

# Convergent Bounds for Stochastic Programs with Expected Value Constraints

Daniel Kuhn<sup>†</sup>

dkuhn@doc.ic.ac.uk

Department of Computing

Imperial College of Science, Technology, and Medicine

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

18th September 2006

**Abstract.** This article elaborates a bounding approximation scheme for convex multistage stochastic programs (MSP) that constrain the conditional expectation of some decision-dependent random variables. Expected value constraints of this type are useful for modelling a decision maker's risk preferences, but they may also arise as artefacts of stage-aggregation. It is shown that the gap between certain upper and lower bounds on the optimal objective value can be made smaller than any prescribed tolerance. Moreover, the solutions of some tractable approximate MSP give rise to a policy which is feasible in the (untractable) original MSP, and this policy's cost differs from the optimal cost at most by the difference between the bounds. The considered problem class comprises models with integrated chance constraints and conditional value-at-risk constraints. No relatively complete recourse is assumed.

**Key words.** stochastic programming, approximation, bounds, expected value constraints, integrated chance constraints

---

<sup>†</sup>On leave from the Institute for Operations Research and Computational Finance, University of St. Gallen, Switzerland.

# 1 Introduction

A multistage stochastic program (MSP) constitutes a mathematical model of a real-life decision problem under uncertainty in which several decisions must be made based on different levels of information [1,26,36]. The correct representation of the underlying uncertainty and the corresponding information structure is crucial. Usually, a suitable set of decision-relevant random parameters is modelled as a vector-valued stochastic process in discrete time. Whenever a decision is due to be taken, complete and accurate information about the past and present realizations of this process is assumed to be available, whereas future realizations are known only in terms of their joint (conditional) probability distribution. Typically, one attempts to find decisions that are subject to a set of explicit constraints and optimal with respect to an appropriate objective criterion. Without much loss of generality, we will assume in this article that the objective criterion is to minimize expected cost.

Stochastic programming techniques have been successfully used in a vast number of very different applications, see e.g. [49] for a recent survey. Although it is relatively easy to formulate a given decision problem as an MSP, its solution is rarely possible without significant approximations. The reason for this is that the original MSP formulation usually involves continuously distributed random parameters and too many decision stages. Traditional approximation schemes replace the underlying stochastic process (with an infinite number of sample paths) by a finitely supported approximate process, which can be represented as a *scenario tree*. This amounts to approximating the original MSP over an infinite-dimensional function space by a numerically tractable MSP over a finite-dimensional Euclidean space.

Different approaches to scenario tree construction or *scenario generation* apply to problems with different structural properties, and they provide different error estimates and convergence characteristics, see also the survey in [27]. Among the

most wide-spread techniques range the moment-matching method [23], the Monte Carlo and Quasi-Monte Carlo schemes [29, 34, 47, 48], a class of methods based on probability metrics [8, 19, 22, 35, 38], as well as the internal sampling methods [5, 6, 12, 21, 24], which generate the scenarios dynamically within the solution procedure of the MSP. The present article further elaborates on yet another class of scenario generation methods, which received continuing attention for several decades: the so-called bounding approximation schemes [2, 9–11, 13–17, 25, 30, 32]. They provide deterministic upper and/or lower bounds on the optimal value of the original MSP, and their applicability relies on certain convexity properties of the recourse (cost-to-go) functions of the underlying problem. Traditionally, these methods were analyzed in a one- or two-stage setting. The fewer multistage extensions were cast in a dynamic programming framework, see e.g. [15, 16, 30]. By using more abstract measure-theoretic methods due to Wright [51], the author reconsidered these bounding approximation schemes in a mathematical programming context [31]. Although addressing multistage problems, the use of dynamic programming recursions was completely avoided. In this setting, the bounding approximations could be combined with another approximation based on stage-aggregation, which resulted in a substantial complexity reduction. The present article discusses further benefits of the bounding approximations introduced in [31], proves their asymptotic consistency, and demonstrates that they apply even if the underlying MSP contains certain expected value constraints.

Expected value constraints have first been studied by Prékopa [36, 37]. They arise in different contexts and prove particularly useful for modelling a decision maker’s risk attitude. An inequality constraint involving present decisions and future random outcomes, for instance, can be assumed to hold with probability 1, with probability  $p < 1$ , or in expectation. In the first case, it might entail a cumbersome feasible set depending largely on certain worst-case scenarios, which are difficult or even impossible to predict. If the cost of violating the constraint in a few unlikely scenarios is not too high, this pessimistic (robust) approach

will lead to overly conservative decisions. A simple remedy to overcome this pessimism is by requiring the constraint to hold with probability  $p \lesssim 1$ , in which case it becomes a so-called *chance constraint* [4]. Although we may now obtain more reasonable decisions, chance constraints have serious mathematical drawbacks, i.e., they can lead to nonconvex or even disconnected feasible sets. They are also qualitative in nature: extreme violations of the constraint are penalized no more than slight violations. Let us assume for a moment that the constraint function under consideration is nonnegative and can be interpreted as a form of loss or shortfall. Depending on the particular application, a more satisfactory approach could be to constrain the expectation of the shortfall. As follows from Markov's inequality, requiring the expected shortfall to be smaller than some (strictly positive) target value implies infinitely many chance constraints to hold for a continuum of significance levels. Conversely, a well-known formula for expected shortfall functions [28, § 5] can be used to prove that an expected value constraint of the above kind corresponds to an integral of chance constraints over a continuum of threshold levels. For this reason, constraints on expected shortfall were termed *integrated chance constraints* (ICC) by Klein Haneveld [18, 28]. They allow the shortfall to become negative in certain scenarios, thereby avoiding overly conservative decisions. Moreover, they are quantitative in nature as they prefer small losses to large losses. Under reasonably general assumptions, they even preserve convexity of the feasible set. ICCs thus share many of the convenient properties of conventional chance constraints without exhibiting their unpleasant side effects.

As discussed above, expected value constraints are useful for controlling the risk associated with certain decision strategies, e.g. by means of ICCs. Another popular way to mitigate any kind of shortfall risk is by imposing constraints on the conditional value-at-risk (CVaR) of the losses [40]. If the losses depend linearly on the decision variables, the CVaR constraints can conveniently be represented as linear expected value constraints [41, § 4]. The relation between ICCs and

CVaR constraints is analyzed in [18, § 4.2].

Expected value constraints can also originate from approximations. Stochastic inventory problems with an indefinite horizon, for instance, are frequently approximated by simpler problems with a finite horizon. The errors due to this simplification are termed *end effects*. Suitable constraints in the last stage of the approximate problem usually help to mitigate such end effects, e.g. by prohibiting depletion of the inventory over the (truncated) planning horizon. Almost sure constraints requiring the terminal inventory level to be higher than some experiential target value tend to be overly restrictive. It is often better to require the *expected* terminal inventory level to exceed a fraction of its historical average. Another reason for our interest in expected value constraints is their appearance in stage-aggregated MSPs, see [31, § 5]. In fact, stochastic programs with many decision stages can occasionally be approximated by two simpler stochastic programs with only few (effective) stages, which provide upper and lower bounds on the true optimal value. The lower aggregated problem contains expected value constraints which originate from specific relaxations.

