Aggregation and Discretization in Multistage Stochastic Programming

Daniel Kuhn

daniel.kuhn@unisg.ch

Institute for Operations Research and Computational Finance

University of St. Gallen

Bodanstrasse 6, 9000 St. Gallen, Switzerland

30th June 2005

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.

Key words. stochastic programming, approximation, bounds, aggregation, discretization

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.

By convention, a sequence of consecutive decisions is referred to as a strategy. In financial applications, one usually attempts to find a strategy which satisfies certain physical or regulatory constraints and minimizes expected cost (i.e. 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, and for the sake of transparent terminology, 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, 29, 37]. 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. 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 binomial 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.[1] In these cases,

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

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 stochastic program while preserving its *temporal* structure (time periods).

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.

A survey and evaluation of popular scenario generation techniques is provided in [30]. Common sampling-based methods [4, Chap. 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. [26]) 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 importance for the present article are the so-called bounding methods [4, Chap. 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, 11, 17, 21, 28, 33]. Other variants, such as Frauendorfer's *barycentric approximation scheme*, apply if the recourse functions represent convex-concave saddle functions [12, 13, 18, 19, 20]. A generalization to specific problems with nonconvex recourse functions is discussed in [32]. Furthermore, the concept of probability metrics is at the heart of several modern scenario generation methods of increasing popularity. *Optimal discretization* [36], for instance, synthesizes a tree-structured approximation to a given stochastic process which minimizes the

Wasserstein distance (transportation metric). *Scenario reduction* [10, 24, 38], 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 in the course of the solution procedure [7, 9, 14, 25, 27].

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 [42] 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 [42] in later sections. Other aggregation schemes have been developed to deal with infinite horizon problems [15, 16, 31]. 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 up-

per and lower bounds on the optimal value of the original problem. Unlike the partially aggregated MSPs in [42], 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.

Section 2 formally introduces the class of stochastic programs under consideration, thereby assuming a Lagrangian perspective. 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 evaluates their compatibility with the discretization scheme of Sect. 3. Finally, Sect. 6 concludes, while Appendix A reviews some useful 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 $t = 1, \ldots, T$. First, we set up a probabilistic model for the underlying uncertainty. All random objects are defined on an abstract probability space (Ω, Σ, P). Adopting the standard terminology of probability theory, we will refer to Ω as the *sample space*. Furthermore, we use the following definition of a stochastic process.

Definition 2.1 (Stochastic Process)**.** *We say that $\boldsymbol{\zeta}$ is a stochastic process with state space Z if $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_T)$ and $Z = \times_{t=1}^{T} Z_t$ such that each random vector $\boldsymbol{\zeta}_t$ 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 $\boldsymbol{\zeta}^t := (\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_t)$ valued in $Z^t := \times_{\tau=1}^{t} Z_\tau$ for all $t = 1, \ldots, T$.*[2]

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

Let $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ be two stochastic processes in the sense of definition 2.1 with state spaces Θ and Ξ, respectively. 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 variables 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 in the sense of definition 2.1 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.[3] Unlike the data processes, which are given exogenously, the decision processes are a priori unspecified and will be determined endogenously in the optimization procedure.

[2]Sometimes, notation is simplified by further introducing a dummy random variable $\boldsymbol{\zeta}^0$ taking the constant value 0.

[3]Decision processes are also referred to as *strategies*, *policies*, or *decision rules*.

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 $t = 1, \ldots, T$ we define

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

By definition, $X(\mathbb{G}) := X^T(\mathbb{G})$ contains all essentially bounded $\mathbb{G}$-adapted primal strategies, whereas $Y(\mathbb{G}) := Y^T(\mathbb{G})$ contains all integrable $\mathbb{G}$-adapted dual strategies. 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 linear spaces $X(\mathbb{F})$ 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. [39].

For each stage $t = 1, \ldots, T$ there is a real-valued cost function c_t as well as a vector-valued constraint function f_t such that both

$$c_t : X^t \times \Theta_t \to \mathbb{R} \quad \text{and} \quad f_t : X^t \times \Xi_t \to \mathbb{R}^{m_t}$$

are (at least) 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

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

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 ($t = 1, \ldots, T$):

(C1) c_t is convex in x^t, concave in η_t, and subdifferentiable on $X^t \times \Theta_t$;

(C2) f_t is additively separable, $f_t = g_t + h_t$, where the mappings $g_t : X^t \to \mathbb{R}^{m_t}$ and $h_t : \Xi^t \to \mathbb{R}^{m_t}$ are componentwise convex and subdifferentiable;

(C3) X_t is a convex compact subset of $\mathbb{R}^{n_t}$, 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 [42], we can now define a family of primal and dual optimization problems, which depend parametrically on the two filtrations $\mathbb{G}$ and $\mathbb{H}$ as well as on the data processes $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$.

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

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

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

[4] However, $\mathcal{P}$ need neither 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.

By the basic regularity conditions, L is convex in (x, ξ), concave in (y, η), and subdifferentiable on its domain. The following proposition establishes a useful reformulation of the parametric stochastic program $\mathcal{P}(\mathbb{G}, \mathbb{H}; \boldsymbol{\eta}, \boldsymbol{\xi})$ in terms of the corresponding Lagrangian.

Proposition 2.2. *Under the conditions (C1), (C2), and (C3) we have*

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

Proof. Extend Wright's argument [42, Sect. 4] to the nonlinear convex case. □

3 Bounds Based on Space-Discretization

Assume that there are stochastic processes $\boldsymbol{\eta}^u$ and $\boldsymbol{\xi}^u$ in the sense of definition 2.1 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.

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

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

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

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

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.

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

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

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

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

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

Theorem 3.1. *Assume the conditions (C1), (C2), and (C3) to hold. If the data process $\boldsymbol{\zeta}$ and its approximation $\boldsymbol{\zeta}^u$ satisfy* (3.1)*, we find*

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

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

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

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 $\boldsymbol{x} \in X(\mathbb{F}^u)$. 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 $\boldsymbol{y}$ and $\boldsymbol{\eta}$ 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 $\boldsymbol{x}$ and $\boldsymbol{\xi}^u$. 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 $\boldsymbol{y} \in Y(\mathbb{F})$ to all decisions in $Y(\mathbb{F}^u)$. □

Theorem 3.2. *Assume the conditions (C1), (C2), and (C3) to hold. If the data process $\boldsymbol{\zeta}$ and its approximation $\boldsymbol{\zeta}^l$ satisfy* (3.2)*, we find*

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

Proof. The proof is analogous to that of theorem 3.1. Without a detailed description of the involved manipulations, we may therefore state the following chain of inequalities

