finite since Ξ is a compact polytope. Observe that, if Ξ is a nondegenerate simplex, the above conditions uniquely determine $P_\Xi(\cdot|\zeta)$ on $\text{ext } \Xi$. In addition, introduce a Borel measurable function $\zeta_\Xi = (\eta_\Xi, \xi_\Xi) : \Xi \rightarrow Z$. Set $\xi_\Xi(e) = e$ and

$$\eta_\Xi(e) = \frac{\int_Z \eta P_\Xi(e|\zeta) P_\zeta(d\zeta)}{\int_Z P_\Xi(e|\zeta) P_\zeta(d\zeta)}$$

if the denominator is nonzero. Otherwise, $\eta_{\Xi}(e)$ is set to $\int_Z \eta P_{\zeta}(d\zeta)$. Using these conventions, we can specify the conditional distribution of ζ^u given $\zeta = \zeta$, i.e.,

$$P_{\zeta^u|\zeta}(B|\zeta) := \sum_{e \in \text{ext } \Xi} P_{\Xi}(e|\zeta) \delta_{\zeta_{\Xi}(e)}(B). \tag{4.3}$$

Here, $\delta_{\zeta_{\Xi}(e)}$ denotes the Dirac measure concentrated at $\zeta_{\Xi}(e)$. It is easily seen that $P_{\zeta^u|\zeta}$ is in fact a probability measure in its first argument and a Borel measurable function in its second argument. Moreover, we have

$$P_{\zeta^u|\zeta}(\cdot|\zeta) \ll \sum_{e \in \text{ext } \Xi} \delta_{\zeta_{\Xi}(e)}(\cdot), \tag{4.4}$$

i.e., the regular conditional distribution of ζ^u given $\zeta = \zeta$ is absolutely continuous with respect to a discrete measure independent of the parameter ζ . The conditional density is $\zeta^u \mapsto P_{\Xi}(\zeta^u|\zeta)$. Having constructed a candidate random vector ζ^u , we now should verify the conditions (3.1a) through (3.1d). In the one-stage case under consideration, the conditions (3.1a) and (3.1b) are trivially satisfied since X and Y are closed convex sets. In fact, the support of the conditional expectation of some random vector with respect to any σ -algebra is necessarily covered by the convex hull of the support of this very random vector. Validation of the conditions (3.1c) and (3.1d) requires some more work. To begin with, let us verify that

$$E(\xi^u|\zeta) = \int_Z \xi^u P_{\zeta^u|\zeta}(d\zeta^u|\zeta) = \sum_{e \in \text{ext } \Xi} e P_{\Xi}(e|\zeta) = \xi \quad P\text{-a.s.}$$

The first equality holds by a standard result in probability theory (1, Sect. 6.6) while the second and the third equalities are due to the defining properties of the regular conditional probability $P_{\zeta^u|\zeta}$ and the measurable function P_{Ξ} , respectively. Thus, (3.1c) follows. In a next step, we will argue that

$$E(\eta|\zeta^u) = \int_Z \eta P_{\zeta|\zeta^u}(d\zeta|\zeta^u) = \eta_{\Xi}(\xi^u) = \eta^u \quad P\text{-a.s.}$$

As before, the first equality holds by a standard result, whereas the second equality follows from the measure-theoretic version of Bayes' theorem (39, Theorem 1.31), which applies due to (4.4). Finally, the third equality is immediate from the construction of ζ^u . This establishes (3.1d).

4.2 Flexible one-stage case

The one-stage case requires some more investigation since the specific construction in Sect. 4.1 does not provide much flexibility in choosing the discrete

approximate process ζ^u . Generally, such ζ^u will not be close to the data process ζ with respect to some \mathcal{L}^p -norm, say. In any case, the conditions (3.1c) and (3.1d) guarantee matching of the first moments, i.e.,

$$E(\xi^u) = E(E(\xi^u|\mathcal{F})) = E(\xi) \quad \text{and} \quad E(\eta^u) = E(E(\eta|\mathcal{F}^u)) = E(\eta).$$

Moreover, equivalence of the second order cross-moments holds,

$$E(\xi^u \eta^{u\top}) = E(\xi^u E(\eta|\mathcal{F}^u)^\top) = E(\xi^u \eta^\top) = E(E(\xi^u|\mathcal{F}) \eta^\top) = E(\xi \eta^\top),$$

but the higher order moments of ζ and ζ^u are generically different. If we want to ensure closeness of the data process and its discrete approximation beyond first order moment matching, we need a more flexible approach for constructing ζ^u . Let us therefore assume that the data process is representable as

$$\zeta = \sum_{\lambda \in \Lambda} 1_{\{\lambda=\lambda\}} \zeta_\lambda, \tag{4.5}$$

where the random variable λ and the random vectors $\{\zeta_\lambda\}_{\lambda \in \Lambda}$ are mutually independent, and Λ is a finite index set. In particular, assume that λ takes values in Λ , while ζ_λ is supported on a compact polytope $Z_\lambda = \Theta_\lambda \times \Xi_\lambda$ with $\Theta_\lambda \subset \Theta$ and $\Xi_\lambda \subset \Xi$ for all $\lambda \in \Lambda$. By (4.5) and the independence assumption, the distribution of ζ can be written as a convex combination of the distributions of the ζ_λ with convex weights $P(\lambda = \lambda)$. Moreover, if some process ζ' with the same distribution as ζ allows for a representation of the form (4.5), then, without loss of generality, we may set $\zeta := \zeta'$. Any such redefinition is unproblematic since only the distribution of ζ has practical relevance for the stochastic program \mathcal{P} . These insights suggest that a decomposition as in (4.5) always exists and that the diameters of the state spaces Z_λ can be made uniformly small.⁶ Next, we apply the method of Sect. 4.1 to each ζ_λ separately. Concretely speaking, we construct random vectors $\{\zeta_\lambda^u\}_{\lambda \in \Lambda}$ with the properties

- (i) $E(\xi_\lambda^u | \zeta_\lambda) = \xi_\lambda$ P -a.s., $\lambda \in \Lambda$;
- (ii) $E(\eta_\lambda | \zeta_\lambda^u) = \eta_\lambda^u$ P -a.s., $\lambda \in \Lambda$;
- (iii) λ and the paired random vectors $\{(\zeta_\lambda, \zeta_\lambda^u)\}_{\lambda \in \Lambda}$ are mutually independent.

Then, we define a candidate process

$$\zeta^u = \sum_{\lambda \in \Lambda} 1_{\{\lambda=\lambda\}} \zeta_\lambda^u, \tag{4.6}$$

which reflects the structure of (4.5). As in Sect. 4.1, we have to verify that this process complies with the conditions (3.1a) through (3.1d). Again, the relations (3.1a) and (3.1b) are trivially satisfied since we operate in a one-stage framework. In order to check the identities (3.1c) and (3.1d), we observe that

⁶ Note that it is possible to represent any compactly supported distribution as a finite convex combination of distributions with arbitrarily small supports; see also the example below.

$$\sigma(\zeta) \subset \sigma(\lambda, \{\zeta_\lambda\}_{\lambda \in \Lambda}) \quad \text{and} \quad \sigma(\zeta^u) \subset \sigma(\lambda, \{\zeta_\lambda^u\}_{\lambda \in \Lambda}). \tag{4.7}$$

For instance, the first inclusion holds because ζ is the image of λ and the ζ_λ 's under a continuous map (with respect to the Euclidean topology on Z on one hand and the product of the discrete topology on Λ and the Euclidean topologies on the Z_λ 's on the other hand). We may thus conclude that

$$\begin{aligned} E(\xi^u | \zeta) &= E\left(E\left(\sum_{\lambda \in \Lambda} 1_{\{\lambda=\lambda\}} \xi_\lambda^u \mid \lambda, \{\zeta_{\lambda'}\}_{\lambda' \in \Lambda} \right) \mid \zeta \right) \\ &= E\left(\sum_{\lambda \in \Lambda} 1_{\{\lambda=\lambda\}} E(\xi_\lambda^u | \zeta_\lambda) \mid \zeta \right) = E(\xi | \zeta) = \xi \quad P\text{-a.s.} \end{aligned}$$

The first equality uses (4.7) while the second and the third equalities are based on the properties (iii) and (i) stated above. By a similar reasoning we obtain

$$\begin{aligned} E(\eta | \zeta^u) &= E\left(E\left(\sum_{\lambda \in \Lambda} 1_{\{\lambda=\lambda\}} \eta_\lambda \mid \lambda, \{\zeta_{\lambda'}^u\}_{\lambda' \in \Lambda} \right) \mid \zeta^u \right) \\ &= E\left(\sum_{\lambda \in \Lambda} 1_{\{\lambda=\lambda\}} E(\eta_\lambda | \zeta_\lambda^u) \mid \zeta^u \right) = E(\eta^u | \zeta^u) = \eta^u \quad P\text{-a.s.} \end{aligned}$$

As before, the first equality is due to (4.7) while the second and the third equalities follow from the properties (iii) and (ii), respectively.

Example 1 We show a possibility to construct λ and $\{\zeta_\lambda\}_{\lambda \in \Lambda}$. Suppose that Λ is a finite index set and that $\{\tilde{Z}_\lambda\}_{\lambda \in \Lambda}$ is a disjoint set partition of Z , i.e.,

$$Z = \bigcup_{\lambda \in \Lambda} \tilde{Z}_\lambda \quad \text{and} \quad \tilde{Z}_\lambda \cap \tilde{Z}_{\lambda'} = \emptyset \quad \text{for } \lambda \neq \lambda'.$$

For simplicity of exposition, let us assume that $P(\zeta \in \tilde{Z}_\lambda) > 0$ for each $\lambda \in \Lambda$. Moreover, assume the closure of \tilde{Z}_λ to be representable as $Z_\lambda = \Theta_\lambda \times \Xi_\lambda$, where $\Theta_\lambda \subset \Theta$ and $\Xi_\lambda \subset \Xi$ are compact polytopes. We can now introduce a discrete random variable λ valued in Λ . Its distribution function P_λ is defined through

$$P_\lambda(B) := P(\zeta \in \cup_{\lambda \in B} \tilde{Z}_\lambda) \quad \text{for } B \subset \Lambda.$$

Moreover, for each $\lambda \in \Lambda$ we can introduce a random vector ζ_λ valued in Z whose distribution P_{ζ_λ} is defined through

$$P_{\zeta_\lambda}(B) := \frac{P(\zeta \in B \cap \tilde{Z}_\lambda)}{P(\zeta \in \tilde{Z}_\lambda)} \quad \text{for } B \in \mathcal{B}(Z).$$

Assuming λ and the $\{\zeta_\lambda\}_{\lambda \in \Lambda}$ to be mutually independent, $\sum_{\lambda \in \Lambda} 1_{\{\lambda=\lambda\}} \zeta_\lambda$ and ζ are equal in distribution and may be identified. With ζ^u as in (4.6), we have

$$\|\zeta - \zeta^u\|_\infty \leq \max\{\text{diam}(Z_\lambda) \mid \lambda \in \Lambda\}$$

since both ζ_λ and ζ_λ^u are supported on Z_λ for each $\lambda \in \Lambda$. By making the diameters of the Z_λ uniformly small, we can thus construct a discrete process ζ^u that approximates ζ arbitrarily well with respect to the \mathcal{L}^∞ -norm.