The above examples highlight the importance of expected value constraints for both modelling and approximation purposes. Even if they arise from preliminary approximations, MSPs with expected value constraints generically constitute infinite-dimensional optimization problems which fail to be numerically tractable. The main goal of this article consists in designing a bounding approximation scheme for a class of convex MSPs with conditional expectations in the constraints. Using the methodology of [31], we construct two finite-dimensional approximate stochastic programs which provide upper and lower bounds on the true optimal value. The main focus of our exposition is on proving convergence of the bounds due to suitable refinements of the underlying discretizations. The use of dynamic programming techniques is completely avoided, which allows us to omit the relatively complete recourse assumptions underlying previous work on bounding approximations [15, 16].

The rest of this article is organized as follows. A rigorous description of convex MSPs with expected value constraints is provided in Section 2. Reconsidering any such decision problem in a conjugate duality framework, as is done in Section 3, facilitates the development and assessment of a flexible approximation scheme in Section 4. We then derive computationally accessible upper and lower bounds on the optimal objective value. Section 5 goes one step further and characterizes a computationally accessible decision strategy which is practically feasible and whose objective value lies between the bounds determined beforehand. A systematic approach to reducing the gap between the bounds is presented in Section 6. In particular, we show that this gap can principally be made smaller than any prescribed tolerance. After an illustrative example in Section 7, conclusions are drawn in Section 8.

## 2 Problem Formulation

Consider a cost minimization problem under uncertainty with expected value constraints. Assume that decisions may be selected at different time points (or stages)  $t = 1, \dots, T$  when new information about the underlying random parameters becomes available. To formalize the treatment of uncertainty, we let all random objects be defined on a complete probability space  $(\Omega, \Sigma, P)$ , which is referred to as the *sample space* in probability theory. As in [31], we further use the following definition of a stochastic process.

**Definition 2.1** (Stochastic Process). *We say that  $\zeta$  is a stochastic process with state space  $Z$  if  $\zeta = (\zeta_1, \dots, \zeta_T)$  and  $Z = \times_{t=1}^T Z_t$  such that each random vector  $\zeta_t$  maps  $(\Omega, \Sigma)$  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  $\zeta^t := (\zeta_1, \dots, \zeta_t)$  valued in  $Z^t := \times_{\tau=1}^t Z_\tau$  for all  $t = 1, \dots, T$ .*

All stochastic processes considered in this article are assumed to be of the

above kind. By convention, random objects (i.e., random variables, random vectors, or stochastic processes) appear in boldface, while their realizations are denoted by the same symbols in normal face. Note that we will mostly deal with stochastic processes taking values in compact state spaces. By boundedness, these processes have finite moments of all orders.

A general nonlinear multistage stochastic program (MSP) with expected value constraints can now be formulated as

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

Here,  $\boldsymbol{\eta}$  and  $\boldsymbol{\xi}$  are two given stochastic processes with state spaces  $\Theta$  and  $\Xi$ , respectively. Note that  $\boldsymbol{\eta}$  impacts only the objective function of  $\mathcal{P}$ , whereas  $\boldsymbol{\xi}$  appears exclusively in the constraints. To keep notation simple, we also 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 observing the data process is formally represented as the  $\sigma$ -algebra induced by the random data revealed from time 1 up to time  $t$ , that is,  $\mathcal{F}^t := \sigma(\boldsymbol{\zeta}^t)$ . Furthermore, we use the convention  $\mathcal{F} := \mathcal{F}^T$  and let  $\mathbb{F} := \{\mathcal{F}^t\}_{t=1}^T$  be the filtration induced by the data process.<sup>1</sup>

Minimization in  $\mathcal{P}$  is over a convex set of stochastic processes  $\mathbf{x}$  with common state space  $X \subset \times_{t=1}^T \mathbb{R}^{n_t}$ . These processes are referred to as strategies, policies, decision rules, or decision processes. The set of admissible strategies is defined as

$$X(\mathbb{F}) := \{\mathbf{x} \in \times_{t=1}^T \mathcal{L}^\infty(\Omega, \mathcal{F}^t, P; \mathbb{R}^{n_t}) \mid \mathbf{x} \text{ is valued in } X \text{ } P\text{-a.s.}\}.$$

By definition,  $X(\mathbb{F})$  merely contains  $\mathbb{F}$ -adapted strategies. This implies that decisions are chosen *non-anticipatively* with respect to the underlying data process, see e.g. [46]. For later use, let  $X_0(\mathbb{F})$  be the linear hull of  $X(\mathbb{F})$ .

The real-valued cost function  $c : X \times \Theta \rightarrow \mathbb{R}$  and the vector-valued constraint functions  $f_t : X \times \Xi \rightarrow \mathbb{R}^{m_t}$  for  $t = 1, \dots, T$  are assumed to be Borel measur-

---

<sup>1</sup>In many applications,  $\boldsymbol{\zeta}_1$  constitutes a deterministic random vector. Our exposition, however, will not rely on this assumption.

able and bounded. This minimal requirement will be further tightened, below, to ensure applicability of the approximation scheme presented in Section 4 and beyond. The objective criterion in  $\mathcal{P}$  is to minimize expected cost. Decisions are subject to a set of almost sure constraints, which are assumed to hold in expectation (conditional on the stagewise information sets, respectively).

By our assumptions on the cost and constraint functions, the minimization problem  $\mathcal{P}$  is in fact well-defined.<sup>2</sup> For the further analysis, we impose the following regularity conditions:

- (C1)  $c$  is convex in  $x$ , concave in  $\eta$ , and continuous on  $X \times \Theta$ ;
- (C2)  $f_t$  is convex and continuous on  $X \times \Xi$  for each  $t = 1, \dots, T$ ;
- (C3)  $X_t$  is a convex compact subset of  $\mathbb{R}^{n_t}$  for each  $t = 1, \dots, T$ ;
- (C4) problem  $\mathcal{P}$  is strictly feasible.

Strict feasibility means that there is an  $\varepsilon > 0$  and a policy  $\mathbf{x}_s \in X(\mathbb{F})$  such that

$$\mathbb{E}(f_t(\mathbf{x}_s, \boldsymbol{\xi}) | \mathcal{F}^t) \leq -\varepsilon \mathbf{1}_t \quad P\text{-a.s.} \quad \forall t = 1, \dots, T,$$

where  $\mathbf{1}_t$  is the element of  $\mathcal{L}^\infty(\Omega, \Sigma, P; \mathbb{R}^{m_t})$  with all components equal to 1. Notice that strict feasibility fails to hold in the presence of equality constraints (which can be represented as two opposing inequality constraints in  $\mathcal{P}$ ). This, however, is no serious deficiency since any model involving equality constraints can be recast as a model without equality constraints by redefining  $X$ .