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

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 $\boldsymbol{\zeta}^u$ 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 $\boldsymbol{\zeta}^l$ represents a finitely supported discrete process. These observations together with theorems 3.1 and 3.2 show that we can (numerically) calculate upper and lower bounds on $\inf \mathcal{P}$ whenever it is possible to find discrete processes $\boldsymbol{\zeta}^u$ and $\boldsymbol{\zeta}^l$ 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 $\boldsymbol{\zeta}$, 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 $\boldsymbol{\zeta}$ and its discrete approximation are close in a certain sense, e.g. with respect to the weak topology. In this section, we focus on the construction of a discrete process $\boldsymbol{\zeta}^u$ subject to the conditions (3.1). The construction of a discrete process $\boldsymbol{\zeta}^l$ 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 $\boldsymbol{\zeta}^u$ may depend on the properties of the underlying data process $\boldsymbol{\zeta}$. Sometimes, finding a suitable $\boldsymbol{\zeta}^u$ may even be impossible. We will show that $\boldsymbol{\zeta}^u$ can systematically be constructed whenever $\boldsymbol{\zeta}$ belongs to some class of autoregressive processes with compact state spaces. 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 4.1. *Let $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ be finite-dimensional random vectors on (Ω, Σ, P) taking values in some Borel sets A and B, respectively. The (marginal) distribution of $\boldsymbol{\beta}$ is denoted by $P_{\boldsymbol{\beta}}$, while $P_{\boldsymbol{\beta}|\boldsymbol{\alpha}}$ stands for the (regular) conditional distribution of $\boldsymbol{\beta}$ given $\boldsymbol{\alpha} = \alpha$. Thus, we have*

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

Note that the data process $\boldsymbol{\zeta}$ and its distribution $P_{\boldsymbol{\zeta}}$ are a priori known. We will construct $\boldsymbol{\zeta}^u$ by specifying the conditional distribution $P_{\boldsymbol{\zeta}^u|\boldsymbol{\zeta}}$. To be a regular conditional distribution, $P_{\boldsymbol{\zeta}^u|\boldsymbol{\zeta}}(B|\zeta)$ 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 $\boldsymbol{\zeta}$ and $\boldsymbol{\zeta}^u$ is uniquely determined by the product measure theorem [1, theorem 2.6.2], while the conditional distribution $P_{\boldsymbol{\zeta}|\boldsymbol{\zeta}^u}$ is obtainable via Bayes' theorem [40, theorem 1.31].

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

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

Thereby, $\operatorname{ext}\Xi$ represents the set of extreme points of Ξ, which 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 $\operatorname{ext}\Xi$. In addition, introduce a Borel measurable function $\zeta_\Xi = (\eta_\Xi, \xi_\Xi) : \Xi \to Z$. Set $\xi_\Xi(e) = e$ and

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

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

$$P_{\boldsymbol{\zeta}^u|\boldsymbol{\zeta}}(B|\zeta) := \sum_{e\in\operatorname{ext}\Xi} P_\Xi(e|\zeta)\, \delta_{\zeta_\Xi(e)}(B). \tag{4.1}$$

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

$$P_{\boldsymbol{\zeta}^u|\boldsymbol{\zeta}}(\cdot|\zeta) \ll \sum_{e\in\operatorname{ext}\Xi} \delta_{\zeta_\Xi(e)}(\cdot), \tag{4.2}$$

i.e. the regular conditional distribution of $\boldsymbol{\zeta}^u$ given $\boldsymbol{\zeta} = \zeta$ is absolutely continuous with respect to a discrete measure independent of the parameter ζ. The conditional density is $\zeta^u \mapsto P_\Xi(\xi^u|\zeta)$. Having constructed a candidate random vector $\boldsymbol{\zeta}^u$, 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

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

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 $P_{\boldsymbol{\zeta}^u|\boldsymbol{\zeta}}$ and the measurable function P_Ξ, respectively.

Thus, (3.1c) follows. In a next step, we will argue that

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

As before, the first equality holds by a standard result, whereas the second equality follows from the measure-theoretic version of Bayes' theorem [40, theorem 1.31], which applies due to (4.2). Finally, the third equality is immediate from the construction of $\boldsymbol{\zeta}^u$. 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 $\boldsymbol{\zeta}^u$. Generally, such $\boldsymbol{\zeta}^u$ will not be close to the data process $\boldsymbol{\zeta}$ with respect to the weak topology of distributions. In any case, the conditions (3.1c) and (3.1d) guarantee matching of the first moments, i.e.

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

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

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

but the higher order moments of $\boldsymbol{\zeta}$ and $\boldsymbol{\zeta}^u$ 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 $\boldsymbol{\zeta}^u$. Let us therefore assume that the data process is representable as

$$\boldsymbol{\zeta} = \sum_{\lambda \in \Lambda} 1_{\{\boldsymbol{\lambda}=\lambda\}} \, \boldsymbol{\zeta}_\lambda, \tag{4.3}$$

where the random variable $\boldsymbol{\lambda}$ and the random vectors $\{\boldsymbol{\zeta}_\lambda\}_{\lambda\in\Lambda}$ are mutually independent, and Λ is a finite index set. In particular, assume that $\boldsymbol{\lambda}$ takes values in Λ, while $\boldsymbol{\zeta}_\lambda$ is supported on a compact polytope $Z_\lambda \subset Z$ for all $\lambda \in \Lambda$. By (4.3)

and the independence assumption, the distribution of $\boldsymbol{\zeta}$ can be written as a convex combination of the distributions of the $\boldsymbol{\zeta}_\lambda$ with convex weights $P(\boldsymbol{\lambda} = \lambda)$. Moreover, if some process $\boldsymbol{\zeta}'$ with the same distribution as $\boldsymbol{\zeta}$ allows for a representation of the form (4.3), then, without loss of generality, we may set $\boldsymbol{\zeta} := \boldsymbol{\zeta}'$. Any such redefinition is unproblematic since only the distribution of $\boldsymbol{\zeta}$ has practical relevance for the stochastic program $\mathcal{P}$. These insights suggest that a decomposition as in (4.3) 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 $\boldsymbol{\zeta}_\lambda$ separately. Concretely speaking, we construct random vectors $\{\boldsymbol{\zeta}_\lambda^u\}_{\lambda\in\Lambda}$ with the properties

(i) $E(\boldsymbol{\xi}_\lambda^u \,|\, \boldsymbol{\zeta}_\lambda) = \boldsymbol{\xi}_\lambda \quad P$-a.s., $\lambda \in \Lambda$;

(ii) $E(\boldsymbol{\eta}_\lambda \,|\, \boldsymbol{\zeta}_\lambda^u) = \boldsymbol{\eta}_\lambda^u \quad P$-a.s., $\lambda \in \Lambda$;

(iii) $\boldsymbol{\lambda}$ and the paired random vectors $\{(\boldsymbol{\zeta}_\lambda, \boldsymbol{\zeta}_\lambda^u)\}_{\lambda\in\Lambda}$ are mutually independent.