4.3 Independent multistage case

Let us now investigate the most simple multistage case where the random vectors $\{\zeta_t\}_{t=1}^T$ are mutually independent. Note that this independence assumption will later be relaxed. In the present section, the random vectors ζ_t^u corresponding to the discrete approximate process ζ^u are constructed successively with increasing t . For notational convenience, let P_t^u be the (regular) conditional distribution of ζ_t^u given $\zeta = \zeta$ and $\zeta^{u,t-1} = \zeta^{u,t-1}$. The mapping P_t^u is chosen such that the following statements hold true:

- (i) $E(\xi_t^u \mid \zeta_t, \zeta^{u,t-1}) = \xi_t$ P -a.s.;
- (ii) $E(\eta_t \mid \zeta_t^u, \zeta^{u,t-1}) = \eta_t^u$ P -a.s.;
- (iii) $\{\zeta_t^u\}$ is conditionally independent of $\{\zeta_\tau\}_{\tau \neq t}$ given $\{\zeta_t\} \cup \{\zeta_\tau^u\}_{\tau < t}$.

The conditions (i) and (ii) are e.g. satisfied if for all fixed $\{\zeta_\tau\}_{\tau \neq t}$ and $\{\zeta_\tau^u\}_{\tau < t}$ the mapping $(B, \zeta_t) \mapsto P_t^u(B \mid \zeta, \zeta^{u,t-1})$ represents a conditional distribution of the form (4.3) or one of its generalizations in the spirit of Sect. 4.2. Note that the discrete conditional scenarios and probabilities corresponding to ζ_t^u may now depend on $(\zeta, \zeta^{u,t-1})$. Furthermore, condition (iii) is e.g. satisfied if for every fixed $B \in \mathcal{B}(Z)$ the Borel measurable function $(\zeta, \zeta^{u,t-1}) \mapsto P_t^u(B \mid \zeta, \zeta^{u,t-1})$ is constant in ζ_τ for $\tau \neq t$ (for a survey of the basic properties of the conditional independence relation we refer to Appendix 7). These insights suggest that a discrete process ζ^u subject to the above conditions can systematically be obtained, and there is considerable flexibility in its construction. In particular, notice that we allow the $\{\zeta_t^u\}_{t=1}^T$ to be mutually dependent, which complicates scenario generation and makes it difficult to check the conditions (3.1). However, this extra flexibility has distinct numerical advantages and may accelerate convergence of solution algorithms; cf. e.g. the related arguments in (20, Sect. 4).

In analogy to the previous sections, we must prove that the exogenous data process ζ and the synthesized approximation ζ^u satisfy the requirements (3.1). In order to prove (3.1a) we choose some $x \in X(\mathbb{F}^u)$. By hypothesis, the support of x is covered by X , which is convex and closed. Moreover, it is known that the support of the conditional expectation $E(x \mid \mathcal{F})$ is a subset of the convex hull of the support of x .⁷ Consequently, $E(x \mid \mathcal{F})$ is valued in X almost surely. It

⁷ Without being rigorous, evaluating the (regular) conditional expectation of x with respect to any σ -algebra and at a fixed $\omega \in \Omega$ can be viewed as taking an infinite convex combination of the points in the support of x .

remains to be shown that $E(\mathbf{x}_t|\mathcal{F})$ is \mathcal{F}^t -measurable almost surely for each t . An equivalent statement is

$$E(\mathbf{x}_t | \zeta) = E(\mathbf{x}_t | \zeta^t) \quad P\text{-a.s. for } t = 1, \dots, T.$$

This, however, is true by Proposition 3 in the appendix, which asserts that the sets of random vectors $\{\zeta_\tau^u\}_{\tau \leq t}$ and $\{\zeta_\tau\}_{\tau > t}$ are conditionally independent given $\{\zeta_\tau\}_{\tau \leq t}$. The proof of (3.1b) is widely parallel. First, choose $\mathbf{y} \in Y(\mathbb{F})$. By repeating the above arguments, it can be shown that the conditional expectation $E(\mathbf{y}|\mathcal{F}^u)$ takes values in Y almost surely. Furthermore, one should verify that $E(\mathbf{y}_t|\mathcal{F}^u)$ is $\mathcal{F}^{u,t}$ -measurable almost surely for t arbitrary. This statement can be reformulated as

$$E(\mathbf{y}_t | \zeta^u) = E(\mathbf{y}_t | \zeta^{u,t}) \quad P\text{-a.s. for } t = 1, \dots, T,$$

and it holds true by Proposition 3, which asserts that the sets of random vectors $\{\zeta_\tau\}_{\tau \leq t}$ and $\{\zeta_\tau^u\}_{\tau > t}$ are conditionally independent given $\{\zeta_\tau^u\}_{\tau \leq t}$. Let us now check the remaining properties (3.1c) and (3.1d). From the above conditions we may deduce the following chain of equalities, which holds almost surely with respect to the probability measure P .

$$\begin{aligned} E(\xi_t^u | \zeta) &= E(E(\xi_t^u | \zeta, \zeta^{u,t-1}) | \zeta) \\ &= E(E(\xi_t^u | \zeta_t, \zeta^{u,t-1}) | \zeta) && \text{by (iii)} \\ &= E(\xi_t | \zeta) = \xi_t && \text{by (i)} \end{aligned}$$

As t is arbitrary, this argument proves (3.1c). Furthermore, we have

$$E(\eta_t | \zeta^u) = E(\eta_t | \zeta^{u,t}) = \eta_t^u \quad P\text{-a.s. for } t = 1, \dots, T.$$

The first equality follows from Proposition 4, which states that $\{\zeta_t\}$ and $\{\zeta_\tau^u\}_{\tau > t}$ are conditionally independent given $\{\zeta_\tau^u\}_{\tau \leq t}$. The second equality is due to (ii). Thus, (3.1d) is established.

One of the simplest ways to discretize a serially independent stochastic process is by discretizing each ζ_t individually as in the example of Sect. 4.2. It is easily seen that the pairs (ζ_t, ζ_t^u) can be chosen mutually independent for $t = 1, \dots, T$. Then, the conditions (i), (ii), and (iii) stated at the beginning of this section are trivially satisfied. If the underlying set partitions of the stagewise state spaces are suitably refined, $\|\zeta - \zeta^u\|_\infty$ can be made arbitrarily small.

4.4 Dependent multistage case

Let $\hat{\eta}$ and $\hat{\xi}$ be two stochastic processes with state spaces $\hat{\Theta}$ and $\hat{\Xi}$, respectively. As usual, we further introduce a combined stochastic process $\hat{\zeta} := (\hat{\eta}, \hat{\xi})$ with state space $\hat{Z} := \hat{\Theta} \times \hat{\Xi}$. Let us assume that the corresponding random vectors $\{\hat{\zeta}_t\}_{t=1}^T$ are mutually independent. Consequently, $\hat{\zeta}$ has the same structure as the

data processes considered in Sect. 4.3. In this section, however, we will study slightly more general data processes of the form

$$\zeta = (\eta, \xi) = (H^o \hat{\eta}, H^c \hat{\xi}), \tag{4.8}$$

where $H^o : \hat{\Theta} \rightarrow \Theta$ and $H^c : \hat{\Xi} \rightarrow \Xi$ are non-anticipative isomorphisms. In other words, these transformations are linear, bijective, and lower block-triangular with respect to the temporal structure. As far as H^o is concerned, for instance, this means that the matrix elements coupling η_t and $\hat{\eta}_s$ are zero for $s > t$. However, they may be nonzero for $s \leq t$, which implies that the η process may be autocorrelated. Since $\hat{\xi}$ consists of mutually independent random vectors, we can proceed as in Sect. 4.3 to construct a process $\hat{\xi}^u$. Thus, the processes $\hat{\xi}$ and $\hat{\xi}^u$ satisfy the conditions (3.1). Next, set

$$\zeta^u := (\eta^u, \xi^u) := (H^o \hat{\eta}^u, H^c \hat{\xi}^u).$$

Using the above definitions, we will prove that the processes ζ and ζ^u also comply with the conditions (3.1). To this end, we first notice that

$$\mathcal{F}^t := \sigma(\zeta^t) = \sigma(\hat{\zeta}^t) =: \hat{\mathcal{F}}^t \quad \text{and} \quad \mathcal{F}^{u,t} := \sigma(\zeta^{u,t}) = \sigma(\hat{\zeta}^{u,t}) =: \hat{\mathcal{F}}^{u,t}$$

for each t since the transformations H^o and H^c are linear, bijective, and non-anticipative. Thus, the filtrations induced by the processes ζ^u and $\hat{\zeta}^u$ are identical, and the proof of (3.1a) is as in Sect. 4.3. Moreover, the filtrations induced by the processes ζ and $\hat{\zeta}$ are identical, too. This implies that the proof of (3.1b) may also be copied from Sect. 4.3. Finally, the conditions (3.1c) and (3.1d) are immediate from the construction of ζ and ζ^u as well as linearity of the transformations H^o and H^c , i.e.,

$$\left. \begin{aligned} E(\xi^u | \mathcal{F}) &= E(H^c \hat{\xi}^u | \mathcal{F}) = H^c E(\hat{\xi}^u | \mathcal{F}) = H^c \hat{\xi} = \xi \\ E(\eta | \mathcal{F}^u) &= E(H^o \hat{\eta} | \mathcal{F}^u) = H^o E(\hat{\eta} | \mathcal{F}^u) = H^o \hat{\eta}^u = \eta^u \end{aligned} \right\} \quad P\text{-a.s.}$$

Notice that the data processes of the form (4.8) cover all ARMA processes and are general enough for many interesting applications. Bounds of the type considered in Sect. 3 are available also for stochastic programs involving more general nonlinear data processes (31). For instance, lognormal stochastic processes with serial correlations are investigated in (31, Sects. 6.3 and 6.4).

It should be mentioned that the approximation error is, again, completely under our control. The remarks at the end of Sect. 4.3 provide some guidelines on how to construct $\hat{\zeta}^u$ such that $\|\hat{\zeta} - \hat{\zeta}^u\|_\infty$ becomes arbitrarily small. Uniform continuity of the linear mappings H^o and H^c thus implies that $\|\zeta - \zeta^u\|_\infty$ can be made arbitrarily small, as well.

5 Bounds based on stage-aggregation

After having studied a particular scenario generation method providing deterministic error bounds, we now turn attention to stage aggregation, which is often inevitable to achieve reasonable problem dimensions. In doing so, we use similar techniques as in Sects. 3 and 4.

Let us introduce two *aggregation operators* \uparrow and \downarrow mapping the finite index set $\{1, \dots, T\}$ to itself (Fig. 1). We will refer to \uparrow and \downarrow as a pair of upper and lower aggregation operators if the following conditions hold:

- (i) **monotonicity:** both \uparrow and \downarrow are monotonically increasing;
- (ii) **idempotence:** $\uparrow \circ \uparrow = \uparrow, \downarrow \circ \downarrow = \downarrow, \uparrow \circ \downarrow = \downarrow,$ and $\downarrow \circ \uparrow = \uparrow$;
- (iii) **ordering:** $\downarrow \leq 1 \leq \uparrow$.