### 3 Lagrangian Reformulation

In order to derive a flexible approximation scheme for problem  $\mathcal{P}$ , we rewrite it in terms of a suitable Lagrangian function, using the elegant language of conjugate duality theory due to Rockafellar [39]. Various duality schemes for convex

---

<sup>2</sup>A priori,  $\mathcal{P}$  neither needs to be solvable nor feasible.



stochastic programs are reported in literature, see e.g. [42–46]. Our approach is strongly inspired by [43] and starts with embedding  $\mathcal{P}$  into a parametric family of convex stochastic programs. The underlying perturbation parameter  $\mathbf{u}$  constitutes a stochastic process which ranges over the linear space

$$U_0(\mathbb{F}) := \times_{t=1}^T \mathcal{L}^\infty(\Omega, \Sigma, P; \mathbb{R}^{m_t}).$$

Concretely speaking, we consider the family of minimization problems

$$\underset{\mathbf{x} \in X_0(\mathbb{F})}{\text{minimize}} F(\mathbf{x}, \mathbf{u}), \quad (\mathcal{P}(\mathbf{u}))$$

where the parametric objective function  $F : X_0(\mathbb{F}) \times U_0(\mathbb{F}) \rightarrow [-\infty, +\infty]$  is defined through

$$F(\mathbf{x}, \mathbf{u}) = \begin{cases} \mathbb{E}(c(\mathbf{x}, \boldsymbol{\eta})) & \text{if } \mathbf{x} \in X(\mathbb{F}) \text{ and for all } t = 1, \dots, T \\ & \mathbb{E}(f_t(\mathbf{x}, \boldsymbol{\xi}) | \mathcal{F}^t) \leq \mathbf{u}_t \text{ } P\text{-a.s.}, \\ +\infty & \text{else.} \end{cases} \quad (3.1)$$

From the regularity conditions (C1) and (C2) it is immediately clear that  $F$  is convex. Note also that  $\mathcal{P}(0)$  is in fact equivalent to the original problem  $\mathcal{P}$ . In analogy to the well-known set  $X(\mathbb{F})$  of admissible *primal* decision processes, we can now define a set  $Y(\mathbb{F})$  of admissible *dual* decision processes,

$$Y(\mathbb{F}) := \{\mathbf{y} \in \times_{t=1}^T \mathcal{L}^1(\Omega, \mathcal{F}^t, P; \mathbb{R}^{m_t}) \mid \mathbf{y} \text{ is valued in } Y \text{ } P\text{-a.s.}\},$$

where  $Y$  denotes the closed nonnegative orthant of  $\times_{t=1}^T \mathbb{R}^{m_t}$ . By construction, every dual strategy  $\mathbf{y} \in Y(\mathbb{F})$  constitutes an integrable  $\mathbb{F}$ -adapted stochastic process with state space  $Y$ . Next, we let  $Y_0(\mathbb{F})$  be the linear hull of  $Y(\mathbb{F})$  and introduce a real-valued bilinear form  $\langle \cdot, \cdot \rangle$  on  $U_0(\mathbb{F}) \times Y_0(\mathbb{F})$  which is defined via

$$\langle \mathbf{y}, \mathbf{u} \rangle := \mathbb{E}(\mathbf{y} \cdot \mathbf{u}).$$

We assume that  $U_0(\mathbb{F})$  is endowed with the  $Y_0(\mathbb{F})$ -topology; see [7, § V.3]. Thus, the linear functionals  $\mathbf{u} \mapsto \langle \mathbf{y}, \mathbf{u} \rangle$  are continuous, and every continuous linear

functional on  $U_0(\mathbb{F})$  is representable in this way for some  $\mathbf{y} \in Y_0(\mathbb{F})$ . Similarly, we assume  $Y(\mathbb{F})$  to be endowed with the  $U_0(\mathbb{F})$ -topology, implying that the linear functionals  $\mathbf{y} \mapsto \langle \mathbf{y}, \mathbf{u} \rangle$  are continuous, and every continuous linear functional on  $Y_0(\mathbb{F})$  is representable in this way for some  $\mathbf{u} \in U_0(\mathbb{F})$ . With these conventions, the Lagrangian function  $\mathcal{L} : X_0(\mathbb{F}) \times Y_0(\mathbb{F}) \rightarrow [-\infty, +\infty]$  corresponding to our embedding can be defined as in [39, § 4].

$$\mathcal{L}(\mathbf{x}, \mathbf{y}) := \inf_{\mathbf{u} \in U_0(\mathbb{F})} F(\mathbf{x}, \mathbf{u}) + \langle \mathbf{y}, \mathbf{u} \rangle$$

It is easy to calculate that

$$\mathcal{L}(\mathbf{x}, \mathbf{y}) = \begin{cases} \mathbb{E}(L(\mathbf{x}, \mathbf{y}; \boldsymbol{\eta}, \boldsymbol{\xi})) & \text{for } \mathbf{x} \in X(\mathbb{F}), \mathbf{y} \in Y(\mathbb{F}), \\ -\infty & \text{for } \mathbf{x} \in X(\mathbb{F}), \mathbf{y} \notin Y(\mathbb{F}), \\ +\infty & \text{for } \mathbf{x} \notin X(\mathbb{F}), \end{cases}$$

where the Lagrangian density  $L : X \times Y \times \Theta \times \Xi \rightarrow \mathbb{R}$  associated with the problem data is defined as

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

By the basic regularity conditions,  $L$  is convex in  $(x, \xi)$ , concave in  $(y, \eta)$ , and continuous on its domain. An elementary calculation yields

$$F(\mathbf{x}, 0) = \sup_{\mathbf{y} \in Y(\mathbb{F})} \mathcal{L}(\mathbf{x}, \mathbf{y}),$$

which implies that the original stochastic program  $\mathcal{P}$  can be rewritten in terms of the Lagrangian function, i.e.,

$$\text{minimize } \sup_{\mathbf{y} \in Y(\mathbb{F})} \mathcal{L}(\mathbf{x}, \mathbf{y}) \quad \text{over all } \mathbf{x} \in X(\mathbb{F}). \quad (\mathcal{P})$$

By interchanging the operations of minimization and maximization in the above *primal* problem, we can introduce a corresponding *dual* problem.

$$\text{maximize } \inf_{\mathbf{x} \in X(\mathbb{F})} \mathcal{L}(\mathbf{x}, \mathbf{y}) \quad \text{over all } \mathbf{y} \in Y(\mathbb{F}) \quad (\mathcal{D})$$

When constructing computationally tractable approximations for problem  $\mathcal{P}$ , we will heavily exploit the structural properties of the underlying Lagrangian function. In the course of a subsequent convergence analysis, moreover, we will study the solution sets of  $\mathcal{P}$ ,  $\mathcal{D}$ , and of the corresponding approximate problems. The properties of the dual solution set,  $\arg \max \mathcal{D}$ , are intimately tied to certain properties of the optimal value function  $\varphi : U_0(\mathbb{F}) \rightarrow [-\infty, +\infty]$ , which is defined through  $\varphi(\mathbf{u}) := \inf \mathcal{P}(\mathbf{u})$ . Note that  $\varphi$  is monotonically decreasing and convex as follows from a simple projection argument [39, Theorem 1]. Our convergence analysis will further be based on the following elementary result.