Then, we define a candidate process

$$\boldsymbol{\zeta}^u = \sum_{\lambda\in\Lambda} 1_{\{\boldsymbol{\lambda}=\lambda\}}\, \boldsymbol{\zeta}_\lambda^u,$$

which reflects the structure of (4.3). 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

$$\sigma(\boldsymbol{\zeta}) \subset \sigma(\boldsymbol{\lambda},\, \{\boldsymbol{\zeta}_\lambda\}_{\lambda\in\Lambda}) \quad \text{and} \quad \sigma(\boldsymbol{\zeta}^u) \subset \sigma(\boldsymbol{\lambda},\, \{\boldsymbol{\zeta}_\lambda^u\}_{\lambda\in\Lambda}). \tag{4.4}$$

For instance, the first inclusion holds because $\boldsymbol{\zeta}$ is the image of $\boldsymbol{\lambda}$ and the $\boldsymbol{\zeta}_\lambda$'s under a continuous map (with respect to the Euclidean topology on Z on one

[6] Note that it is possible to represent any compactly supported distribution as a finite convex combination of distributions with arbitrarily small supports.

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

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

The first equality uses (4.4) while the second and the third equalities are based on the properties (iii) and (i) stated above. By a similar reasoning we obtain

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

As before, the first equality is due to (4.4) while the second and the third equalities follow from the properties (iii) and (ii), respectively

4.3 Independent Multistage Case

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

(i) $E(\boldsymbol{\xi}_t^u \,|\, \boldsymbol{\zeta}_t,\, \boldsymbol{\zeta}^{u,t-1}) = \boldsymbol{\xi}_t \quad P\text{-a.s.};$

(ii) $E(\boldsymbol{\eta}_t \,|\, \boldsymbol{\zeta}_t^u,\, \boldsymbol{\zeta}^{u,t-1}) = \boldsymbol{\eta}_t^u \quad P\text{-a.s.};$

(iii) $\{\boldsymbol{\zeta}_t^u\}$ is conditionally independent of $\{\boldsymbol{\zeta}_\tau\}_{\tau\neq t}$ given $\{\boldsymbol{\zeta}_t\} \cup \{\boldsymbol{\zeta}_\tau^u\}_{\tau<t}$.

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 $(B,\zeta_t)\mapsto P_t^u(B|\zeta,\zeta^{u,t-1})$ represents a conditional distribution of the form (4.1) or one of its generalizations in the spirit of Sect. 4.2. Note that the discrete conditional scenarios and probabilities corresponding to $\boldsymbol{\zeta}_t^u$ 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 $(\zeta,\zeta^{u,t-1})\mapsto P_t^u(B|\zeta,\zeta^{u,t-1})$ 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 A). These insights suggest that a discrete process $\boldsymbol{\zeta}^u$ subject to the above conditions can systematically be obtained, and there is considerable flexibility in its construction. In particular, notice that we allow the $\{\boldsymbol{\zeta}_t^u\}_{t=1}^T$ 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 $\boldsymbol{\zeta}$ and the synthesized approximation $\boldsymbol{\zeta}^u$ satisfy the requirements (3.1). In order to prove (3.1a) we choose some $\boldsymbol{x}\in X(\mathbb{F})$. By hypothesis, the support of $\boldsymbol{x}$ is covered by X, which is convex and closed. Moreover, it is known that the support of the conditional expectation $E(\boldsymbol{x}|\mathcal{F}^u)$ is a subset of the convex hull of the support of $\boldsymbol{x}$.[7] Consequently, $E(\boldsymbol{x}|\mathcal{F}^u)$ is valued in X almost surely. It remains to be shown that $E(\boldsymbol{x}_t|\mathcal{F}^u)$ is $\mathcal{F}^{u,t}$-measurable almost surely for each t. An equivalent statement is

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

This, however, is true by proposition A.2 (iii) in the appendix since the sets of random vectors $\{\boldsymbol{\zeta}_\tau\}_{\tau\leq t}$ and $\{\boldsymbol{\zeta}_\tau^u\}_{\tau>t}$ are conditionally independent given $\{\boldsymbol{\zeta}_\tau^u\}_{\tau\leq t}$;

[7] Without being rigorous, evaluating the (regular) conditional expectation of $\boldsymbol{x}$ 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 $\boldsymbol{x}$.

conditional independence is implied by proposition A.6. The proof of (3.1b) is widely parallel. First, choose $\boldsymbol{y} \in Y(\mathbb{F}^u)$. By repeating the above arguments, it can be shown that the conditional expectation $E(\boldsymbol{y}|\mathcal{F})$ takes values in Y almost surely. Furthermore, one should verify that $E(\boldsymbol{y}_t|\mathcal{F})$ is $\mathcal{F}^t$-measurable almost surely for t arbitrary. This statement can be reformulated as

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

which is an immediate consequence of proposition A.2 (iii). Thereby, we use the fact that the sets of random vectors $\{\boldsymbol{\zeta}_\tau^u\}_{\tau \le t}$ and $\{\boldsymbol{\zeta}_\tau\}_{\tau > t}$ are conditionally independent given $\{\boldsymbol{\zeta}_\tau\}_{\tau \le t}$, as implied by proposition A.6. 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.

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

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

$$E(\boldsymbol{\eta}_t \,|\, \boldsymbol{\zeta}^u) = E(\boldsymbol{\eta}_t \,|\, \boldsymbol{\zeta}^{u,t}) = \boldsymbol{\eta}_t^u \quad P\text{-a.s.} \quad \text{for} \quad t = t, \ldots, T.$$

The first equality holds by proposition A.2 (iii) since $\{\boldsymbol{\zeta}_t\}$ is conditionally independent of $\{\boldsymbol{\zeta}_\tau^u\}_{\tau > t}$ given $\{\boldsymbol{\zeta}_\tau^u\}_{\tau \le t}$; conditional independence follows from proposition A.7. The second equality is due to (ii). Thus, (3.1d) is established.

4.4 Dependent Multistage Case

Let $\hat{\boldsymbol{\eta}}$ and $\hat{\boldsymbol{\xi}}$ be two stochastic processes in the sense of definition 2.1 with state spaces $\hat{\Theta}$ and $\hat{\Xi}$, respectively. As usual, we further introduce a combined stochastic process $\hat{\boldsymbol{\zeta}} := (\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\xi}})$ with state space $\hat{Z} := \hat{\Theta} \times \hat{\Xi}$. Let us assume that the corresponding random vectors $\{\hat{\boldsymbol{\zeta}}_t\}_{t=1}^T$ are mutually independent. Consequently,

$\hat{\boldsymbol{\zeta}}$ 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

$$\boldsymbol{\zeta} = (\boldsymbol{\eta}, \boldsymbol{\xi}) = (H^{\mathrm{o}}\,\hat{\boldsymbol{\eta}}, H^{\mathrm{c}}\,\hat{\boldsymbol{\xi}}), \tag{4.5}$$

where $H^{\mathrm{o}} : \hat{\Theta} \to \Theta$ and $H^{\mathrm{c}} : \hat{\Xi} \to \Xi$ 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 $\boldsymbol{\eta}_t$ and $\hat{\boldsymbol{\eta}}_s$ are zero for $s > t$. However, they may be nonzero for $s \leq t$, which implies that the $\boldsymbol{\eta}$ process may be autocorrelated. Since $\hat{\boldsymbol{\zeta}}$ consists of mutually independent random vectors, we can proceed as in Sect. 4.3 to construct a process $\hat{\boldsymbol{\zeta}}^u$. Thus, the processes $\hat{\boldsymbol{\zeta}}$ and $\hat{\boldsymbol{\zeta}}^u$ satisfy the conditions (3.1). Next, set

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

Using the above definitions, we will prove that the processes $\boldsymbol{\zeta}$ and $\boldsymbol{\zeta}^u$ also comply with the conditions (3.1). To this end, we first notice that

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

for each t, since the transformations H^{o} and H^{c} are linear, bijective, and non-anticipative. Thus, the filtrations induced by the processes $\boldsymbol{\zeta}^u$ and $\hat{\boldsymbol{\zeta}}^u$ are identical, and the proof of (3.1a) is as in Sect. 4.3. Moreover, the filtrations induced by the processes $\boldsymbol{\zeta}$ and $\hat{\boldsymbol{\zeta}}$ 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 $\boldsymbol{\zeta}$ and $\boldsymbol{\zeta}^u$ as well as linearity of the transformations H^{o} and H^{c}, i.e.