Here, 1 denotes the identity mapping on $\{1, \dots, T\}$. As follows from the defining conditions, the two aggregation operators are uniquely determined by their fixed point sets $\{t \mid \uparrow(t) = t\}$ and $\{t \mid \downarrow(t) = t\}$. Note that these sets coincide with the ranges of \uparrow and \downarrow , respectively, and are equal by the idempotence property (ii). Next, introduce σ -algebras $\mathcal{G}^{\uparrow,t} := \mathcal{F}^{\uparrow(t)}$ and $\mathcal{G}^{\downarrow,t} := \mathcal{F}^{\downarrow(t)}$ for all t , and define $\mathbb{G}^{\uparrow} := \{\mathcal{G}^{\uparrow,t}\}_{t=1}^T$ and $\mathbb{G}^{\downarrow} := \{\mathcal{G}^{\downarrow,t}\}_{t=1}^T$. Note that, by monotonicity of the aggregation operators, \mathbb{G}^{\uparrow} and \mathbb{G}^{\downarrow} represent specific filtrations on the sample space. By the ordering property (iii), the filtration \mathbb{F} induced by the data process is a subfiltration of \mathbb{G}^{\uparrow} in the sense that $\mathcal{F}^t \subset \mathcal{G}^{\uparrow,t}$ for each t . Moreover, \mathbb{F} is a superfiltration of \mathbb{G}^{\downarrow} in the sense that $\mathcal{F}^t \supset \mathcal{G}^{\downarrow,t}$ for each t .

In the present section, we will assume that the state spaces Z_t are equal for all t . This may be postulated without loss of generality if the dimension of the random vectors η_t and ξ_t is constant over time. Furthermore, assume that there are stochastic processes $\zeta^{\uparrow} = (\eta^{\uparrow}, \xi^{\uparrow})$ and $\zeta^{\downarrow} = (\eta^{\downarrow}, \xi^{\downarrow})$, both of which are valued in the common state space Z . These processes are determined through the relations $\zeta_t^{\uparrow} := \zeta_{\uparrow(t)}$ and $\zeta_t^{\downarrow} := \zeta_{\downarrow(t)}$ for $t = 1, \dots, T$, and they will be viewed as approximations of the data process ζ . The corresponding induced filtrations are constructed in the usual way, i.e., set

$$\begin{aligned} \mathcal{F}^{\uparrow,t} &:= \sigma(\zeta^{\uparrow,t}) \quad \text{for } t = 1, \dots, T, & \mathcal{F}^{\uparrow} &:= \mathcal{F}^{\uparrow,T}, & \mathbb{F}^{\uparrow} &:= \{\mathcal{F}^{\uparrow,t}\}_{t=1}^T, \\ \mathcal{F}^{\downarrow,t} &:= \sigma(\zeta^{\downarrow,t}) \quad \text{for } t = 1, \dots, T, & \mathcal{F}^{\downarrow} &:= \mathcal{F}^{\downarrow,T}, & \mathbb{F}^{\downarrow} &:= \{\mathcal{F}^{\downarrow,t}\}_{t=1}^T. \end{aligned}$$

Notice that \mathbb{F}^{\uparrow} is a subfiltration of \mathbb{G}^{\uparrow} , while \mathbb{F}^{\downarrow} is a subfiltration of \mathbb{G}^{\downarrow} ; equality holds if \uparrow or \downarrow coincides with the identity mapping, respectively. Moreover, \mathbb{F}^{\downarrow} is a subfiltration of the natural filtration \mathbb{F} induced by the data process, though \mathbb{F}^{\uparrow} is neither a sub- nor a superfiltration of \mathbb{F} . The following proposition establishes another useful connection between the newly introduced filtrations given that ζ follows a Markov process. The insights gained will be important for proving the main results in Sect. 5.1.

Proposition 2 *If the data process ζ is Markovian, then the following conditions hold for suitable versions of the conditional expectations, respectively.*

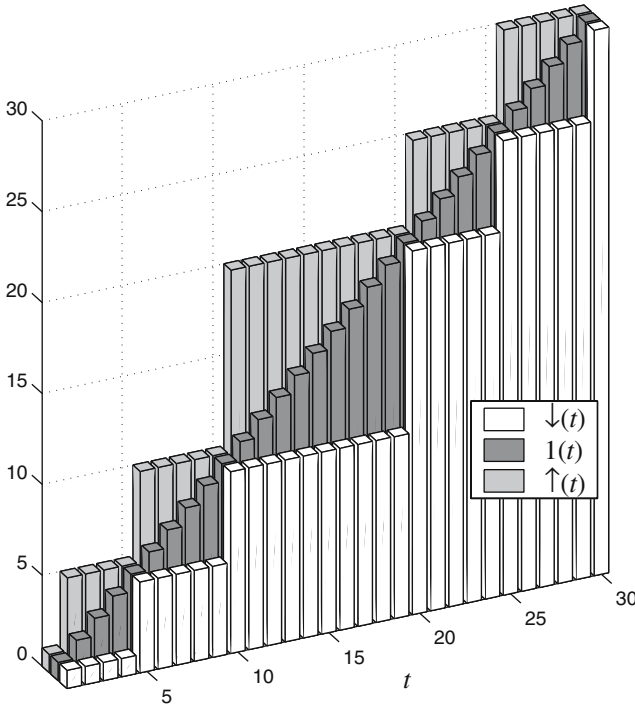


Fig. 1 Aggregation operators on a 30-point index set

- (i) $E(\mathbf{x}|\mathcal{F}^\downarrow) \in X(\mathbb{F}^\downarrow)$ for all $\mathbf{x} \in X(\mathbb{G}^\downarrow)$
- (ii) $E(\mathbf{y}|\mathcal{F}^\downarrow) \in Y(\mathbb{F}^\downarrow)$ for all $\mathbf{y} \in Y(\mathbb{G}^\downarrow)$
- (iii) $E(\mathbf{x}|\mathcal{F}^\uparrow) \in X(\mathbb{F}^\uparrow)$ for all $\mathbf{x} \in X(\mathbb{G}^\uparrow)$
- (iv) $E(\mathbf{y}|\mathcal{F}^\uparrow) \in Y(\mathbb{F}^\uparrow)$ for all $\mathbf{y} \in Y(\mathbb{G}^\uparrow)$

Proof The proof is based on a familiar argument known from Sect. 4.3. Choose an arbitrary $\mathbf{x} \in X(\mathbb{G}^\downarrow)$. The support of the conditional expectation $E(\mathbf{x}|\mathcal{F}^\downarrow)$ is a subset of the convex hull of the support of \mathbf{x} , which in turn is covered by X . Consequently, $E(\mathbf{x}|\mathcal{F}^\downarrow)$ is valued in X almost surely. It remains to be shown that $E(\mathbf{x}_t|\mathcal{F}^\downarrow)$ is $\mathcal{F}^{\downarrow,t}$ -measurable almost surely for each t . An equivalent statement is

$$E(\mathbf{x}_t | \zeta^\downarrow) = E(\mathbf{x}_t | \zeta^{\downarrow,t}) \quad P\text{-a.s. for } t = 1, \dots, T.$$

This, however, is true by Proposition 5 asserting that the sets of random vectors $\{\zeta_\tau\}_{\tau \leq \downarrow(t)}$ and $\{\zeta_\tau^\downarrow\}_{\downarrow(\tau) > \downarrow(t)}$ are conditionally independent given the set $\{\zeta_\tau^\downarrow\}_{\downarrow(\tau) \leq \downarrow(t)}$. Consequently, assertion (i) is established. Notice that the proofs of the other statements (ii) through (iv) are widely parallel and may thus be omitted. □

5.1 Makrov-martingale processes

In this section we will derive bounds on $\inf \mathcal{P}$ via stage-aggregation. Numerical complexity of the aggregated problems will be considerably reduced due to lower dimensionality. This feature can be exploited in optimization algorithms which would fail to cope with the original unaggregated problem. For technical reasons, we have to impose suitable regularity conditions on the underlying data process. In a first step, it is convenient to restrict attention to the class of Markov-martingale processes. Thus, we assume that

$$E(\xi_t | \mathcal{F}^s) = \xi_s \quad P\text{-a.s. for } 1 \leq s \leq t \leq T. \tag{5.9}$$

Notice that the martingale requirement can later be relaxed. The additional requirement that the data process be Markovian is necessary since we want to use Proposition 2 when constructing bounds (see Theorems 3 and 4).

Theorem 3 *Assume the conditions (C1), (C2), and (C3) hold. If the data process ζ represents a Markov-martingale, then*

$$\inf \mathcal{P}(\mathbb{F}^\downarrow, \mathbb{F}^\uparrow; \eta^\downarrow, \xi^\uparrow) \geq \inf \mathcal{P}.$$

Proof The claim is proved by using the martingale property of the data process, Jensen’s inequality, and specific relations between the relevant σ -algebras. In a preliminary step, we obtain

$$\begin{aligned} \inf \mathcal{P} &\leq \inf_{\mathbf{x} \in X(\mathbb{F})} \sup_{\mathbf{y} \in Y(\mathbb{F})} E \left(\sum_{t=1}^T c_t(\mathbf{x}^t, \eta_t) + \mathbf{y}_t \cdot f_t(\mathbf{x}^t, \xi_t^\uparrow) \right) \\ &\leq \inf_{\mathbf{x} \in X(\mathbb{G}^\downarrow)} \sup_{\mathbf{y} \in Y(\mathbb{F})} E \left(\sum_{t=1}^T c_t(\mathbf{x}^t, \eta_t) + \mathbf{y}_t \cdot f_t(\mathbf{x}^t, \xi_t^\uparrow) \right) \\ &\leq \inf_{\mathbf{x} \in X(\mathbb{G}^\downarrow)} \sup_{\mathbf{y} \in Y(\mathbb{F})} E \left(\sum_{t=1}^T c_t(\mathbf{x}^t, \eta_t^\downarrow) + \mathbf{y}_t \cdot f_t(\mathbf{x}^t, \xi_t^\uparrow) \right) \\ &\leq \inf_{\mathbf{x} \in X(\mathbb{G}^\downarrow)} \sup_{\mathbf{y} \in Y(\mathbb{G}^\uparrow)} E \left(\sum_{t=1}^T c_t(\mathbf{x}^t, \eta_t^\downarrow) + \mathbf{y}_t \cdot f_t(\mathbf{x}^t, \xi_t^\uparrow) \right). \end{aligned}$$

The first inequality is based on the fact that, due to the martingale property, the random vector ξ_t can be written as $E(\xi_t^\uparrow | \mathcal{F}^t)$. Application of the conditional Jensen inequality then yields the desired result. The second inequality is due to restriction of the primal feasible set. In order to justify the third inequality, we apply the conditional Jensen inequality once again and replace $E(\eta_t | \mathcal{G}^{\downarrow,t})$ by the random vector η_t^\downarrow . Finally, the fourth inequality is due to relaxation of the dual feasible set. Thus, we have shown

$$\inf \mathcal{P}(\mathbb{G}^\downarrow, \mathbb{G}^\uparrow; \eta^\downarrow, \xi^\uparrow) \geq \inf \mathcal{P}.$$