**Theorem 3.1** ([39, Theorem 2]). *A pair  $(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \in X_0(\mathbb{F}) \times Y_0(\mathbb{F})$  satisfies the so-called saddle-point condition*

$$\mathcal{L}(\mathbf{x}, \bar{\mathbf{y}}) \geq \mathcal{L}(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \geq \mathcal{L}(\bar{\mathbf{x}}, \mathbf{y}) \quad \text{for all } \mathbf{x} \in X_0(\mathbb{F}), \mathbf{y} \in Y_0(\mathbb{F})$$

*if and only if  $\bar{\mathbf{x}}$  solves  $\mathcal{P}$ ,  $\bar{\mathbf{y}}$  solves  $\mathcal{D}$ , and one has  $\inf \mathcal{P} = \sup \mathcal{D}$ .*

## 4 Discretization

In the remainder of this article we will assume that  $\zeta$  constitutes an admissible process in the following sense.

**Definition 4.1.** *The stochastic process  $\zeta = (\boldsymbol{\eta}, \boldsymbol{\xi})$  with state space  $Z = \Theta \times \Xi$  is called admissible if  $Z$  is compact and*

$$\zeta = (\boldsymbol{\eta}, \boldsymbol{\xi}) = (H^\circ \hat{\boldsymbol{\eta}}, H^c \hat{\boldsymbol{\xi}}), \tag{4.1}$$

where  $\hat{\zeta} := (\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\xi}})$  is a serially independent process with state space  $\hat{Z} := \hat{\Theta} \times \hat{\Xi}$ . The transformations  $H^\circ : \hat{\Theta} \rightarrow \Theta$  and  $H^c : \hat{\Xi} \rightarrow \Xi$  are linear and lower block-triangular with respect to the temporal structure, i.e., the matrix elements of  $H^\circ$  coupling  $\boldsymbol{\eta}_t$  and  $\hat{\boldsymbol{\eta}}_s$  (and those of  $H^c$  coupling  $\boldsymbol{\xi}_t$  and  $\hat{\boldsymbol{\xi}}_s$ ) are zero for all  $s > t$ .<sup>3</sup>

---

<sup>3</sup>In this sense, the transformations  $H^\circ$  and  $H^c$  are ‘non-anticipative’.

It is worthwhile to remark that the admissible processes cover all ARMA processes and are suitable for modelling many real-life stochastic phenomena. Under the assumption that the data process belongs to the class of admissible processes, we now address the approximation of multistage stochastic programs subject to the regularity conditions (C1)–(C4). To this end, we first construct discrete stochastic processes  $\boldsymbol{\eta}^u$  and  $\boldsymbol{\xi}^u$  with state spaces  $\Theta$  and  $\Xi$ , respectively. Thus,  $\boldsymbol{\eta}^u$  has the same range as the data process  $\boldsymbol{\eta}$  of Section 2, and  $\boldsymbol{\xi}^u$  has the same range as  $\boldsymbol{\xi}$ . As for the original data processes, we simplify notation by introducing a combined process  $\boldsymbol{\zeta}^u = (\boldsymbol{\eta}^u, \boldsymbol{\xi}^u)$  with state space  $Z$ . One should think of  $\boldsymbol{\zeta}^u$  as an approximation for  $\boldsymbol{\zeta}$ . Next, we denote by  $\mathbb{F}^u$  the filtration induced by  $\boldsymbol{\zeta}^u$ , 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  $\mathcal{F}^u := \mathcal{F}^{u,T}$ . Having introduced the necessary notation, we can now state the basic properties of  $\boldsymbol{\zeta}^u$ . Concretely speaking, we require the following relations to hold for suitable versions of the conditional expectations, respectively.

$$\mathbb{E}(\boldsymbol{x}|\mathcal{F}) \in X(\mathbb{F}) \text{ for all } \boldsymbol{x} \in X(\mathbb{F}^u) \quad (4.2a)$$

$$\mathbb{E}(\boldsymbol{y}|\mathcal{F}^u) \in Y(\mathbb{F}^u) \text{ for all } \boldsymbol{y} \in Y(\mathbb{F}) \quad (4.2b)$$

$$\mathbb{E}(\boldsymbol{\xi}^u|\mathcal{F}) = \boldsymbol{\xi} \quad (4.2c)$$

$$\mathbb{E}(\boldsymbol{\eta}|\mathcal{F}^u) = \boldsymbol{\eta}^u \quad (4.2d)$$

It is shown in [31, § 4] that for every given tolerance  $\epsilon > 0$  there exists a discrete processes  $\boldsymbol{\zeta}^u$  subject to (4.2a)–(4.2d) with  $\|\boldsymbol{\zeta} - \boldsymbol{\zeta}^u\|_\infty \leq \epsilon$ . Such  $\boldsymbol{\zeta}^u$  can systematically be constructed. The construction in [31] further implies that

$$\mathbb{E}(\boldsymbol{y}|\mathcal{F}) \in Y(\mathbb{F}) \text{ for all } \boldsymbol{y} \in Y(\mathbb{F}^u) \quad (4.2e)$$

$$\mathbb{E}(\boldsymbol{x}|\mathcal{F}^u) \in X(\mathbb{F}^u) \text{ for all } \boldsymbol{x} \in X(\mathbb{F}) \quad (4.2f)$$

for suitable versions of the conditional expectations, respectively.<sup>4</sup>

---

<sup>4</sup>These additional conditions are not needed in [31] but will be necessary for our convergence proof in Section 6. For the processes  $\boldsymbol{\zeta}^u$  constructed in [31, § 4] verification of (4.2e) and (4.2f) is straightforward and completely analogous to verification of (4.2a) and (4.2b), respectively.