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

Notice that the data processes of the form (4.5) 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 [32]. For instance, lognormal stochastic processes with serial correlations are investigated in [32, Sects. 6.3 and 6.4].

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. Thereby, we use similar techniques as in Sects. 3 and 4.

Let us introduce two *aggregation operators* $\uparrow$ and $\downarrow$ mapping the finite index set $\{1, \ldots, T\}$ to itself. 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 $\downarrow$ ar monotonously increasing;

(ii) idempotence: $\uparrow \circ \uparrow = \uparrow$, $\downarrow \circ \downarrow = \downarrow$, $\uparrow \circ \downarrow = \downarrow$, and $\downarrow \circ \uparrow = \uparrow$;

(iii) ordering: $\downarrow \leq 1 \leq \uparrow$.

Thereby, 1 denotes the identity mapping on $\{1, \ldots, T\}$. As follows from the defining conditions, the two aggregation operators are uniquely determined by their fixed point sets $\{t \,|\, \uparrow(t) = t\}$ and $\{t \,|\, \downarrow(t) = t\}$. 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

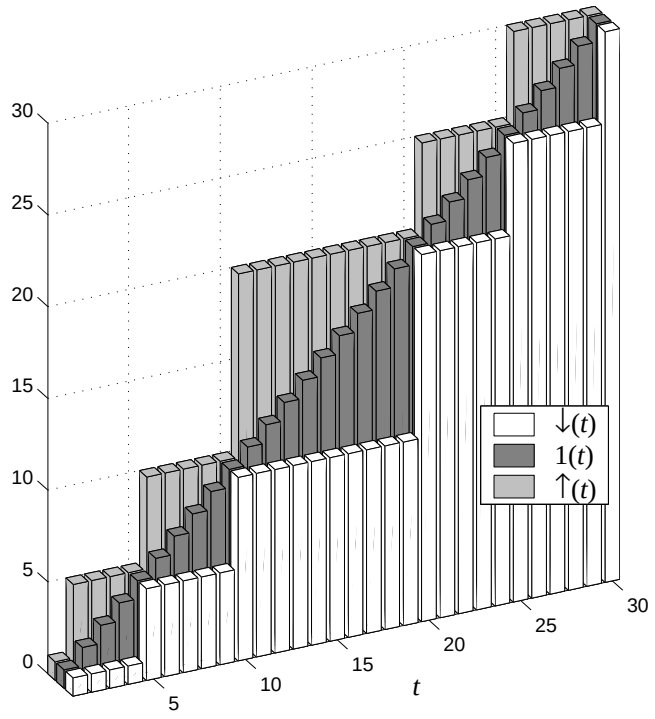

Figure 1: Exemplary aggregation operators on a 30-point index set

of the random vectors $\boldsymbol{\eta}_t$ and $\boldsymbol{\xi}_t$ is constant over time. Furthermore, assume that there are stochastic processes $\boldsymbol{\zeta}^{\uparrow} = (\boldsymbol{\eta}^{\uparrow}, \boldsymbol{\xi}^{\uparrow})$ and $\boldsymbol{\zeta}^{\downarrow} = (\boldsymbol{\eta}^{\downarrow}, \boldsymbol{\xi}^{\downarrow})$ according to definition 2.1, both of which are valued in the common state space Z. These processes are determined through the relations

$$\boldsymbol{\zeta}_t^{\uparrow} := \boldsymbol{\zeta}_{\uparrow(t)} \quad \text{and} \quad \boldsymbol{\zeta}_t^{\downarrow} := \boldsymbol{\zeta}_{\downarrow(t)} \quad \text{for all} \quad t = 1, \ldots, T.$$

Having in mind the reasoning of Sect. 3, $\boldsymbol{\zeta}^{\uparrow}$ and $\boldsymbol{\zeta}^{\downarrow}$ can be viewed as approximations of data process $\boldsymbol{\zeta}$. The corresponding induced filtrations are constructed in the usual way, i.e. set

$$\mathcal{F}^{\uparrow,t} := \sigma(\boldsymbol{\zeta}^{\uparrow,t}) \text{ for } t = 1, \ldots, T, \quad \mathcal{F}^{\uparrow} := \mathcal{F}^{\uparrow,T}, \quad \mathbb{F}^{\uparrow} := \{\mathcal{F}^{\uparrow,t}\}_{t=1}^{T},$$
$$\mathcal{F}^{\downarrow,t} := \sigma(\boldsymbol{\zeta}^{\downarrow,t}) \text{ for } t = 1, \ldots, T, \quad \mathcal{F}^{\downarrow} := \mathcal{F}^{\downarrow,T} \quad \mathbb{F}^{\downarrow} := \{\mathcal{F}^{\downarrow,t}\}_{t=1}^{T}.$$

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 $\boldsymbol{\zeta}$ follows a Markov process. The insights gained will be important for proving the main results in Sect. 5.1.

Proposition 5.1. *If the data process $\boldsymbol{\zeta}$ is Markovian, then the following conditions hold for suitable versions of the conditional expectations, respectively.*

(i) $E(\boldsymbol{x}|\mathcal{F}^{\downarrow}) \in X(\mathbb{F}^{\downarrow})$ *for all* $\boldsymbol{x} \in X(\mathbb{G}^{\downarrow})$

(ii) $E(\boldsymbol{y}|\mathcal{F}^{\downarrow}) \in Y(\mathbb{F}^{\downarrow})$ *for all* $\boldsymbol{y} \in Y(\mathbb{G}^{\downarrow})$

(iii) $E(\boldsymbol{x}|\mathcal{F}^{\uparrow}) \in X(\mathbb{F}^{\uparrow})$ *for all* $\boldsymbol{x} \in X(\mathbb{G}^{\uparrow})$

(iv) $E(\boldsymbol{y}|\mathcal{F}^{\uparrow}) \in Y(\mathbb{F}^{\uparrow})$ *for all* $\boldsymbol{y} \in Y(\mathbb{G}^{\uparrow})$

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

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

This, however, is true by proposition A.2 (iii) since the sets of random vectors $\{\boldsymbol{\zeta}_\tau\}_{\tau \leq t}$ and $\{\boldsymbol{\zeta}^{\downarrow}_\tau\}_{\downarrow(\tau) > t}$ are conditionally independent given $\{\boldsymbol{\zeta}^{\downarrow}_\tau\}_{\tau \leq t}$ for each fixed point $t = \downarrow(t)$. Conditional independence follows from proposition A.9. Consequently, assertion (i) is established. Notice that the proofs of the 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

$$E(\boldsymbol{\zeta}_t \,|\mathcal{F}^s) = \boldsymbol{\zeta}_s \quad P\text{-a.s.} \quad \text{for} \quad 1 \leq s \leq t \leq T. \tag{5.1}$$

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 5.1 when constructing bounds (see theorems 5.2 and 5.3).