Next, we will use the Markov property of the data process to prove that

$$\inf \mathcal{P}(\mathbb{G}^\downarrow, \mathbb{G}^\uparrow; \boldsymbol{\eta}^\downarrow, \boldsymbol{\xi}^\uparrow) = \inf \mathcal{P}(\mathbb{F}^\downarrow, \mathbb{F}^\uparrow; \boldsymbol{\eta}^\downarrow, \boldsymbol{\xi}^\uparrow). \quad (5.10)$$

In order to show that the left hand side is no larger than the right hand side, we proceed as follows:

$$\begin{aligned} & \inf \mathcal{P}(\mathbb{G}^\downarrow, \mathbb{G}^\uparrow; \boldsymbol{\eta}^\downarrow, \boldsymbol{\xi}^\uparrow) \\ & \leq \inf_{\mathbf{x} \in X(\mathbb{F}^\downarrow)} \sup_{\mathbf{y} \in Y(\mathbb{G}^\uparrow)} E \left(\sum_{t=1}^T c_t(\mathbf{x}^t, \boldsymbol{\eta}_t^\downarrow) + \mathbf{y}_t \cdot f_t(\mathbf{x}^t, \boldsymbol{\xi}_t^\uparrow) \right) \\ & = \inf_{\mathbf{x} \in X(\mathbb{F}^\downarrow)} \sup_{\mathbf{y} \in Y(\mathbb{G}^\uparrow)} E \left(\sum_{t=1}^T c_t(\mathbf{x}^t, \boldsymbol{\eta}_t^\downarrow) + E(\mathbf{y}_t | \mathcal{F}^\uparrow) \cdot f_t(\mathbf{x}^t, \boldsymbol{\xi}_t^\uparrow) \right) \\ & \leq \inf_{\mathbf{x} \in X(\mathbb{F}^\downarrow)} \sup_{\mathbf{y} \in Y(\mathbb{F}^\uparrow)} E \left(\sum_{t=1}^T c_t(\mathbf{x}^t, \boldsymbol{\eta}_t^\downarrow) + \mathbf{y}_t \cdot f_t(\mathbf{x}^t, \boldsymbol{\xi}_t^\uparrow) \right). \end{aligned}$$

The first inequality is due to restriction of the primal feasible set, while the equality follows from the law of iterated conditional expectations and linearity of the Lagrangian in the dual decisions. By Proposition 2 (iv), for each $\mathbf{y} \in Y(\mathbb{G}^\uparrow)$ there exists a version of $E(\mathbf{y}_t | \mathcal{F}^\uparrow)$ which is an element of $Y(\mathbb{F}^\uparrow)$. Thus, the last inequality holds by relaxation of the dual feasible set. In order to prove that the left hand side of (5.10) is no less than the right hand side, we use an analogous argument, i.e., we basically interchange the manipulations with respect to primal and dual decisions.

$$\begin{aligned} & \inf \mathcal{P}(\mathbb{G}^\downarrow, \mathbb{G}^\uparrow; \boldsymbol{\eta}^\downarrow, \boldsymbol{\xi}^\uparrow) \\ & \geq \inf_{\mathbf{x} \in X(\mathbb{G}^\downarrow)} \sup_{\mathbf{y} \in Y(\mathbb{F}^\uparrow)} E \left(\sum_{t=1}^T c_t(\mathbf{x}^t, \boldsymbol{\eta}_t^\downarrow) + \mathbf{y}_t \cdot f_t(\mathbf{x}^t, \boldsymbol{\xi}_t^\uparrow) \right) \\ & \geq \inf_{\mathbf{x} \in X(\mathbb{G}^\downarrow)} \sup_{\mathbf{y} \in Y(\mathbb{F}^\uparrow)} E \left(\sum_{t=1}^T c_t(E(\mathbf{x}^t | \mathcal{F}^\downarrow), \boldsymbol{\eta}_t^\downarrow) + \mathbf{y}_t \cdot f_t(E(\mathbf{x}^t | \mathcal{F}^\downarrow), \boldsymbol{\xi}_t^\uparrow) \right) \\ & \geq \inf_{\mathbf{x} \in X(\mathbb{F}^\downarrow)} \sup_{\mathbf{y} \in Y(\mathbb{F}^\uparrow)} E \left(\sum_{t=1}^T c_t(\mathbf{x}^t, \boldsymbol{\eta}_t^\downarrow) + \mathbf{y}_t \cdot f_t(\mathbf{x}^t, \boldsymbol{\xi}_t^\uparrow) \right). \end{aligned}$$

Here, the first inequality is due to restriction of the dual feasible set. Then, we apply the conditional version of Jensen's inequality (notice that $\mathcal{F}^\downarrow = \mathcal{F}^\uparrow$ by the defining properties of the aggregation operators). The third inequality follows from Proposition 2 (i) and relaxation of the primal feasible set, i.e., for every $\mathbf{x} \in X(\mathbb{G}^\downarrow)$ there exists a version of $E(\mathbf{x} | \mathcal{F}^\downarrow)$ in $X(\mathbb{F}^\downarrow)$. Consequently, we have proved (5.10), and thus the claim is established. \square

Unlike in Theorem 1, applicability of the conditional Jensen inequality relies on the internal structure of the Lagrangian and not just its convexity properties.

Theorem 4 *Assume the conditions (C1), (C2), and (C3) hold. If the data process ζ represents a Markov-martingale, then*

$$\inf \mathcal{P}(\mathbb{F}^\uparrow, \mathbb{F}^\downarrow; \eta^\uparrow, \xi^\downarrow) \leq \inf \mathcal{P}.$$

Proof Apart from minor exceptions, the proof is analogous to that of Theorem 3. Without a detailed description of the involved manipulations, we may therefore state the following chain of inequalities

$$\begin{aligned} \inf \mathcal{P} &\geq \inf_{\mathbf{x} \in X(\mathbb{F})} \sup_{\mathbf{y} \in Y(\mathbb{F})} E \left(\sum_{t=1}^T c_t(\mathbf{x}^t, \eta_t^\uparrow) + \mathbf{y}_t \cdot f_t(\mathbf{x}^t, \xi_t) \right) \\ &\geq \inf_{\mathbf{x} \in X(\mathbb{F})} \sup_{\mathbf{y} \in Y(\mathbb{G}^\downarrow)} E \left(\sum_{t=1}^T c_t(\mathbf{x}^t, \eta_t^\uparrow) + \mathbf{y}_t \cdot f_t(\mathbf{x}^t, \xi_t) \right) \\ &\geq \inf_{\mathbf{x} \in X(\mathbb{F})} \sup_{\mathbf{y} \in Y(\mathbb{G}^\downarrow)} E \left(\sum_{t=1}^T c_t(\mathbf{x}^t, \eta_t^\uparrow) + \mathbf{y}_t \cdot f_t(\mathbf{x}^t, \xi_t^\downarrow) \right) \\ &\geq \inf_{\mathbf{x} \in X(\mathbb{G}^\uparrow)} \sup_{\mathbf{y} \in Y(\mathbb{G}^\downarrow)} E \left(\sum_{t=1}^T c_t(\mathbf{x}^t, \eta_t^\uparrow) + \mathbf{y}_t \cdot f_t(\mathbf{x}^t, \xi_t^\downarrow) \right). \end{aligned}$$

When applying the conditional Jensen inequality in the third line, we explicitly use separability of the constraint functions⁸ with respect to decisions and random parameters (note that \mathbf{x}^t is generally not $\mathcal{G}^{\downarrow,t}$ -measurable). Moreover, we substitute ξ_t^\downarrow for $E(\xi_t | \mathcal{G}^{\downarrow,t})$, which is allowed by the martingale property of the data process. In summary, this yields

$$\inf \mathcal{P}(\mathbb{G}^\uparrow, \mathbb{G}^\downarrow; \eta^\uparrow, \xi^\downarrow) \leq \inf \mathcal{P}.$$

As in the proof of Theorem 3, we will now use the Markov property of the data process to justify the equality

$$\inf \mathcal{P}(\mathbb{G}^\uparrow, \mathbb{G}^\downarrow; \eta^\uparrow, \xi^\downarrow) = \inf \mathcal{P}(\mathbb{F}^\uparrow, \mathbb{F}^\downarrow; \eta^\uparrow, \xi^\downarrow). \tag{5.11}$$

⁸ It should be emphasized that separability of the constraint functions is not needed in the proof of Theorem 3.

By restricting the primal feasible set, using the law of iterated conditional expectations, and relaxing the dual feasible set, we obtain

$$\begin{aligned}
 & \inf \mathcal{P}(\mathbb{G}^\uparrow, \mathbb{G}^\downarrow; \boldsymbol{\eta}^\uparrow, \boldsymbol{\xi}^\downarrow) \\
 & \leq \inf_{\mathbf{x} \in X(\mathbb{F}^\uparrow)} \sup_{\mathbf{y} \in Y(\mathbb{G}^\downarrow)} E \left(\sum_{t=1}^T c_t(\mathbf{x}^t, \boldsymbol{\eta}_t^\uparrow) + \mathbf{y}_t \cdot f_t(\mathbf{x}^t, \boldsymbol{\xi}_t^\downarrow) \right) \\
 & = \inf_{\mathbf{x} \in X(\mathbb{F}^\uparrow)} \sup_{\mathbf{y} \in Y(\mathbb{G}^\downarrow)} E \left(\sum_{t=1}^T c_t(\mathbf{x}^t, \boldsymbol{\eta}_t^\uparrow) + E(\mathbf{y}_t | \mathcal{F}^\downarrow) \cdot f_t(\mathbf{x}^t, \boldsymbol{\xi}_t^\downarrow) \right) \\
 & \leq \inf_{\mathbf{x} \in X(\mathbb{F}^\uparrow)} \sup_{\mathbf{y} \in Y(\mathbb{F}^\downarrow)} E \left(\sum_{t=1}^T c_t(\mathbf{x}^t, \boldsymbol{\eta}_t^\uparrow) + \mathbf{y}_t \cdot f_t(\mathbf{x}^t, \boldsymbol{\xi}_t^\downarrow) \right).
 \end{aligned}$$

In the last line we indirectly use Proposition 2 (ii), which guarantees that for each $\mathbf{y} \in Y(\mathbb{G}^\downarrow)$ there exists a version of $E(\mathbf{y} | \mathcal{F}^\downarrow)$ which is in $Y(\mathbb{F}^\downarrow)$. To prove the converse inequality in (5.11), we first restrict the dual feasible set, then use the conditional Jensen inequality, and finally relax the primal feasible set:

$$\begin{aligned}
 & \inf \mathcal{P}(\mathbb{G}^\uparrow, \mathbb{G}^\downarrow; \boldsymbol{\eta}^\uparrow, \boldsymbol{\xi}^\downarrow) \\
 & \geq \inf_{\mathbf{x} \in X(\mathbb{G}^\uparrow)} \sup_{\mathbf{y} \in Y(\mathbb{F}^\downarrow)} E \left(\sum_{t=1}^T c_t(\mathbf{x}^t, \boldsymbol{\eta}_t^\uparrow) + \mathbf{y}_t \cdot f_t(\mathbf{x}^t, \boldsymbol{\xi}_t^\downarrow) \right) \\
 & \geq \inf_{\mathbf{x} \in X(\mathbb{G}^\uparrow)} \sup_{\mathbf{y} \in Y(\mathbb{F}^\downarrow)} E \left(\sum_{t=1}^T c_t(E(\mathbf{x}^t | \mathcal{F}^\uparrow), \boldsymbol{\eta}_t^\uparrow) + \mathbf{y}_t \cdot f_t(E(\mathbf{x}^t | \mathcal{F}^\uparrow), \boldsymbol{\xi}_t^\downarrow) \right) \\
 & \geq \inf_{\mathbf{x} \in X(\mathbb{F}^\uparrow)} \sup_{\mathbf{y} \in Y(\mathbb{F}^\downarrow)} E \left(\sum_{t=1}^T c_t(\mathbf{x}^t, \boldsymbol{\eta}_t^\uparrow) + \mathbf{y}_t \cdot f_t(\mathbf{x}^t, \boldsymbol{\xi}_t^\downarrow) \right).
 \end{aligned}$$