Next, introduce another discrete process  $\zeta^l = (\eta^l, \xi^l)$  such that  $\eta^l$  and  $\xi^l$  are valued in  $\Theta$  and  $\Xi$ , respectively. Again,  $\zeta^l$  is supposed to approximate the original data process  $\zeta$ . The induced filtration  $\mathbb{F}^l$  is constructed as usual, i.e.,  $\mathbb{F}^l := \{\mathcal{F}^{l,t}\}_{t=1}^T$  where  $\mathcal{F}^{l,t} := \sigma(\zeta^{l,t})$ , and we use the convention  $\mathcal{F}^l := \mathcal{F}^{l,T}$ . Using these notational conventions, we require  $\zeta^u$  to satisfy the following conditions for suitable versions of the conditional expectations, respectively.

$$\mathbb{E}(\mathbf{x}|\mathcal{F}^l) \in X(\mathbb{F}^l) \text{ for all } \mathbf{x} \in X(\mathbb{F}) \quad (4.3a)$$

$$\mathbb{E}(\mathbf{y}|\mathcal{F}) \in Y(\mathbb{F}) \text{ for all } \mathbf{y} \in Y(\mathbb{F}^l) \quad (4.3b)$$

$$\mathbb{E}(\xi|\mathcal{F}^l) = \xi^l \quad (4.3c)$$

$$\mathbb{E}(\eta^l|\mathcal{F}) = \eta \quad (4.3d)$$

From the argumentation in [31] it follows that for every given tolerance  $\epsilon > 0$  we can construct a discrete processes  $\zeta^l$  subject to (4.3a)–(4.3d) with  $\|\zeta - \zeta^l\|_\infty \leq \epsilon$ . In addition, this  $\zeta^l$  may be assumed to satisfy the conditions

$$\mathbb{E}(\mathbf{y}|\mathcal{F}^l) \in Y(\mathbb{F}^l) \text{ for all } \mathbf{y} \in Y(\mathbb{F}) \quad (4.3e)$$

$$\mathbb{E}(\mathbf{x}|\mathcal{F}) \in X(\mathbb{F}) \text{ for all } \mathbf{x} \in X(\mathbb{F}^l) \quad (4.3f)$$

for suitable versions of the conditional expectations, respectively. It should be pointed out that the construction of  $\zeta^l$  depends in no way on the construction of  $\zeta^u$  and vice versa. In mathematical terms, we may thus assume  $\zeta^u$  and  $\zeta^l$  to be conditionally independent given  $\zeta$ . This assumption will be needed for our convergence proof in Section 6.

If we replace the true data process  $\zeta$  by  $\zeta^u$  and the true filtration  $\mathbb{F}$  by  $\mathbb{F}^u$  in  $\mathcal{P}$ , then we obtain an approximate optimization problem denoted  $\mathcal{P}^u$ . Another approximate problem  $\mathcal{P}^l$  is obtained by substituting  $\zeta^l$  for  $\zeta$  and  $\mathbb{F}^l$  for  $\mathbb{F}$ . Note that replacing the filtrations has a primal and a dual effect, that is, after substitution the primal decisions as well as the constraints are adapted to the approximate filtration. With these conventions, we are now ready to generalize a main result of [31].

**Theorem 4.2.** *Assume the conditions (C1), (C2), and (C3) hold. If the data process  $\zeta$  and its approximations  $\zeta^u$  and  $\zeta^l$  satisfy (4.2) and (4.3), respectively, then  $\inf \mathcal{P}^l \leq \inf \mathcal{P} \leq \inf \mathcal{P}^u$ .*

*Proof.* Although here we work in a more general setting with expected value constraints, the proofs of the corresponding Theorems 1 and 2 in [31] remain valid without any change.  $\square$

As  $\zeta^u$  represents a finitely supported stochastic process, any  $\mathbb{F}^u$ -adapted primal or dual strategy is finitely supported, as well. Thus,  $\mathcal{P}^u$  essentially constitutes a convex program over some finite-dimensional Euclidean space with a finite number of constraints, and it principally allows for numerical solution. For similar reasons,  $\mathcal{P}^l$  is computationally tractable, too. These observations together with Theorem 4.2 show that, given enough memory and CPU power, we can numerically calculate upper and lower bounds on  $\inf \mathcal{P}$ .

## 5 A Posteriori Error Estimate

In this section we will assume that suitable approximate processes  $\zeta^u$  and  $\zeta^l$  subject to the conditions (4.2) and (4.3), respectively, have been chosen. We will also assume that these processes are finitely supported, such that the upper and lower approximate problems  $\mathcal{P}^u$  and  $\mathcal{P}^l$  are indeed equivalent to finite-dimensional convex mathematical programs.

**Theorem 5.1.** *Assume the conditions (C1), (C2), and (C3) hold. If the approximate problem  $\mathcal{P}^u$  has a finite infimum and is solved by some strategy  $\mathbf{x}^u$ , then  $\hat{\mathbf{x}} := \mathbb{E}(\mathbf{x}^u | \mathcal{F})$  is feasible in  $\mathcal{P}$  and*

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

*Proof.* The first inequality follows immediately from Theorem 4.2. By using the Lagrangian reformulation of  $\mathcal{P}$ , the saddle structure of the Lagrangian density  $L$ ,

and the basic properties of the approximate process  $\zeta^u$ , we further obtain

$$\begin{aligned}
\inf \mathcal{P} &\leq \sup_{\mathbf{y} \in Y(\mathbb{F})} \mathbb{E} (L(\mathbb{E}(\mathbf{x}^u | \mathcal{F}), \mathbf{y}; \boldsymbol{\eta}, \mathbb{E}(\boldsymbol{\xi}^u | \mathcal{F}))) \\
&\leq \sup_{\mathbf{y} \in Y(\mathbb{F})} \mathbb{E} (L(\mathbf{x}^u, \mathbf{y}; \boldsymbol{\eta}, \boldsymbol{\xi}^u)) \\
&\leq \sup_{\mathbf{y} \in Y(\mathbb{F})} \mathbb{E} (L(\mathbf{x}^u, \mathbb{E}(\mathbf{y} | \mathcal{F}^u); \mathbb{E}(\boldsymbol{\eta} | \mathcal{F}^u), \boldsymbol{\xi}^u)) \tag{5.4} \\
&\leq \sup_{\mathbf{y} \in Y(\mathbb{F}^u)} \mathbb{E} (L(\mathbf{x}^u, \mathbf{y}; \boldsymbol{\eta}^u, \boldsymbol{\xi}^u)) \\
&= \inf_{\mathbf{x} \in X(\mathbb{F}^u)} \sup_{\mathbf{y} \in Y(\mathbb{F}^u)} \mathbb{E} (L(\mathbf{x}, \mathbf{y}; \boldsymbol{\eta}^u, \boldsymbol{\xi}^u)) .
\end{aligned}$$

The first inequality in (5.4) follows from (4.2a) and (4.2c). Next, we use the conditional version of Jensen's inequality for moving the conditional expectations out of the Lagrangian density. This is allowed by convexity of the Lagrangian density in the first and the fourth arguments, and since  $\mathbf{y}$  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 density in the second and the third arguments together with the  $\mathcal{F}^u$ -measurability of  $\mathbf{x}^u$  and  $\boldsymbol{\xi}^u$ . Finally, the fourth inequality holds by the assumptions (4.2b) and (4.2d). It entails relaxation of the dual feasible set from those decisions which are representable as the conditional expectation (given  $\mathcal{F}^u$ ) of some  $\mathbf{y} \in Y(\mathbb{F})$  to all decisions in  $Y(\mathbb{F}^u)$ .