Theorem 5.2. *Assume the conditions (C1), (C2), and (C3) to hold. If the data process $\boldsymbol{\zeta}$ represents a Markov-martingale, we find*

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

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

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

The first inequality is based on the fact that, due to the martingale property, the random vector $\boldsymbol{\xi}_t$ can be written as $E(\boldsymbol{\xi}_t^{\uparrow}|\mathcal{F}^t)$. 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 $E(\boldsymbol{\eta}_t|\mathcal{G}^{\downarrow,t})$ by the random vector $\boldsymbol{\eta}_t^{\downarrow}$. Finally, the fourth inequality is due to relaxation of the dual feasible set. Thus, we have shown

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

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

$$\inf \mathcal{P}(\mathbb{G}^{\downarrow}, \mathbb{G}^{\uparrow}; \boldsymbol{\eta}^{\downarrow}, \boldsymbol{\xi}^{\uparrow}) = \inf \mathcal{P}(\mathbb{F}^{\downarrow}, \mathbb{F}^{\uparrow}; \boldsymbol{\eta}^{\downarrow}, \boldsymbol{\xi}^{\uparrow}). \tag{5.2}$$

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

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

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 5.1 (iv), for each $\boldsymbol{y} \in Y(\mathbb{G}^{\uparrow})$ there exists a version of $E(\boldsymbol{y}_t|\mathcal{F}^{\uparrow})$ 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 (5.2) 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.

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

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 5.1 (i) and relaxation of the primal feasible set, i.e. for every $\boldsymbol{x} \in X(\mathbb{G}^{\downarrow})$ there exists a version of $E(\boldsymbol{x}|\mathcal{F}^{\downarrow})$ in $X(\mathbb{F}^{\downarrow})$. Consequently, we have proved (5.2), and thus the claim is established. □

Unlike in theorem 3.1, applicability of the conditional Jensen inequality relies on the internal structure of the Lagrangian and not just its convexity properties.

Theorem 5.3. *Assume the conditions (C1), (C2), and (C3) to hold. If the data process $\boldsymbol{\zeta}$ represents a Markov-martingale, we find*

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

Proof. Apart from minor exceptions, the proof is analogous to that of theorem 5.2. Without a detailed description of the involved manipulations, we may therefore

state the following chain of inequalities

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

When applying the conditional Jensen inequality in the third line, we explicitly use separability of the constraint functions[8] with respect to decisions and random parameters (note that $\boldsymbol{x}^t$ is generally not $\mathcal{G}^{\downarrow,t}$-measurable). Moreover, we substitute $\boldsymbol{\xi}_t^{\downarrow}$ for $E(\boldsymbol{\xi}_t|\mathcal{G}^{\downarrow,t})$, which is allowed by the martingale property of the data process. In summary, this yields

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

As in the proof of theorem 5.2, we will now use the Markov property of the data process to justify the equality

$$\inf \mathcal{P}(\mathbb{G}^{\uparrow}, \mathbb{G}^{\downarrow}; \boldsymbol{\eta}^{\uparrow}, \boldsymbol{\xi}^{\downarrow}) = \inf \mathcal{P}(\mathbb{F}^{\uparrow}, \mathbb{F}^{\downarrow}; \boldsymbol{\eta}^{\uparrow}, \boldsymbol{\xi}^{\downarrow}). \tag{5.3}$$

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

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

[8]It should be emphasized that separability of the constraint functions is not needed in the proof of theorem 5.2.

In the last line we indirectly use proposition 5.1 (ii), which guarantees that for each $\boldsymbol{y} \in Y(\mathbb{G}^{\downarrow})$ there exists a version of $E(\boldsymbol{y}|\mathcal{F}^{\downarrow})$ which is in $Y(\mathbb{F}^{\downarrow})$. To prove the converse inequality in (5.3), we first relax the primal feasible set, then use the conditional Jensen inequality, and finally restrict the dual feasible set:

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

Observe that the last inequality is based on proposition 5.1 (iii), i.e. for all primal decision vectors $\boldsymbol{x} \in X(\mathbb{G}^{\uparrow})$ there is a version of the conditional expectation such that $E(\boldsymbol{x}|\mathcal{F}^{\uparrow}) \in X(\mathbb{F}^{\uparrow})$. Combining the above results proves (5.3), and thus the claim follows. □

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 5.2 and 5.3 critically rely on the martingale property of the data process. Although martingales enjoy wide popularity both in discrete and continuous time finance, see e.g. [22] and [23], 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 $\boldsymbol{\zeta}$ is a Markov process satisfying the conditions

$$\left.\begin{array}{rcl} E(\boldsymbol{\eta}_t \,|\mathcal{F}^s) & = & H^{\mathrm{o}}_{t,s}(\boldsymbol{\eta}_s) \\ E(\boldsymbol{\xi}_t \,|\mathcal{F}^s) & = & H^{\mathrm{c}}_{t,s}(\boldsymbol{\xi}_s) \end{array}\right\} \quad P\text{-a.s.} \quad \text{for} \quad 1 \leq s \leq t \leq T. \tag{5.4}$$

The mappings $H^{\mathrm{o}}_{t,s}$ and $H^{\mathrm{c}}_{t,s}$ are linear affine and invertible. The corresponding inverse mappings will be denoted by $H^{\mathrm{o}}_{s,t}$ and $H^{\mathrm{c}}_{s,t}$, 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

$$H^{\mathrm{i},\alpha} \;:=\; H^{\mathrm{i}}_{1,\alpha(1)} \times \cdots \times H^{\mathrm{i}}_{T,\alpha(T)} \quad (\mathrm{i},\alpha) \in \{\mathrm{o},\mathrm{c}\} \times \{\uparrow,\downarrow\},$$

which depend on the aggregation operators. The next theorem generalizes the theorems 5.2 and 5.3 by allowing for data processes which need not be martingales.

Theorem 5.4. *Assume the conditions (C1), (C2), and (C3) to hold. If the data process $\boldsymbol{\zeta}$ represents a linear Markov process, we find*

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

Proof. The statements are proved exactly as in theorems 5.2 and 5.3. However, the relations (5.4) are used instead of the martingale property. □

It is worthwhile to remark that the mappings $H^{\mathrm{o},\uparrow}$ and $H^{\mathrm{c},\uparrow}$ will diverge with respect to the matrix 2-norm, say, if $H^{\mathrm{o},\downarrow}$ and $H^{\mathrm{c},\downarrow}$ become singular. This can

happen if correlation between the random vectors $\boldsymbol{\zeta}_t$ is low across neighboring stages. In any such situation, stage-aggregation may not be justifiable, and the bounds proposed in theorem 5.4 may become very coarse. In contrast, the bounds are expected to be tight in case the $\boldsymbol{\zeta}_t$'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 [34, Chap. 2].