Observe that the last inequality is based on Proposition 2 (iii), i.e., for all primal decision vectors $\mathbf{x} \in X(\mathbb{G}^\uparrow)$ there is a version of the conditional expectation such that $E(\mathbf{x} | \mathcal{F}^\uparrow) \in X(\mathbb{F}^\uparrow)$. Combining the above results proves (5.11), and thus the claim follows. \square

Notice that the approximate problem $\mathcal{P}(\mathbb{F}^\downarrow, \mathbb{F}^\uparrow; \boldsymbol{\eta}^\downarrow, \boldsymbol{\xi}^\uparrow)$, which provides an upper bound on the original unaggregated stochastic program, is built on the probability space induced by the random vectors $\{\boldsymbol{\zeta}_t | 1 \leq \downarrow(t) = t = \uparrow(t) \leq T\}$. Depending on the specific design of the aggregation operators, the approximate problem may thus have considerably lower dimension than the original problem \mathcal{P} , which is built on the probability space induced by all random vectors $\{\boldsymbol{\zeta}_t\}_{t=1}^T$. Using similar arguments, we may claim that $\mathcal{P}(\mathbb{F}^\uparrow, \mathbb{F}^\downarrow; \boldsymbol{\eta}^\uparrow, \boldsymbol{\xi}^\downarrow)$, which provides a lower bound on the original stochastic program, has typically much lower dimension than \mathcal{P} . As a consequence, the aggregated problems might occasionally allow for numerical solution even in cases where the original problem is computationally untractable.

5.2 Linear Markov processes

The bounds provided by Theorems 3 and 4 critically rely on the martingale property of the data process. Although martingales enjoy wide popularity both in discrete and continuous time finance, it is desirable to extend the results of the previous section to more general stochastic models. Here, we will study the class of linear Markov processes. By definition, a linear Markov process ζ is a Markov process satisfying the conditions

$$\left. \begin{aligned} E(\eta_t | \mathcal{F}^s) &= H_{t,s}^o(\eta_s) \\ E(\xi_t | \mathcal{F}^s) &= H_{t,s}^c(\xi_s) \end{aligned} \right\} \quad P\text{-a.s. for } 1 \leq s \leq t \leq T. \tag{5.12}$$

The mappings $H_{t,s}^o$ and $H_{t,s}^c$ are linear affine and invertible. The corresponding inverse mappings will be denoted by $H_{s,t}^o$ and $H_{s,t}^c$, respectively. Notice that the linear Markov processes cover the class of Markov martingales considered in the previous section. Thus, we effectively work in a more general setting, here. In order to keep notation simple, we introduce the combined mappings

$$H^{i,\alpha} := H_{1,\alpha(1)}^i \times \dots \times H_{T,\alpha(T)}^i \quad (i, \alpha) \in \{o, c\} \times \{\uparrow, \downarrow\},$$

which depend on the aggregation operators. The next result generalizes Theorems 3 and 4 by allowing for data processes which need not be martingales.

Theorem 5 *Assume the conditions (C1), (C2), and (C3) hold. If the data process ζ represents a linear Markov process, then*

$$\begin{aligned} \inf \mathcal{P}(\mathbb{F}^\downarrow, \mathbb{F}^\uparrow; H^{o,\downarrow}(\eta^\downarrow), H^{c,\uparrow}(\xi^\uparrow)) &\geq \inf \mathcal{P}, \\ \inf \mathcal{P}(\mathbb{F}^\uparrow, \mathbb{F}^\downarrow; H^{o,\uparrow}(\eta^\uparrow), H^{c,\downarrow}(\xi^\downarrow)) &\leq \inf \mathcal{P}. \end{aligned}$$

Proof The statements are proved exactly as in Theorems 3 and 4. However, the relations (5.12) are used instead of the martingale property. □

It is worthwhile to remark that the mappings $H^{o,\uparrow}$ and $H^{c,\uparrow}$ will diverge with respect to the matrix 2-norm, say, if $H^{o,\downarrow}$ and $H^{c,\downarrow}$ become singular. This can happen if correlation between the random vectors ζ_t is low across neighboring stages. In any such situation, stage-aggregation may not be justifiable, and the bounds proposed in Theorem 5 may become very coarse. In contrast, the bounds are expected to be tight in case the ζ_t 's are strongly correlated across several stages.

Furthermore, it should be mentioned that the class of linear Markov processes is general enough to cover many stochastic processes of practical interest. In fact, recall that any autoregressive process can be represented as a (higher-dimensional) linear Markov process, as is shown e.g. in (33, Sect. 2).

6 Joint aggregation and discretization

Stage-aggregation may considerably reduce the dimensionality of some given stochastic optimization problem. However, numerical solution still requires discretization of the (reduced) probability space. Any attempt to aggregate certain decision stages should thus be complemented by a suitable space-discretization scheme as proposed in Sect. 3.

In order to formally elaborate the combination of aggregation and discretization schemes, we have to introduce some additional notation. First, assume that suitable aggregation operators \uparrow and \downarrow have been selected, and consider the associated stochastic processes ζ^\uparrow and ζ^\downarrow defined in Sect. 5. Then, let $\zeta^{\downarrow,u} = (\eta^{\downarrow,u}, \xi^{\downarrow,u})$ be a discrete process approximating ζ^\downarrow in the sense of Sect. 3, that is, assume $\zeta^{\downarrow,u}$ and ζ^\downarrow to satisfy the conditions (3.1). As demonstrated earlier, $\zeta^{\downarrow,u}$ can systematically be constructed if, for instance, ζ^\downarrow has a compact state space and is representable as some linear affine transformation of a serially independent noise process; see Sect. 4. We denote by $\mathbb{F}^{\downarrow,u} = \{\mathcal{F}^{\downarrow,u,t}\}_{t=1}^T$ the filtration induced by $\zeta^{\downarrow,u}$, and set $\mathcal{F}^{\downarrow,u} := \mathcal{F}^{\downarrow,u,T}$. Furthermore, we introduce a compatible discrete stochastic process $\zeta^{\uparrow,u} = (\eta^{\uparrow,u}, \xi^{\uparrow,u})$ approximating ζ^\uparrow . The idempotence property of the aggregation operators implies

$$\zeta_t^\uparrow = \zeta_{\uparrow(t)} = \zeta_{\downarrow\circ\uparrow(t)} = \zeta_{\uparrow(t)}^\downarrow, \quad t = 1, \dots, T.$$

Thus, it is natural to define $\zeta^{\uparrow,u}$ through $\zeta_t^{\uparrow,u} := \zeta_{\uparrow(t)}^{\downarrow,u}$ for $t = 1, \dots, T$. Using this definition, it is easily seen that ζ^\uparrow and its discrete approximation $\zeta^{\uparrow,u}$ also satisfy the conditions (3.1). As usual, we denote by $\mathbb{F}^{\uparrow,u} = \{\mathcal{F}^{\uparrow,u,t}\}_{t=1}^T$ the filtration induced by $\zeta^{\uparrow,u}$, and set $\mathcal{F}^{\uparrow,u} := \mathcal{F}^{\uparrow,u,T}$.

For the further argumentation we need another discrete stochastic process $\zeta^{\uparrow,l} = (\eta^{\uparrow,l}, \xi^{\uparrow,l})$ which approximates ζ^\uparrow in the sense that $\zeta^{\uparrow,l}$ and ζ^\uparrow satisfy the conditions (3.2). Moreover, we define a compatible process $\zeta^{\downarrow,l} = (\eta^{\downarrow,l}, \xi^{\downarrow,l})$ via $\zeta_t^{\downarrow,l} := \zeta_{\downarrow(t)}^{\uparrow,l}$ for $t = 1, \dots, T$. It can be verified that $\zeta^{\downarrow,l}$ and ζ^\downarrow also satisfy the conditions (3.2). Finally, let $\mathbb{F}^{\uparrow,l} = \{\mathcal{F}^{\uparrow,l,t}\}_{t=1}^T$ and $\mathbb{F}^{\downarrow,l} = \{\mathcal{F}^{\downarrow,l,t}\}_{t=1}^T$ stand for the filtrations induced by $\zeta^{\uparrow,l}$ and $\zeta^{\downarrow,l}$, respectively, and set $\mathcal{F}^{\downarrow,l} := \mathcal{F}^{\downarrow,l,T}$ and $\mathcal{F}^{\uparrow,l} := \mathcal{F}^{\uparrow,l,T}$. Now we are prepared to state our main theorem.

Theorem 6 *Assume the conditions (C1), (C2), and (C3) hold. If the data process ζ represents a linear Markov process, and the processes $\zeta^{\uparrow,u}$, $\zeta^{\uparrow,l}$, $\zeta^{\downarrow,u}$, and $\zeta^{\downarrow,l}$ have the postulated properties, then*

$$\begin{aligned} \inf \mathcal{P}(\mathbb{F}^{\downarrow,u}, \mathbb{F}^{\uparrow,u}; H^{0,\downarrow}(\eta^{\downarrow,u}), H^{c,\uparrow}(\xi^{\uparrow,u})) &\geq \inf \mathcal{P}, \\ \inf \mathcal{P}(\mathbb{F}^{\uparrow,l}, \mathbb{F}^{\downarrow,l}; H^{0,\uparrow}(\eta^{\uparrow,l}), H^{c,\downarrow}(\xi^{\downarrow,l})) &\leq \inf \mathcal{P}. \end{aligned}$$

Proof In order to keep notation simple, we prove the theorem under the stronger assumption that ζ is a Markov martingale. Generalization to the case of linear Markov processes is straightforward. First, by the Theorems 3 and 4 we have

$$\inf \mathcal{P}(\mathbb{F}^\downarrow, \mathbb{F}^\uparrow; \boldsymbol{\eta}^\downarrow, \boldsymbol{\xi}^\uparrow) \geq \inf \mathcal{P} \geq \inf \mathcal{P}(\mathbb{F}^\uparrow, \mathbb{F}^\downarrow; \boldsymbol{\eta}^\uparrow, \boldsymbol{\xi}^\downarrow).$$