The last line of (5.4) corresponds to the optimal value of the upper approximate problem and is finite by assumption. This implies that the supremum over  $Y(\mathbb{F})$  in the first line of (5.4) is attained by  $\mathbf{y} = \mathbf{0}$ , and

$$\sup_{\mathbf{y} \in Y(\mathbb{F})} \mathbb{E} (L(\mathbb{E}(\mathbf{x}^u | \mathcal{F}), \mathbf{y}; \boldsymbol{\eta}, \mathbb{E}(\boldsymbol{\xi}^u | \mathcal{F}))) = \mathbb{E}(c(\hat{\mathbf{x}}, \boldsymbol{\eta})).$$

We can therefore conclude that  $\hat{\mathbf{x}}$  is feasible in  $\mathcal{P}$ , and the corresponding objective value  $\mathbb{E}(c(\hat{\mathbf{x}}, \boldsymbol{\eta}))$  satisfies the postulated inequalities.  $\square$

Whereas Theorem 4.2 provides an a priori estimate of the minimal cost achievable in theory under the (unknown) true optimal policy, Theorem 5.1 suggests a computationally accessible near-optimal policy and provides an a posteriori estimate of the minimal cost achievable in practice. In fact, if  $\mathbf{x}^\star$  denotes an optimal

solution to problem  $\mathcal{P}$ , we may conclude from Theorem 5.1 that

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

Thus, the error of implementing the suboptimal policy  $\hat{\mathbf{x}}$  instead of some true optimal policy  $\mathbf{x}^*$  is bounded by  $\inf \mathcal{P}^u - \inf \mathcal{P}^l$ . This difference is computationally accessible and estimates the approximation error as defined by Pflug [35, Section 2].

Evaluation of  $\hat{\mathbf{x}}$  for some given realization of  $\boldsymbol{\zeta}$  reduces to the evaluation of a weighted sum over some subset of the (finite) support of  $\mathbf{x}^u$ .<sup>5</sup> Of particular interest is the recommended first stage decision  $\hat{\mathbf{x}}_1$ . If  $\boldsymbol{\zeta}_1$  constitutes a degenerate deterministic random vector, as is usually assumed in practice, it is reasonable to set  $\boldsymbol{\zeta}_1^u = \boldsymbol{\zeta}_1$ . This choice is consistent with the conditions (4.2) and the construction of  $\boldsymbol{\zeta}^u$  suggested in [31, Section 4]. Then, it immediately follows that the recommended first stage decision  $\hat{\mathbf{x}}_1$  coincides with the optimal first stage decision  $\mathbf{x}_1^u$  of the approximate problem  $\mathcal{P}^u$ , that is, no conditional expectation needs to be evaluated.

A major drawback of most solution schemes for stochastic programs based on discretization is that only the first stage decisions can be implemented. The recourse decisions associated with later stages are given only for few discrete scenarios and are therefore not implementable (almost surely). The discretization method presented here does not suffer from this shortcoming since an implementable strategy  $\hat{\mathbf{x}}$  can be derived from any solution  $\mathbf{x}^u$  of the upper approximate problem. As shown above,  $\hat{\mathbf{x}}$  is implementable for every possible realization of  $\boldsymbol{\zeta}$ .

Finally, it is worthwhile to remark that  $\hat{\mathbf{y}} = E(\mathbf{y}^l | \mathcal{F})$  is feasible in  $\mathcal{D}$  if  $\mathbf{y}^l$  solves the dual  $\mathcal{D}^l$  corresponding to the lower approximate problem  $\mathcal{P}^l$ . The dual

---

<sup>5</sup>As only the marginal distribution of  $\boldsymbol{\zeta}^u$  is relevant for the solution of problem  $\mathcal{P}^u$ , the optimal policy  $\mathbf{x}^u$  can be chosen conditionally independent of  $\boldsymbol{\zeta}$  given  $\boldsymbol{\zeta}^u$ . Under this premise, calculation of the expectation  $E(\mathbf{x}^u | \mathcal{F})$  is straightforward.



policy  $\hat{\boldsymbol{y}}$  is computationally accessible, and its objective value lies between  $\inf \mathcal{P}^l$  and  $\inf \mathcal{P}^u$ . This result can be proved in a similar same manner as Theorem 5.1. Details are omitted for brevity of exposition.

## 6 Convergence

The asymptotic consistency of MSP approximations has been the subject of several investigations. Pennanen [33] establishes very general conditions which guarantee that a sequence of finite-dimensional problems epi-converges to the original infinite-dimensional MSP. More quantitative results are reported, among others, by Heitsch et al. [20] for probability metric based approximations and by Frauendorfer [16] and Casey and Sen [3] for bounding approximations. Neither of these recent approaches is directly applicable in the present situation since we consider problems with induced constraints and with expected value constraints. A few nonrestrictive assumptions, however, enable us to show asymptotic consistency of the approximation scheme under consideration. Our proof is self-contained and provides additional insights.

For each  $J \in \mathbb{N}$  introduce a pair of discrete stochastic processes  $\zeta_J^u$  and  $\zeta_J^l$  with induced filtrations  $\mathbb{F}_J^u = \{\mathcal{F}_J^{u,t}\}_{t=1}^T$  and  $\mathbb{F}_J^l = \{\mathcal{F}_J^{l,t}\}_{t=1}^T$ , respectively. As usual, use the convention  $\mathcal{F}_J^u := \mathcal{F}_J^{u,T}$  and  $\mathcal{F}_J^l := \mathcal{F}_J^{l,T}$  in order to simplify notation. Assume that  $\zeta$  and  $\zeta_J^u$  satisfy the conditions (4.2),  $\zeta$  and  $\zeta_J^l$  satisfy (4.3), and  $\zeta_J^u$  is conditionally independent of  $\zeta_J^l$  given  $\zeta$  for all  $J \in \mathbb{N}$ . Moreover, assume that

$$\lim_{J \rightarrow \infty} \|\zeta_J^u - \zeta\|_\infty = 0 \quad \text{and} \quad \lim_{J \rightarrow \infty} \|\zeta_J^l - \zeta\|_\infty = 0. \quad (6.5)$$

By the argumentation in Section 4 and in [31, § 4], such sequences of discrete stochastic processes exist and can systematically be constructed whenever  $\zeta$  constitutes an admissible process in the sense of Definition 4.1. Unless otherwise noted, the above assumptions are supposed to hold throughout the rest of this article.