5.3 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.

To be specific, assume that the given original stochastic program complies with the regularity conditions (C1), (C2), and (C3). Furthermore, let $\boldsymbol{\zeta}$ be a linear Markov process. Then, decision stages may be aggregated as in theorem 5.4. Denote by t_s the (common) fixed points of the aggregation operators $\uparrow$ and $\downarrow$, where the integer s ranges from 1 to S. Here, it should be emphasized again that S is usually much smaller than T (cf. also Fig. 1). Moreover, observe that the approximate stochastic programs of theorem 5.4 merely depend on the random vectors $\{\boldsymbol{\zeta}_{t_s}\}_{s=1}^{S}$, which define a reduced stochastic process $\boldsymbol{\zeta}'$ comprising S stages. The related processes $\boldsymbol{\eta}'$ and $\boldsymbol{\xi}'$ are defined in the intuitive way. Both aggregated minimization problems of theorem 5.4 can then be recast as S-stage stochastic programs with primal and dual strategies adapted to the filtration induced by $\boldsymbol{\zeta}'$. It is important to notice that these S-stage problems have a slightly

different structure than the original problem $\mathcal{P}$ since the constraint function of any stage s may depend on the decisions associated with stage $s+1$. However, the corresponding Lagrangians still represent subdifferentiable saddle functions being jointly convex in ξ' and the primal decisions while being jointly concave in η' and the dual decisions. If the reduced process $\boldsymbol{\zeta}'$ can be approximated by discrete processes $\boldsymbol{\zeta}'^{,u}$ and $\boldsymbol{\zeta}'^{,l}$ as in Sect. 3, then the theorems 3.1 and 3.2 remain valid. In this case, we can establish two discrete S-stage stochastic programs providing upper and lower bounds on the optimal value of the original problem $\mathcal{P}$.

6 Conclusions

This article addresses the approximation of convex multistage stochastic programs via aggregation of decision stages and discretization of the underlying probability space. In other words, the temporal and spacial granularity of some given stochastic program is coarsened in a sophisticated way. Thereby, *deterministic* bounds on the optimal objective value are constructed (as opposed to a *probabilistic* confidence interval). By adapting the reasoning in [32, 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 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. Thereby, 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 [42]. 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 *sub*filtrations. Conversely, the stage-aggregation scheme presented here (see Sect. 5) also involves *refining*. In fact, certain filtrations are replaced by suitable *super*filtrations. 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.

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

A 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 A.1 (Conditional Independence)**.** *We say that the sets of random vectors A and B are conditionally independent given C if*

$$P(\mathsf{A} \cap \mathsf{B} \,|\, \sigma(C)) = P(\mathsf{A} \,|\, \sigma(C))\, P(\mathsf{B} \,|\, \sigma(C)) \quad P\text{-a.s.}$$

For all $\mathsf{A} \in \sigma(A)$ and $\mathsf{B} \in \sigma(B)$. We will use the shorthand notation $A \perp\!\!\!\perp B \,|\, C$ 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 $A \perp\!\!\!\perp B$.

Observe that conditional independence of sets of random variables depends on the choice of the probability measure P.

Proposition A.2 (Equivalent Characterizations)**.** *Conditional independence of A and B given C is equivalent to any one of the following statements:*

(i) for all $\mathsf{A} \in \sigma(A)$ we have

$$P(\mathsf{A} \,|\, \sigma(B \cup C)) = P(\mathsf{A} \,|\, \sigma(C)) \quad P\text{-a.s.}$$

(ii) for every $\boldsymbol{\alpha} \in \mathcal{L}^\infty(\Omega, \sigma(A), P; \mathbb{R})$ and $\boldsymbol{\beta} \in \mathcal{L}^\infty(\Omega, \sigma(B), P; \mathbb{R})$ we have

$$E(\boldsymbol{\alpha}\,\boldsymbol{\beta} \,|\, \sigma(C)) = E(\boldsymbol{\alpha} \,|\, \sigma(C))\, E(\boldsymbol{\beta} \,|\, \sigma(C)) \quad P\text{-a.s.}$$

(iii) for every $\boldsymbol{\alpha} \in \mathcal{L}^\infty(\Omega, \sigma(A), P; \mathbb{R})$ we have

$$E(\boldsymbol{\alpha} \,|\, \sigma(B \cup C)) = E(\boldsymbol{\alpha} \,|\, \sigma(C)) \quad P\text{-a.s.}$$

Proof. The claim follows from standard measure-theoretic arguments as exemplified e.g. in [6, Sect. 7.3]. □

Proposition A.3 (Basic Properties)**.** *The conditional independence relation has the following basic properties:*

(i) Symmetry:

$$A \perp\!\!\!\perp B \,|\, C \quad \Rightarrow \quad B \perp\!\!\!\perp A \,|\, C;$$

(ii) Decomposition:

$$A \perp\!\!\!\perp B \cup D \,|\, C \quad \Rightarrow \quad A \perp\!\!\!\perp B \,|\, C;$$

(iii) Weak Union:

$$A \perp\!\!\!\perp B \cup D \,|\, C \quad \Rightarrow \quad A \perp\!\!\!\perp B \,|\, C \cup D;$$

(iv) Contraction:

$$A \perp\!\!\!\perp B \,|\, C \quad \textit{and} \quad A \perp\!\!\!\perp D \,|\, B \cup C \quad \Rightarrow \quad A \perp\!\!\!\perp B \cup D \,|\, C.$$

Proof. [8, Sects. 5 and 6] □

Furthermore, it can easily be checked from the definitions that the conditional independence relation exhibits the *trivial independence* property $A \perp\!\!\!\perp B \,|\, B$. In the sequel 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. [35] and the references therein. To this end, assume that the set of random vectors $V = \{\boldsymbol{\zeta}_t\}_{t=1}^{T}$ is totally ordered.[9] Let V_t be the set of the first $t-1$ elements[10] with respect to the given ordering, and let B_t be a subset of V_t satisfying the *a priori* conditional independence relationship

$$\{\boldsymbol{\zeta}_t\} \perp\!\!\!\perp V_t \backslash B_t \mid B_t.$$

[9] 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.

[10] Assume that V_1 is the empty set.