Thus, the claim follows if we can show that

$$\begin{aligned} \inf \mathcal{P}(\mathbb{F}^{\downarrow,u}, \mathbb{F}^{\uparrow,u}; \boldsymbol{\eta}^{\downarrow,u}, \boldsymbol{\xi}^{\uparrow,u}) &\geq \inf \mathcal{P}(\mathbb{F}^\downarrow, \mathbb{F}^\uparrow; \boldsymbol{\eta}^\downarrow, \boldsymbol{\xi}^\uparrow), \\ \inf \mathcal{P}(\mathbb{F}^{\uparrow,l}, \mathbb{F}^{\downarrow,l}; \boldsymbol{\eta}^{\uparrow,l}, \boldsymbol{\xi}^{\downarrow,l}) &\leq \inf \mathcal{P}(\mathbb{F}^\uparrow, \mathbb{F}^\downarrow; \boldsymbol{\eta}^\uparrow, \boldsymbol{\xi}^\downarrow). \end{aligned}$$

The proofs of these inequalities follow exactly the same pattern as the proofs of Theorems 1 and 2, relying only on the conditions (3.1) and (3.2) as well as the convexity properties of the Lagrangian. There is one subtle point, however, which deserves special mention. When taking conditional expectations, it should be noticed that the σ -algebras \mathcal{F}^\uparrow and \mathcal{F}^\downarrow are equal. This follows from the fact that both aggregation operators have the same range. Similarly, we have $\mathcal{F}^{\uparrow,u} = \mathcal{F}^{\downarrow,u}$ and $\mathcal{F}^{\uparrow,l} = \mathcal{F}^{\downarrow,l}$. \square

Recall that the cardinality of the range of the aggregation operators determines the number of (effective) decision stages in the approximate models. This number can be much smaller than T . The random vectors which are not eliminated in the aggregation step are further approximated in the discretization step. Thus, the approximate stochastic programs suggested in Theorem 6 are, in principle, numerically tractable. The computational costs for their solution can be significantly lower than the cost for solving the discretized problems of Sect. 3 (without stage-aggregation).

Example 2 We illustrate the proposed approximation scheme with the following three-stage stochastic program.

$$\begin{aligned} &\underset{\mathbf{x} \in X(\mathbb{F})}{\text{minimize}} && E(-\eta_1 \mathbf{x}_1 + \eta_2 \mathbf{x}_2 - \eta_3 \mathbf{x}_3), \\ &\text{s.t.} && \left. \begin{aligned} \mathbf{x}_1 &\leq \boldsymbol{\xi}_1 \\ \mathbf{x}_1 + \mathbf{x}_2 &\leq \boldsymbol{\xi}_2 \\ -\mathbf{x}_2 + \mathbf{x}_3 &\leq \boldsymbol{\xi}_3 \end{aligned} \right\} \quad P\text{-a.s.} \end{aligned} \tag{P}$$

In this example, the state space X of the primal decision process is set to the three-dimensional unit cube $[0, 1]^3$, and the data processes $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are given by real-valued Markov martingales with independent identically distributed increments. Concretely speaking, we set

$$\begin{aligned} \eta_1 &= 1, \eta_2 = \eta_1 + \hat{\eta}_2, \eta_3 = \eta_2 + \hat{\eta}_3, \\ \xi_1 &= 1, \xi_2 = \xi_1 + \hat{\xi}_2, \xi_3 = \xi_2 + \hat{\xi}_3, \end{aligned}$$

where the independent random variables $\hat{\eta}_2, \hat{\eta}_3, \hat{\xi}_2,$ and $\hat{\xi}_3$ are uniformly distributed on $[-\frac{1}{4}, \frac{1}{4}]$. It is easily seen that \mathcal{P} satisfies (C1), (C2), and (C3).

First, we discuss stage-aggregation. To this end, we select a pair of aggregation operators satisfying the axioms of monotonicity, idempotence, and ordering. Our choice is $\uparrow(1) = 1, \uparrow(2) = 3, \uparrow(3) = 3$ and $\downarrow(1) = 1, \downarrow(2) = 1, \downarrow(3) = 3$.

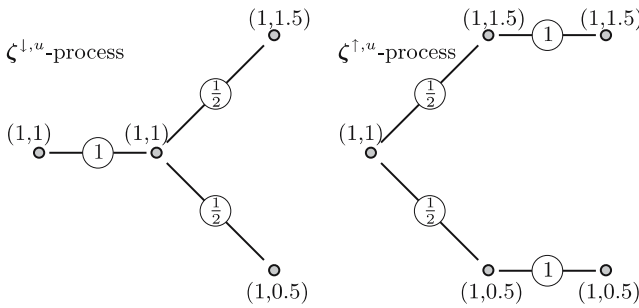


Fig. 2 Tree representation of the discretized processes $\zeta^{\downarrow,u}$ and $\zeta^{\uparrow,u}$: the state of the processes is indicated next to each node, and the transition probabilities are indicated by circled numbers assigned to the arcs

This results in approximating the original three-stage problem by effective two-stage problems. The upper aggregated problem is given by $\mathcal{P}(\mathbb{F}^\downarrow, \mathbb{F}^\uparrow; \eta^\downarrow, \xi^\uparrow)$. Note that $\eta_1^\downarrow = \eta_2^\downarrow = \xi_1^\uparrow = 1$ are deterministic while η_3^\downarrow and $\xi_2^\uparrow = \xi_3^\uparrow$ follow independent identical triangular distributions with lower limit 0.5, mode 1, and upper limit 1.5. Similarly, we introduce the lower aggregated problem $\mathcal{P}(\mathbb{F}^\uparrow, \mathbb{F}^\downarrow; \eta^\uparrow, \xi^\downarrow)$. Here, $\xi_1^\downarrow = \xi_2^\downarrow = \eta_1^\uparrow = 1$ are deterministic, while ξ_3^\downarrow and $\eta_2^\uparrow = \eta_3^\uparrow$ follow independent identical triangular distributions with lower limit 0.5, mode 1, and upper limit 1.5.

Next, we address discretization. Notice that discretizing the aggregated processes ζ^\uparrow and ζ^\downarrow reduces to discretizing the two-dimensional random vector $\zeta_3 = (\eta_3, \xi_3)$, which coincides with $\zeta_2^\uparrow = \zeta_3^\uparrow = \zeta_3^\downarrow$. All other components of the aggregated processes are deterministic. Observe that the support of ζ_3 is given by $Z_3 = \Theta_3 \times \Xi_3 = [0.5, 1.5]^2$, which represents a product of compact polytopes. Thus, we may proceed as in Sect. 4.1 to construct ζ_3^u . By (4.3) the conditional distribution of ζ_3^u given $\zeta_3 = \zeta_3$ reads

$$P_{\zeta_3^u | \zeta_3}(B | \zeta_3) = (\xi_3 - 0.5) \delta_{(1,1.5)}(B) + (1.5 - \xi_3) \delta_{(1,0.5)}(B) \quad \forall B \in \mathcal{B}(Z_3).$$

It should be remarked that the involved weighting factors $P_{\Xi_3}(e | \zeta_3)$ for $e \in \text{ext } \Xi_3$ are uniquely determined since the interval Ξ_3 represents a nondegenerate simplex on the real line. The marginal distribution of ζ_3^u then becomes

$$P_{\zeta_3^u}(B) = 0.5 \delta_{(1,1.5)}(B) + 0.5 \delta_{(1,0.5)}(B) \quad \forall B \in \mathcal{B}(Z_3).$$

According to the general theory, we set $\zeta^{\uparrow,u} := (\zeta_1, \zeta_3^u, \zeta_3^u)$ and $\zeta^{\downarrow,u} := (\zeta_1, \zeta_1, \zeta_3^u)$. These discrete stochastic processes are visualized in Fig. 2. By interchanging the roles of η and ξ in the above derivation, we can furthermore construct the two processes $\zeta^{\uparrow,l} := (\zeta_1, \zeta_3^l, \zeta_3^l)$ and $\zeta^{\downarrow,l} := (\zeta_1, \zeta_1, \zeta_3^l)$, where the conditional and marginal distributions of ζ_3^l are given by

Table 1 Numerical results

Problem	Minimum
$\mathcal{P}(\mathbb{F}^{\downarrow,u}, \mathbb{F}^{\uparrow,u}; \eta^{\downarrow,u}, \xi^{\uparrow,u})$	-1.250
$\mathcal{P}(\mathbb{F}^{\downarrow}, \mathbb{F}^{\uparrow}; \eta^{\downarrow}, \xi^{\uparrow})$	-1.417
\mathcal{P}	-1.667
$\mathcal{P}(\mathbb{F}^{\uparrow}, \mathbb{F}^{\downarrow}; \eta^{\uparrow}, \xi^{\downarrow})$	-1.917
$\mathcal{P}(\mathbb{F}^{\uparrow,l}, \mathbb{F}^{\downarrow,l}; \eta^{\uparrow,l}, \xi^{\downarrow,l})$	-2.000

$$P_{\zeta_3^l | \zeta_3}(B | \zeta_3) = (\eta_3 - 0.5) \delta_{(1.5,1)}(B) + (1.5 - \eta_3) \delta_{(0.5,1)}(B) \quad \text{and}$$

$$P_{\zeta_3^l}(B) = 0.5 \delta_{(1.5,1)}(B) + 0.5 \delta_{(0.5,1)}(B) \quad \forall B \in \mathcal{B}(Z_3).$$

After having found two approximate scenario tree representations for the data process, the upper aggregated and discretized problem $\mathcal{P}(\mathbb{F}^{\downarrow,u}, \mathbb{F}^{\uparrow,u}; \eta^{\downarrow,u}, \xi^{\uparrow,u})$ is constructed in the usual way. The number of nontrivial decision variables in this stochastic program equals the number of nodes of the $\zeta^{\downarrow,u}$ -tree (that is 4), whereas the number of nontrivial constraints equals the number of nodes of the $\zeta^{\uparrow,u}$ -tree (that is 5). Next, we construct the lower aggregated and discretized problem $\mathcal{P}(\mathbb{F}^{\uparrow,l}, \mathbb{F}^{\downarrow,l}; \eta^{\uparrow,l}, \xi^{\downarrow,l})$, where the number of nontrivial decision variables corresponds to the number of nodes in the $\zeta^{\uparrow,l}$ -tree (that is 5), and the number of nontrivial constraints corresponds to the number of nodes in the $\zeta^{\downarrow,l}$ -tree (that is 4).⁹ Both discretized problems are computationally tractable. Furthermore, the original and the aggregated (undiscretized) problems are analytically solvable since the chosen example is particularly simple. Table 1 provides a survey of the minimal values associated with the five stochastic programs under consideration. The order of these values is consistent with the general theory presented above.

7 Conclusions and outlook

This article addresses the approximation of convex multistage stochastic programs via aggregation of decision stages and discretization of the underlying probability space. Put differently, the temporal and spacial granularity of some given stochastic program is coarsened in a systematic way. In doing so, *deterministic* bounds on the optimal objective value are constructed (as opposed to a *probabilistic* confidence interval). By adapting the reasoning in (31, Sect. 4.6), the proposed bounds could principally be used to construct deterministic bounding sets for the optimal first stage decisions.