If we replace the true data process  $\zeta$  by  $\zeta_J^u$  and the true filtration  $\mathbb{F}$  by  $\mathbb{F}_J^u$  in the derivations of Section 3, then, for each  $J \in \mathbb{N}$ , we obtain a parametric family of approximate optimization problems denoted by  $\mathcal{P}_J^u(\mathbf{u})$ , where the perturbation parameter  $\mathbf{u}$  ranges over  $U_0(\mathbb{F}_J^u)$ . Another parametric family of approximate optimization problems  $\mathcal{P}_J^l(\mathbf{u})$ ,  $\mathbf{u} \in U_0(\mathbb{F}_J^l)$ , is obtained by substituting  $\zeta_J^l$  for  $\zeta$  and  $\mathbb{F}_J^l$  for  $\mathbb{F}$ . Note that replacing the filtration in (3.1) affects the measurability properties of the admissible  $\mathbf{x}$  as well as the conditional expectations in the explicit constraints. For each  $J \in \mathbb{N}$ , we introduce the shorthand notation  $\mathcal{P}_J^u$  and  $\mathcal{P}_J^l$  for the problems  $\mathcal{P}_J^u(0)$  and  $\mathcal{P}_J^l(0)$ , respectively. Moreover, we denote by  $\mathcal{D}_J^u$  the maximization problem dual to  $\mathcal{P}_J^u$ , while  $\mathcal{D}_J^l$  refers to the maximization problem dual to  $\mathcal{P}_J^l$ ; see Section 3 for the definition of the dual problems.

We are now prepared to show that the bounds of Theorem 4.2 (associated with the discretizations  $\zeta_J^u$  and  $\zeta_J^l$ ,  $J \in \mathbb{N}$ ) converge to the optimal value of the original problem as  $J$  tends to infinity. The proof is based on five preliminary lemmas. The first lemma establishes a relation between strategies that are feasible for the upper and lower approximate problems, respectively, while the second lemma provides a general result about conditional expectations. This result is used in the third lemma to prove that the approximate problems  $\mathcal{P}_J^u$  and  $\mathcal{P}_J^l$  are strictly feasible for  $J$  large enough. Primal and dual solvability of these finite-dimensional problems is established in the fourth lemma. Finally, dual solvability and strict primal feasibility are used to prove that the dual solutions of the approximate problems are uniformly bounded for large  $J$ . Uniform boundedness of the (primal *and* dual) approximate solutions, in turn, are crucial ingredients for our convergence proof presented in Theorem 6.6.

**Lemma 6.1.** *The following relations hold for suitable versions of the conditional expectations.*

- (i)  $E(\mathbf{x}|\mathcal{F}_J^u) \in X(\mathbb{F}_J^u)$  for all  $\mathbf{x} \in X(\mathbb{F}_J^l)$
- (ii)  $E(\mathbf{y}|\mathcal{F}_J^l) \in X(\mathbb{F}_J^l)$  for all  $\mathbf{y} \in Y(\mathbb{F}_J^u)$

*Proof.* First, we select an arbitrary primal policy  $\mathbf{x} \in X(\mathbb{F}_J^l)$ . By conditional independence of  $\zeta_J^u$  and  $\zeta_J^l$  given  $\zeta$  and by (4.3f), we have

$$\mathbb{E}(\mathbf{x}|\zeta, \zeta_J^u) = \mathbb{E}(\mathbf{x}|\zeta) =: \mathbf{x}' \in X(\mathbb{F})$$

for suitable versions of the conditional expectations. The law of iterated conditional expectations and (4.3f) then imply that

$$\mathbb{E}(\mathbf{x}|\zeta_J^u) = \mathbb{E}(\mathbb{E}(\mathbf{x}|\zeta, \zeta_J^u)|\zeta_J^u) = \mathbb{E}(\mathbf{x}'|\zeta_J^u) \in X(\mathbb{F}_J^u)$$

for suitable versions of the conditional expectations. This observation establishes assertion (i). The proof of assertion (ii) is analogous.  $\square$

**Lemma 6.2.** *If  $\zeta$  and  $\zeta'$  are random vectors in  $\mathcal{L}^p(\Omega, \Sigma, P; Z)$ , then*

$$\|\mathbb{E}(\zeta|\zeta') - \zeta'\|_p \leq \|\zeta - \zeta'\|_p \quad \text{for all } 1 \leq p \leq \infty.$$

*Proof.* Assume first that  $p < \infty$ . By using the conditional Jensen inequality and the law of iterated conditional expectations, we find

$$\begin{aligned} \|\mathbb{E}(\zeta|\zeta') - \zeta'\|_p^p &= \|\mathbb{E}(\zeta - \zeta'|\zeta')\|_p^p \\ &= \mathbb{E}(|\mathbb{E}(\zeta - \zeta'|\zeta')|^p) \\ &\leq \mathbb{E}(\mathbb{E}(|\zeta - \zeta'|^p|\zeta')) = \|\zeta - \zeta'\|_p^p. \end{aligned}$$

Thus, the claim is established for  $p < \infty$ . For  $p = \infty$  the statement follows from the above argument and the well-known relation  $\lim_{p \rightarrow \infty} \|\hat{\zeta}\|_p = \|\hat{\zeta}\|_\infty$ , where  $\hat{\zeta}$  is any random vector in  $\mathcal{L}^\infty(\Omega, \Sigma, P; Z)$ .  $\square$

In the next lemma we use this elementary result to prove strict feasibility of the approximate problems. For the sake of simple notation, let  $\mathbf{1} := (\mathbf{1}_1, \dots, \mathbf{1}_T)$  be the element of  $\times_{t=1}^T \mathcal{L}^\infty(\Omega, \Sigma, P; \mathbb{R}^{m_t})$  whose components are all 1's.

**Lemma 6.3.** *Assume the conditions (C1)–(C4) hold, and let  $\mathbf{x}_s$  be strictly feasible in  $\mathcal{P}$ . Then there exists  $\varepsilon > 0$  such that  $\mathbb{E}(\mathbf{x}_s|\mathcal{F}_J^u)$  is feasible in  $\mathcal{P}_J^u(-\varepsilon\mathbf{1})$  for all  $J$  large enough, and  $\mathbb{E}(\mathbf{x}_s|\mathcal{F}_J^l)$  is feasible in  $\mathcal{P}_J^l(-\varepsilon\mathbf{1})$  for all  $J \in \mathbb{N}$ .*

*Proof.* By strict feasibility, there exists  $\varepsilon > 0$  with

$$\mathbb{E}(f_t(\mathbf{x}_s, \boldsymbol{\xi}) | \mathcal{F}^t) \leq -2\varepsilon \mathbf{1}_t \quad P\text{-a.s.} \quad (6.6)$$

for all  $t = 1, \dots, T$ . For the further argumentation we fix some  $t$  and define

$$\hat{Y}_t(\mathbb{G}) := \{\mathbf{y}_t \in Y_t(\mathbb{G}) \mid \mathbb{E}(\mathbf{1}_t \cdot \mathbf{y}_t) = 1\},$$

where  $\mathbb{G}$  can be either of the filtrations  $\mathbb{F}$  or  $\mathbb{F}_J^u$  for  $J \in \mathbb{N}$ , and  $Y_t(\mathbb{G})$  is the projection of  $Y(\mathbb{G})$  to the space of dual stage  $t$  decisions. Then, the (vectorial) inequality (6.6) involving a conditional expectation is equivalent to a family of (scalar) inequalities involving only unconditional expectations, i.e.,

$$\mathbb{E}(\mathbf{y}_t \cdot f_t(\mathbf{x}_s, \boldsymbol{\xi})) \leq -2\varepsilon \quad \forall \mathbf{y}_t \in \hat{Y}_t(\mathbb{F}). \quad (6.7)$$