By proposition A.2 (i) this means that the conditional distribution of $\boldsymbol{\zeta}_t$ given $\sigma(V_t)$ almost surely coincides with the conditional distribution of $\boldsymbol{\zeta}_t$ 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 $\boldsymbol{\zeta}_t$, i.e. from each vertex in B_t draw an arc terminating in $\boldsymbol{\zeta}_t$. The set of edges E constructed in this manner is considered as a subset of $V \times V$.

$$E \;=\; \bigcup_{t=1}^{T} \{(\boldsymbol{\zeta}_s, \boldsymbol{\zeta}_t) \,|\, \boldsymbol{\zeta}_s \in B_t\}$$

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 A.4 (d-separation)**.** *(Pearl [35, Sect. 3.3.1]) Let A, B, and C be three disjoint sets of nodes in the graph G. We say that C d-separates*[11] *A and B if along every path between a node in A and a node in B there is a node $\boldsymbol{\zeta}_t$ satisfying one of the following conditions: (i) $\boldsymbol{\zeta}_t$ is a head-to-head node along the path and neither $\boldsymbol{\zeta}_t$ nor any of its descendants are in C, or (ii) $\boldsymbol{\zeta}_t$ is not head-to-head but is in C.*

Theorem A.5. *Let A, B, and C be disjoint sets of nodes in the graph G. If C d-separates A and B, then A and B are conditionally independent given C.*

Proof. [41, Theorem 2] □

[11] d-separation stands for direction-dependent separation.

The converse statement need not be true, i.e. if A and B are conditionally independent given C, we may not conclude that C necessarily d-separates the sets A and B; see the discussion at the end of [35, Sect. 3.3.1]. Moreover, it is worthwhile to mention that the proof of theorem A.5 exclusively relies on the basic properties established in proposition A.3.

Proposition A.6. *Assume that* $V = \{\boldsymbol{\zeta}_t\}_{t=1}^T$ *and* $V^* = \{\boldsymbol{\zeta}_t^*\}_{t=1}^T$ *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*

$$\{\boldsymbol{\zeta}_t^*\} \perp\!\!\!\perp V \backslash \{\boldsymbol{\zeta}_t\} \mid V_t^* \cup \{\boldsymbol{\zeta}_t\}, \quad t = 1, \ldots, T.$$

Then, we find

$$V_t \perp\!\!\!\perp V^* \backslash V_t^* \mid V_t^* \quad \textit{and} \quad V_t^* \perp\!\!\!\perp V \backslash V_t \mid V_t, \quad t = 1, \ldots, T.$$

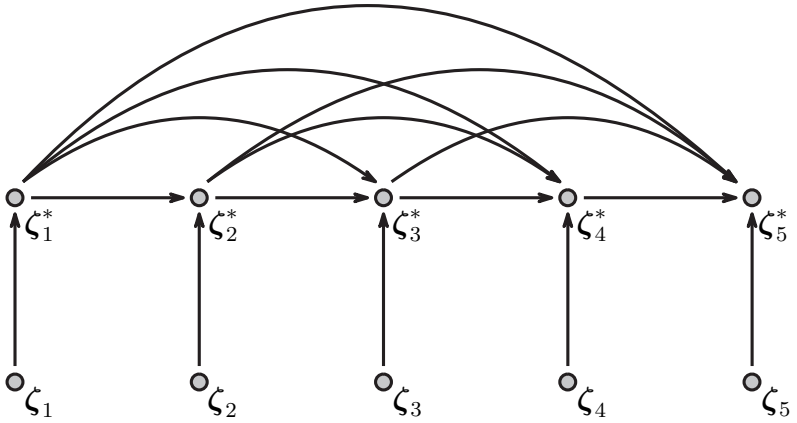

Figure 2: Directed acyclic graph

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 $\boldsymbol{\zeta}_t'$ where the index t ranges from 1 to $2T$ and

$$\boldsymbol{\zeta}_t' := \begin{cases} \boldsymbol{\zeta}_t & \text{for } t = 1, \ldots, T, \\ \boldsymbol{\zeta}_{t-T}^* & \text{for } t = T+1, \ldots, 2T. \end{cases}$$

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

$$B'_t := \begin{cases} \emptyset & \text{for } t = 1, \ldots, T, \\ V^*_{t-T} \cup \{\boldsymbol{\zeta}_{t-T}\} & \text{for } t = T+1, \ldots, 2T. \end{cases}$$

By construction, the postulated conditional and unconditional independence relationships translate to

$$\{\boldsymbol{\zeta}'_t\} \perp\!\!\!\perp V'_t \backslash B'_t \mid B'_t, \quad t = 1, \ldots, 2T.$$

As before, we can now construct a directed acyclic graph $G' = (V', E')$ with vertices V' and directed edges

$$E' \;=\; \bigcup_{t=1}^{2T} \{(\boldsymbol{\zeta}'_s, \boldsymbol{\zeta}'_t) \,|\, \boldsymbol{\zeta}'_s \in B'_t\}.$$

Figure 2 visualizes G' for the special case $T = 5$. The first statement now follows by application of theorem A.5 if we can show that V^*_t d-separates V_t and $V^* \backslash V^*_t$ (all sets are considered as subsets of V'). In fact, every path from V_t to $V^* \backslash V^*_t$ 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 A.5 if we can show that V_t d-separates V^*_t and $V \backslash V_t$. It turns out that every path from V^*_t to $V \backslash V_t$ passes by some node in $V^* \backslash V^*_t$ 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. □

Proposition A.7. *Under the assumptions of proposition A.6 we also have*

$$\{\boldsymbol{\zeta}_t\} \perp\!\!\!\perp \{\boldsymbol{\zeta}^*_\tau\}_{\tau=t+1}^{T} \mid \{\boldsymbol{\zeta}^*_\tau\}_{\tau=1}^{t}, \quad t = 1, \ldots, T.$$

Proof. Bearing in mind the special topology of the graph $G' = (V', E')$ considered in the proof of proposition A.6, it is clear that every path from $\boldsymbol{\zeta}_t$ to the set $\{\boldsymbol{\zeta}^*_\tau\}_{\tau=t+1}^{T}$ passes by the node $\boldsymbol{\zeta}^*_t$, which is not head-to-head along the path. This implies d-separation of $\{\boldsymbol{\zeta}_t\}$ and $\{\boldsymbol{\zeta}^*_\tau\}_{\tau=t+1}^{T}$ given $\{\boldsymbol{\zeta}^*_\tau\}_{\tau=1}^{t}$. The claim now follows from theorem A.5. □

Definition A.8 (Markov Process)**.** *Let $V = \{\boldsymbol{\zeta}_t\}_{t=1}^T$ 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*

$$\{\boldsymbol{\zeta}_{t+1}\} \perp\!\!\!\perp V_t \mid \{\boldsymbol{\zeta}_t\}, \quad t = 2, \ldots, T-1.$$

For the further argumentation, let $\alpha : \{1, \ldots, T\} \to \{1, \ldots, T\}$ be a generalized aggregation operator, i.e. we require that α be monotonously 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

$$U_t^\alpha := \{\boldsymbol{\zeta}_{\alpha(\tau)}\}_{\tau=1}^t, \quad t = 1, \ldots, T.$$

Moreover, define $U := U_T$ and $U^\alpha := U_T^\alpha$.