We interpret stochastic programs as abstract optimization problems over infinite-dimensional Lebesgue spaces. These problems are conveniently analyzed in a Lagrangian framework where the underlying data and information

⁹ Observe that the $\zeta^{\uparrow,l}$ -tree is isomorphic to the $\zeta^{\uparrow,u}$ -tree, and the $\zeta^{\downarrow,l}$ -tree is isomorphic to the $\zeta^{\downarrow,u}$ -tree.

processes (i.e., the filtrations governing the primal and dual strategies) represent exogenous parameters. Computationally accessible bounds on the optimal objective value are obtained by slightly perturbing these parameters. In doing so, we employ the conditional Jensen inequality together with some suitable restrictions or relaxations of the primal and dual feasible sets. Notice that the bounds based on space-discretization (cf. Sects. 3 and 4) can also be derived via Frauendorfer's barycentric approximation scheme (19,20). Our approach, however, is inspired by Theorem 2 in (4, Sect. 11.1) and avoids the use of dynamic programming techniques.

The idea of simplifying the primal and dual information processes of some given MSP goes back to Wright (44). We extend this idea by jointly controlling the information *and* the data processes. Notice that Wright concentrates on *coarsening* of the information processes only, which means that the involved filtrations are replaced by suitable *subfiltrations*. Conversely, the stage-aggregation scheme presented here (see Sect. 5) also involves *refining*. In fact, certain filtrations are replaced by suitable *superfiltrations*. In a Markovian framework, the refined information processes can later be 're-coarsened' without affecting the optimal objective value. Our approach to stage-aggregation basically plays with the timing of data revelation. If the observation of new data is delayed to some extent, we end up with an upper bound on the true objective value. Conversely, if future observations are foreseen some time ahead, we obtain a lower bound.

Joint stage-aggregation and discretization may significantly reduce the computational complexity of some given MSP. The resulting approximate problems will generally exhibit few (effective) decision stages and a finite number of scenarios. However, in case of extensive aggregation, the number of decision variables per (effective) stage may become very large. Consequently, one might eventually be forced to reduce the number of decision variables by using a suitable heuristics and, of course, without sacrificing too much accuracy.

Future research will have to address the question of how (and under what additional assumptions) the proposed bounds can be made tighter. In order to study convergence of the bounds based on discretization, one should consider two sequences of discrete stochastic processes $\{\zeta_J^u\}_{J \in \mathbb{N}}$ and $\{\zeta_J^l\}_{J \in \mathbb{N}}$ such that ζ and ζ_J^u satisfy the conditions (3.1) while ζ and ζ_J^l satisfy the conditions (3.2) for all $J \in \mathbb{N}$. If both sequences converge to ζ with respect to the \mathcal{L}^∞ -norm, and if problem \mathcal{P} is strictly feasible, then the optimal values of the approximate problems associated with ζ_J^u and ζ_J^l can be shown to converge to the optimal value of the original problem. Details will be provided in forthcoming publications.

As for the bounds based on stage-aggregation, it is useful to consider two sequences of aggregation operators $\{\uparrow_J\}_{J \in \mathbb{N}}$ and $\{\downarrow_J\}_{J \in \mathbb{N}}$ such that each pair $(\uparrow_J, \downarrow_J)$ satisfies the axioms of monotonicity, idempotence, and ordering. If these sequences are monotonic, that is, if $\uparrow_J \geq \uparrow_{J+1}$ and $\downarrow_J \leq \downarrow_{J+1}$ for all $J \in \mathbb{N}$, then the optimal values of the upper (lower) aggregated problems corresponding to the pair $(\uparrow_J, \downarrow_J)$ are monotonically decreasing (increasing) in J . If both sequences of aggregation operators converge to the identity opera-

tor,¹⁰ we have $\uparrow_J = \downarrow_J = 1$ for all J large enough. Then, the optimal values of the stage-aggregated problems trivially converge to the optimal value of the original problem. This theoretical result is of little practical relevance unless the gap between the bounds can be shown to be small for pairs of aggregation operators with considerably less than T fixed points.

A first numerical convergence analysis is provided in (22) for a real-life decision problem with more than 2,000 stages and with random parameters appearing only in the objective. It could be demonstrated that the lower bound based on stage aggregation (in a maximization framework) saturates at around ten effective stages. Additional computational experiments, however, will be necessary to assess the practical value of the proposed aggregation and discretization scheme.

Acknowledgments The author thanks the Swiss National Science Foundation for financial support.

Appendix: Conditional independence

Let V be a finite set of random vectors on the probability space (Ω, Σ, P) . By assumption, all random vectors considered in this appendix are valued in convex closed subsets of some finite-dimensional Euclidean spaces. For $A \subset V$ we denote by $\sigma(A)$ the sub- σ -algebra of Σ induced by the random vectors contained in A . In the following, A, B, C , and D stand for arbitrary subsets of V .

Definition 3 (Conditional independence) *We say that the sets of random vectors A and B are conditionally independent given C if*

$$E(\alpha \mid \sigma(B \cup C)) = E(\alpha \mid \sigma(C)) \quad P\text{-a.s.} \quad \forall \alpha \in \mathcal{L}^\infty(\Omega, \sigma(A), P; \mathbb{R}).$$

We will use the shorthand notation $A \perp\!\!\!\perp B \mid C$ for referring to conditional independence of A and B given C . If $C = \emptyset$, we say that A and B are unconditionally independent, and we will write $A \perp\!\!\!\perp B$.

Note that several equivalent definitions of conditional independence are available, see e.g. (7, Sect. 7.3). As shown in (9, Sects. 5 and 6), the conditional independence relation has the following basic properties:

- (i) Symmetry: $A \perp\!\!\!\perp B \mid C \Rightarrow B \perp\!\!\!\perp A \mid C$;
- (ii) Decomposition: $A \perp\!\!\!\perp B \cup D \mid C \Rightarrow A \perp\!\!\!\perp B \mid C$;
- (iii) Weak Union: $A \perp\!\!\!\perp B \cup D \mid C \Rightarrow A \perp\!\!\!\perp B \mid C \cup D$;
- (iv) Contraction: $A \perp\!\!\!\perp B \mid C$ and $A \perp\!\!\!\perp D \mid B \cup C \Rightarrow A \perp\!\!\!\perp B \cup D \mid C$.

It can easily be checked from the definitions that the conditional independence relation also exhibits the *trivial independence* property $A \perp\!\!\!\perp B \mid B$. In the sequel

¹⁰ Here, convergence of aggregation operators is with respect to the discrete topology on the index set of decision stages.

we describe a method to detect unobvious conditional independence relationships among certain subsets of V . In other words, given arbitrary disjoint subsets A, B , and C of V we would like to establish an easily checkable criterion to decide whether A is independent of B conditional on C . We will present a graph-theoretic approach that has been developed in the context of artificial intelligence research; see e.g. (34) and the references therein. To this end, assume that the set of random vectors $V = \{\zeta_t\}_{t=1}^T$ is totally ordered.¹¹ Let V_t be the set of the first $t - 1$ elements¹² with respect to the given ordering, and let B_t be a subset of V_t satisfying the *a priori* conditional independence relationship $\{\zeta_t\} \perp\!\!\!\perp V_t \setminus B_t \mid B_t$. This means that the conditional distribution of ζ_t given $\sigma(V_t)$ almost surely coincides with the conditional distribution of ζ_t given $\sigma(B_t)$. We will now construct a directed acyclic graph $G = (V, E)$ which reflects some conditional independence relationships of the underlying probabilistic model (including at least all *a priori* relationships). To this end, interpret the random vectors in V as the nodes or vertices of G . Moreover, create the directed edges or arcs of G by designating each B_t as the set of parents of the vertex ζ_t , i.e., from each vertex in B_t draw an arc terminating in ζ_t . The set of edges E constructed in this manner is considered as a subset of $V \times V$.

$$E = \bigcup_{t=1}^T \{(\zeta_s, \zeta_t) \mid \zeta_s \in B_t\}$$

A sequence of arcs such that every arc has exactly one vertex in common with the previous arc is called a path. A node along a path is head-to-head if the node before it and after it along the path both point to it in the graph. A directed path is a path in which the terminal node of each arc is identical to the initial node of the next arc. A node is a descendant of another if there is a directed path from the latter to the former.

Definition 4 (*d*-separation, Pearl (34, Sect. 3.3.1)) *Let A, B , and C be three disjoint sets of nodes in the graph G . We say that C *d*-separates¹³ A and B if along every path between a node in A and a node in B there is a node ζ_t satisfying one of the following conditions: (i) ζ_t is a head-to-head node along the path and neither ζ_t , nor any of its descendants are in C , or (ii) ζ_t is not head-to-head but is in C .*

Theorem 7 (42, Theorem 2) *Let A, B , and C be disjoint sets of nodes in the graph G . If C *d*-separates A and B , then $A \perp\!\!\!\perp B \mid C$.*

The converse statement need not be true, i.e., if A and B are conditionally independent given C , then C can fail to *d*-separate the sets A and B ; see the

¹¹ Notice that there are $T!$ different orderings on V . In principle, we are free to choose any ordering, but frequently there is a natural choice.

¹² Assume that V_1 is the empty set.

¹³ *d*-separation stands for direction-dependent separation.

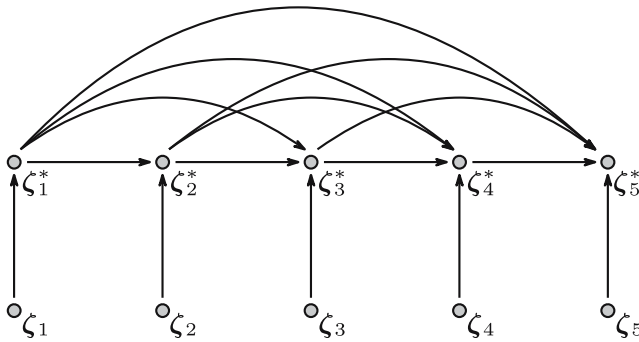


Fig. 3 Directed acyclic graph $G' = (V', E')$

discussion at the end of (34, Sect. 3.3.1). Note that the proof of Theorem 7 exclusively relies on the four basic properties of the conditional independence relation.