Proposition A.9 (Locality)**.** *For each $t = \alpha(t)$ we have $U_t \perp\!\!\!\perp U^\alpha \backslash U_t^\alpha \mid U_t^\alpha$.*

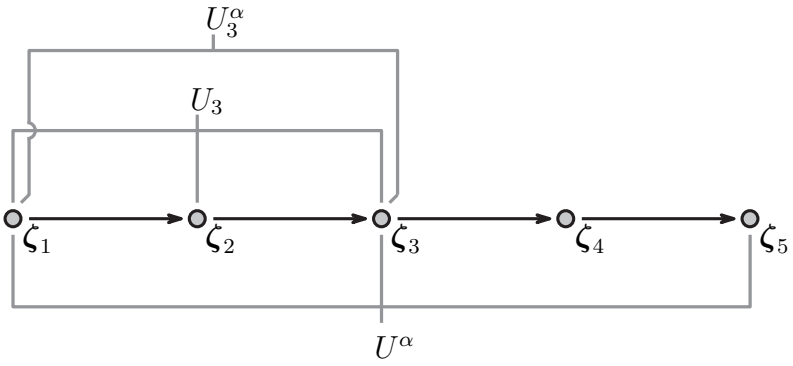

Figure 3: Graphical representation of a Markov process

Proof. We represent the Markov process in the usual way as a directed acyclic graph $G = (V, E)$. By definition A.8, the sets of parents of the individual vertices can be chosen to be $B_1 := \emptyset$ and $B_t := \{\boldsymbol{\zeta}_{t-1}\}$ for all $t > 1$. Figure 3 visualizes this simple 'linear' graph for the special case $T = 5$. Next, choose an arbitrary

fixed point $t = \alpha(t)$. Since every path from $U_t \backslash U_t^\alpha$ to $U^\alpha \backslash U_t^\alpha$ passes by the node $\boldsymbol{\zeta}_t \in U_t^\alpha$, which is not head-to-head along the path, we may invoke theorem A.5 to conclude that

$$U_t \backslash U_t^\alpha \perp\!\!\!\perp U^\alpha \backslash U_t^\alpha \mid U_t^\alpha.$$

The claim now follows from the trivial independence and contraction properties of the conditional independence relation (cf. proposition A.3). □

References

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

[2] Birge, J. Aggregation in stochastic production models. In *Proceedings of the 11th IFIP Conference on System Modelling and Optimization* (New York, 1984), Springer-Verlag.

[3] Birge, J. Aggregation in stochastic linear programming. *Mathematical Programming 31* (1985), 25–41.

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

[5] Birge, J., and Wets, R.-B. Computing bounds for stochastic programming problems by means of a generalized moment problem. *Mathematics of Operations Research 12* (1987), 149–162.

[6] Chow, Y., and Teicher, H. *Probability Theory*, 3rd ed. Springer-Verlag, New York, 1997.

[7] Dantzig, G., and Infanger, G. Large-scale stochastic linear programs–importance sampling and Benders decomposition. *Computational and Applied Mathematics I* (1992), 111–120.

[8] Dawid, A. Conditional independence for statistical operations. *The Annals of Statistics 8*, 3 (1980), 598–617.

[9] Dempster, M., and Thompson, R. EVPI-based importance sampling solution procedures for multistage stochastic linear programmes on parallel MIMD architectures. *Annals of Operations Research 90* (1999), 161–184.

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

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

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

[13] Edirisinghe, N., and Ziemba, W. Bounds for two-stage stochastic programs with fixed recourse. *Mathematics of Operations Research 19* (1994), 292–313.

[14] Ermoliev, Y., and Gaivoronski, A. Stochastic quasigradient methods for optimization of discrete event systems. *Annals of Operations Research 39* (1992), 1–39.

[15] Flåm, S., and Wets, R.-B. Finite horizon approximates of infinite horizon stochastic programs. *Stochastic Optimization 81* (1986), 337–350.

[16] Flåm, S., and Wets, R.-B. Existence results and finite horizon approximates for infinite horizon optimization problems. *Econometrica 55* (1987), 1187–1209.

[17] Frauendorfer, K. Solving SLP recourse problems with arbitrary multivariate distributions — the dependent case. *Mathematics of Operations Research 13* (1988), 377–394.

[18] Frauendorfer, K. *Stochastic two-stage programming*, vol. 392 of *Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag, Berlin, 1992.

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

[20] Frauendorfer, K. Barycentric scenario trees in convex multistage stochastic programming. *Mathematical Programming 75*, 2 (1996), 277–294.

[21] Gassmann, H., and Ziemba, W. A tight upper bound for the expectation of a convex function of a multivariate random variable. *Mathematical Programming Study 27* (1986), 39–53.

[22] Harrison, J., and Kreps, D. Martingales and arbitrage in multi-period securities markets. *Journal of Economic Theory 20* (1979), 381–408.

[23] Harrison, J., and Pliska, S. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications 11* (1981), 215–260.

[24] Heitsch, H., and Römisch, W. Scenario reduction algorithms in stochastic programming. *Computational Optimimzation and Applications 24* (2003), 187–206.

[25] Higle, J., and Sen, S. Stochastic decomposition: An algorithm for two-stage linear programs with recourse. *Mathematics of Operations Research 16* (1991), 650—-669.

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

[27] Infanger, G. *Planning under Uncertainty: Solving Large-Scale Stochastic Linear Programs*. Boyd and Fraser, Danvers, 1994.

[28] Kall, P. An upper bound for SLP using first and total second moments. *Annals of Operations Research 30* (1991), 267–276.

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

[30] Kaut, M., and Wallace, S. Evaluation of scenario-generation methods for stochastic programming. The Stochastic Programming E-Print Series (SPEPS), May 2003.

[31] 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.

[32] Kuhn, D. *Generalized Bounds for Convex Multistage Stochastic Programs*, vol. 548 of *Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag, Berlin, 2004.

[33] Madansky, A. Inequalities for stochastic linear programming problems. *Management Science 6* (1960), 197–204.

[34] Meyn, S., and Tweedie, R. *Markov Chains and Stochastic Stability*. Springer-Verlag, New York, 1996.

[35] Pearl, J. *Probabilistic Reasoning in Intelligent Systems*, 2nd ed. Morgan Kaufman, San Mateo, CA, 1991.

[36] Pflug, G. Scenario tree generation for multiperiod financial optimization by optimal discretization. *Mathematical Programming, Ser. B 89* (2001), 251–271.

[37] Prekopa, A. *Stochastic Programming*. Kluwer Academic Publishers, Dordrecht, 1995.

[38] Rachev, S., and Römisch, W. Quantitative stability in stochastic programming: the method of probability metrics. *Mathematics of Operations Research 27* (2002), 792–818.

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

[40] Schervish, M. *Theory of Statistics*. Springer-Verlag, New York, 1995.

[41] Verma, T., and Pearl, J. Causal networks and expressiveness. In *Proceedings of the 4^{th} Workshop on Uncertainty in Artificial Intelligence* (Mountain View, CA, 1988), pp. 352–359.

[42] Wright, S. Primal-dual aggregation and disaggregation for stochastic linear programs. *Mathematics of Operations Research 19*, 4 (1994), 893–908.