Proposition 3 Assume that $V = \{\xi_t\}_{t=1}^T$ and $V^* = \{\xi_t^*\}_{t=1}^T$ are ordered sets of random vectors with the same number of elements. Moreover, let V_t and V_t^* be the sets of the first $t - 1$ elements with respect to the given orderings, respectively. Assume that the random vectors in V are mutually independent and that $\{\xi_t^*\} \perp\!\!\!\perp V \setminus \{\xi_t\} \mid V_t^* \cup \{\xi_t\}$ for $t = 1, \dots, T$. Then, we find

$$V_t \perp\!\!\!\perp V^* \setminus V_t^* \mid V_t^* \quad \text{and} \quad V_t^* \perp\!\!\!\perp V \setminus V_t \mid V_t, \quad t = 1, \dots, T.$$

Proof Construct an ordering on $V' := V \cup V^*$ by first enumerating the elements of V and then those of V^* using the respective individual orderings. Denote the elements of V' by ξ'_t where the index t ranges from 1 to $2T$ and

$$\xi'_t := \begin{cases} \xi_t & \text{for } t = 1, \dots, T, \\ \xi_{t-T}^* & \text{for } t = T + 1, \dots, 2T. \end{cases}$$

As usual, Let V'_t be the set of the first $t - 1$ elements of V' . Furthermore, set

$$B'_t := \begin{cases} \emptyset & \text{for } t = 1, \dots, T, \\ V_{t-T}^* \cup \{\xi_{t-T}\} & \text{for } t = T + 1, \dots, 2T. \end{cases}$$

By construction, the postulated conditional and unconditional independence relationships translate to $\{\xi'_t\} \perp\!\!\!\perp V'_t \setminus B'_t \mid B'_t$ for $t = 1, \dots, 2T$. As before, we can now construct a directed acyclic graph $G' = (V', E')$ with vertices V' and directed edges (Fig. 3 visualizes G' for the special case $T = 5$)

$$E' = \bigcup_{t=1}^{2T} \{(\xi'_s, \xi'_t) \mid \xi'_s \in B'_t\}.$$

The first statement now follows by application of Theorem 7 if we can show that V_t^* d -separates V_t and $V^* \setminus V_t^*$ (all sets are considered as subsets of V'). In fact, every path from V_t to $V^* \setminus V_t^*$ passes by some node in V_t^* which is not head-to-head along the path. Thus, the claim follows. Conversely, the second statement holds by Theorem 7 if we can show that V_t d -separates V_t^* and $V \setminus V_t$. It turns out that every path from V_t^* to $V \setminus V_t$ passes by some node in $V^* \setminus V_t^*$ which is head-to-head along the path. Neither this specific node nor any of its descendants belong to V_t . Therefore, the second claim is established. \square

Proposition 4 *Under the assumptions of Proposition 3 we also have*

$$\{\zeta_t\} \perp\!\!\!\perp \{\zeta_\tau^*\}_{\tau=t+1}^T \mid \{\zeta_\tau^*\}_{\tau=1}^t, \quad t = 1, \dots, T.$$

Proof Bearing in mind the special topology of the graph $G' = (V', E')$ considered in the proof of Proposition 3, it is clear that every path from ζ_t to the set $\{\zeta_\tau^*\}_{\tau=t+1}^T$ passes by the node ζ_t^* , which is not head-to-head along the path. This implies d -separation of $\{\zeta_t\}$ and $\{\zeta_\tau^*\}_{\tau=t+1}^T$ given $\{\zeta_\tau^*\}_{\tau=1}^t$. The claim now follows from Theorem 7. \square

Definition 5 (Markov Process) *Let $V = \{\zeta_t\}_{t=1}^T$ be an ordered set of random vectors and let V_t be the set of its first $t - 1$ elements, respectively. V is called a Markov process if $\{\zeta_{t+1}\} \perp\!\!\!\perp V_t \mid \{\zeta_t\}$ for each $t = 2, \dots, T - 1$.*

For the further argumentation, let $\alpha : \{1, \dots, T\} \rightarrow \{1, \dots, T\}$ be a generalized aggregation operator, i.e., we require that α be monotonically increasing and idempotent ($\alpha \circ \alpha = \alpha$). For instance, one can think of α as the upper or lower aggregation operators \uparrow or \downarrow introduced in Sect. 5, or as the trivial aggregation operator 1. Then, let U_t be the set of the first t elements of V , and set $U_t^\alpha := \{\zeta_{\alpha(\tau)}\}_{\tau=1}^t$ for $t = 1, \dots, T$. Moreover, define $U := U_T$ and $U^\alpha := U_T^\alpha$.

Proposition 5 (Locality) *For each $t = \alpha(t)$ we have $U_t \perp\!\!\!\perp U^\alpha \setminus U_t^\alpha \mid U_t^\alpha$.*

Proof We represent the Markov process in the usual way as a directed acyclic graph $G = (V, E)$. By Definition 5, the sets of parents of the individual vertices can be chosen to be $B_1 := \emptyset$ and $B_t := \{\zeta_{t-1}\}$ for all $t > 1$. Next, choose an arbitrary fixed point $t = \alpha(t)$. Since every path from $U_t \setminus U_t^\alpha$ to $U^\alpha \setminus U_t^\alpha$ passes by the node $\zeta_t \in U_t^\alpha$, which is not head-to-head along the path, we may invoke Theorem 7 to conclude that $U_t \setminus U_t^\alpha \perp\!\!\!\perp U^\alpha \setminus U_t^\alpha \mid U_t^\alpha$. The claim now follows from the trivial independence and contraction properties of the conditional independence relation. \square

References

1. Ash, R.: Real Analysis and Probability. Probability and Mathematical Statistics. Academic Berlin Heidelberg Newyork (1972)
2. Birge, J.: Aggregation in stochastic production models. In: Proceedings of the 11th IFIP Conference on System Modelling and Optimization. Springer Berlin Heidelberg Newyork, New York (1984)

3. Birge, J.: Aggregation in stochastic linear programming. *Math. Program.* **31**, 25–41 (1985)
4. Birge, J., Louveaux, F.: *Introduction to Stochastic Programming*. Springer Berlin Heidelberg New York (1997)
5. Birge, J., Wets, R.B.: Computing bounds for stochastic programming problems by means of a generalized moment problem. *Math. Oper. Res.* **12**, 149–162 (1987)
6. Casey, M., Sen, S.: The scenario generation algorithm for multistage stochastic linear programming. *Math. Oper. Res.* **30**(3), 615–631 (2005)
7. Chow, Y., Teicher, H.: *Probability Theory*, 3rd edn. Springer Berlin Heidelberg, New York (1997)
8. Dantzig, G., Infanger, G.: Large-scale stochastic linear programs—importance sampling and Benders decomposition. *Comput. Appl. Math.* **1**, 111–120 (1992)
9. Dawid, A.: Conditional independence for statistical operations. *Ann. Stat.* **8**(3), 598–617 (1980)
10. Dempster, M., Thompson, R.: EVPI-based importance sampling solution procedures for multistage stochastic linear programmes on parallel MIMD architectures. *Ann. Oper. Res.* **90**, 161–184 (1999)
11. Dupačová, J., Gröwe-Kuska, N., Römisch, W.: Scenario reduction in stochastic programming: an approach using probability metrics. *Math. Program. Ser. A* **95**, 493–511 (2003)
12. Edirisinghe, N., Ziemba, W.: Bounding the expectation of a saddle function with application to stochastic programming. *Math. Oper. Res.* **19**, 314–340 (1994)
13. Edirisinghe, N., Ziemba, W.: Bounds for two-stage stochastic programs with fixed recourse. *Math. Oper. Res.* **19**, 292–313 (1994)
14. Ermoliev, Y., Gaivoronski, A.: Stochastic quasigradient methods for optimization of discrete event systems. *Ann. Oper. Res.* **39**, 1–39 (1992)
15. Flåm, S., Wets, R.B.: Finite horizon approximates of infinite horizon stochastic programs. *Stochas. Optim.* **81**, 337–350 (1986)
16. Flåm, S., Wets, R.B.: Existence results and finite horizon approximates for infinite horizon optimization problems. *Econometrica* **55**, 1187–1209 (1987)
17. Frauendorfer, K.: Solving SLP recourse problems with arbitrary multivariate distributions – the dependent case. *Math. Oper. Res.* **13**, 377–394 (1988)
18. Frauendorfer, K.: *Stochastic two-stage programming*, Lect. Notes Econ. Math. Syst., vol. 392. Springer, Berlin Heidelberg New York (1992)
19. Frauendorfer, K.: Multistage stochastic programming: Error analysis for the convex case. *Z. Oper. Res.* **39**(1), 93–122 (1994)
20. Frauendorfer, K.: Barycentric scenario trees in convex multistage stochastic programming. *Math. Program.* **75**(2), 277–294 (1996)
21. Gassmann, H., Ziemba, W.: A tight upper bound for the expectation of a convex function of a multivariate random variable. *Math. Program. Study* **27**, 39–53 (1986)
22. Haarbrücker, G., Kuhn, D.: Valuation of electricity swing options by multistage stochastic programming. Working paper (2004)
23. Heitsch, H., Römisch, W.: Scenario reduction algorithms in stochastic programming. *Comput. Optim. Appl.* **24**, 187–206 (2003)
24. Higle, J., Sen, S.: Stochastic decomposition: An algorithm for two-stage linear programs with recourse. *Math. Oper. Res.* **16**, 650fb–669 (1991)
25. Høyland, K., Wallace, S.: Generating scenario trees for multistage decision problems. *Manage. Sci.* **47**(2), 295–307 (2001)
26. Infanger, G.: *Planning under Uncertainty: Solving Large-Scale Stochastic Linear Programs*. Boyd and Fraser, Danvers (1994)
27. Kall, P.: An upper bound for SLP using first and total second moments. *Ann. Oper. Res.* **30**, 267–276 (1991)
28. Kall, P., Wallace, S.: *Stochastic Programming*. Wiley, Chichester (1994)
29. Kaut, M., Wallace, S.: Evaluation of scenario-generation methods for stochastic programming. *The Stochastic Programming E-Print Series (SPEPS)* (2003)
30. Korf, L.: An approximation framework for infinite horizon stochastic dynamic optimization problems with discounted cost. Research report, Department of Mathematics, Washington University, Seattle, USA (2000)
31. Kuhn, D.: Generalized Bounds for Convex Multistage Stochastic Programs. *Lect. Notes Econ. Math. Syst.*, vol. 548. Springer, Berlin Heidelberg New York (2004)

32. Madansky, A.: Inequalities for stochastic linear programming problems. *Manage. Sci.* **6**, 197–204 (1960)
33. Meyn, S., Tweedie, R.: *Markov Chains and Stochastic Stability*. Springer, Berlin Heidelberg New York (1996)
34. Pearl, J.: *Probabilistic Reasoning in Intelligent Systems*, 2nd edn. Morgan Kaufman, San Mateo (1991)
35. Pflug, G.: Scenario tree generation for multiperiod financial optimization by optimal discretization. *Math. Program., Ser. B* **89**, 251–271 (2001)
36. Prékopa, A.: *Stochastic Programming*. Kluwer, Dordrecht (1995)
37. Rachev, S., Römisch, W.: Quantitative stability in stochastic programming: the method of probability metrics. *Math. Oper. Res.* **27**, 792–818 (2002)
38. Rockafellar, R., Wets, R.B.: The optimal recourse problem in discrete time: L^1 -multipliers for inequality constraints. *SIAM J. Control Optim.* **16**, 16–36 (1978)
39. Schervish, M.: *Theory of Statistics*. Springer Berlin Heidelberg New York (1995)
40. Shapiro, A.: On complexity of multistage stochastic programs. *Oper. Res. Lett.* **34**, 1–8 (2006)
41. Shapiro, A., Nemirovski, A.: On complexity of stochastic programming problems. In: Jeyakumar, V., Rubinov, A.: (eds.) *Continuous Optimization: Current Trends and Applications*, pp. 111–144. Springer Berlin Heidelberg New York (2005)
42. Verma, T., Pearl, J.: Causal networks and expressiveness. In: *Proceedings of the 4th Workshop on Uncertainty in Artificial Intelligence*, pp. 352–359. Mountain View, CA (1988)
43. Dupačová (as Žáčková), J.: On minimax solutions of stochastic linear programming problems. *Časopis pro Pěstování Matematiky* **91**, 423–429 (1966)
44. Wright, S.: Primal-dual aggregation and disaggregation for stochastic linear programs. *Math. Oper. Res.* **19**(4), 893–908 (1994)