From the relation (4.2e) it is easily seen that  $\mathbb{E}(\mathbf{y}_t | \mathcal{F})$  is an element of  $\hat{Y}_t(\mathbb{F})$  whenever  $\mathbf{y}_t$  is an element of  $\hat{Y}_t(\mathbb{F}_J^u)$ . Thus, (6.7) implies

$$\mathbb{E}(\mathbb{E}(\mathbf{y}_t | \mathcal{F}) \cdot f_t(\mathbf{x}_s, \boldsymbol{\xi})) \leq -2\varepsilon \quad \forall \mathbf{y}_t \in \hat{Y}_t(\mathbb{F}_J^u). \quad (6.8)$$

The law of iterated conditional expectations and the  $\mathcal{F}$ -measurability of  $\mathbf{x}_s$  and  $\boldsymbol{\xi}$  enable us to eliminate the conditional expectation with respect to  $\mathcal{F}$ . The system of inequalities (6.8) is therefore equivalent to

$$\mathbb{E}(\mathbf{y}_t \cdot f_t(\mathbf{x}_s, \boldsymbol{\xi})) \leq -2\varepsilon \quad \forall \mathbf{y}_t \in \hat{Y}_t(\mathbb{F}_J^u).$$

Next, the conditional Jensen inequality, convexity of  $f_t$ , and nonnegativity of  $\mathbf{y}_t$  can be used to conclude that

$$\mathbb{E}(\mathbf{y}_t \cdot f_t(\mathbb{E}(\mathbf{x}_s | \mathcal{F}_J^u), \mathbb{E}(\boldsymbol{\xi} | \mathcal{F}_J^u))) \leq -2\varepsilon \quad \forall \mathbf{y}_t \in \hat{Y}_t(\mathbb{F}_J^u). \quad (6.9)$$

Notice that  $\|\mathbb{E}(\boldsymbol{\xi} | \mathcal{F}_J^u) - \boldsymbol{\xi}_J^u\|_\infty$  converges to zero as  $J$  tends to infinity. This follows from Lemma 6.2 and convergence of  $\|\boldsymbol{\xi}_J^u - \boldsymbol{\xi}\|_\infty$  to zero. Thus, by uniform continuity of  $f_t$  on its compact domain, there is some  $J_0 \in \mathbb{N}$  with

$$f_t(\mathbb{E}(\mathbf{x}_s | \mathcal{F}_J^u), \mathbb{E}(\boldsymbol{\xi} | \mathcal{F}_J^u)) \leq f_t(\mathbb{E}(\mathbf{x}_s | \mathcal{F}_J^u), \boldsymbol{\xi}_J^u) + \varepsilon \quad P\text{-a.s. for } J \geq J_0.$$

Plugging this estimate into (6.9) leads to the conclusion that

$$\mathbb{E} \left( f_t(\mathbb{E}(\mathbf{x}_s | \mathcal{F}_J^u), \boldsymbol{\xi}_J^u) | \mathcal{F}_J^{u,t} \right) \leq -\varepsilon \mathbf{1}_t \quad P\text{-a.s. for } J \geq J_0.$$

As  $t$  was arbitrary, we find that  $\mathbb{E}(\mathbf{x}_s | \mathcal{F}_J^u)$  is feasible in  $\mathcal{P}_J^u(-\varepsilon \mathbf{1})$  for  $J \geq J_0$ . To prove that  $\mathbb{E}(\mathbf{x}_s | \mathcal{F}_J^l)$  is feasible in  $\mathcal{P}_J^l(-\varepsilon \mathbf{1})$  for all  $J \in \mathbb{N}$ , we use an analogous argument. However, the proof is simplified by the fact that  $\mathbb{E}(\boldsymbol{\xi} | \mathcal{F}_J^l)$  exactly equals  $\boldsymbol{\xi}_J^l$  for each  $J \in \mathbb{N}$ ; it not only comes close to  $\boldsymbol{\xi}_J^l$  for  $J$  large.  $\square$

The discrete approximate problems are not only feasible, but also solvable for  $J$  large enough. This is a consequence of their finite dimensionality.

**Lemma 6.4.** *If the conditions (C1)–(C4) hold, then  $\mathcal{P}_J^u$ ,  $\mathcal{D}_J^u$ ,  $\mathcal{P}_J^l$ , and  $\mathcal{D}_J^l$  are solvable, while  $\min \mathcal{P}_J^u = \max \mathcal{D}_J^u$  and  $\min \mathcal{P}_J^l = \max \mathcal{D}_J^l$  for all  $J$  large enough.*

*Proof.* Solvability of  $\mathcal{P}_J^u$  and  $\mathcal{D}_J^u$  relies on the fact that the approximate process  $\zeta_J^u$  is finitely supported. Consequently, any  $\mathbb{F}_J^u$ -adapted primal or dual strategy is finitely supported, too. The conditions (C1)–(C4) thus imply via Lemma 6.3 that, for  $J$  large enough,  $\mathcal{P}_J^u$  essentially constitutes a strictly feasible convex program of the type [39, Example 1]. It has finitely many decision variables and constraints, a continuous objective function, and a compact feasible set. Primal solvability follows now from the Weierstrass maximum theorem, while dual solvability follows from Theorem 17 and Example 1” in [39]. Furthermore, [39, Theorem 17] implies strong duality, that is,  $\min \mathcal{P}_J^u = \max \mathcal{D}_J^u$ . Solvability of  $\mathcal{P}_J^l$  and  $\mathcal{D}_J^l$  and the relation  $\min \mathcal{P}_J^l = \max \mathcal{D}_J^l$  are proved in exactly the same way.  $\square$

It is worthwhile to remark that dual solvability could be lost in an infinite-dimensional setting.<sup>6</sup> This might happen if the stochastic program at hand fails to have relatively complete recourse; see e.g. [43, § 5]. In the remainder of this section, we fix any solutions of  $\mathcal{P}_J^u$  and  $\mathcal{P}_J^l$ , which will be denoted by  $\mathbf{x}_J^u$  and  $\mathbf{x}_J^l$ , respectively. Furthermore, we fix any solutions of  $\mathcal{D}_J^u$  and  $\mathcal{D}_J^l$ , which will

---

<sup>6</sup>Imagine  $\zeta_J^u$  or  $\zeta_J^l$  to be continuously distributed for some  $J$ .

be denoted by  $\mathbf{y}_J^u$  and  $\mathbf{y}_J^l$ , respectively. For notational convenience, we also introduce the optimal value functions  $\varphi_J^u(\mathbf{u}) := \inf \mathcal{P}_J^u(\mathbf{u})$  defined on  $U_0(\mathbb{F}_J^u)$  and  $\varphi_J^l(\mathbf{u}) := \inf \mathcal{P}_J^l(\mathbf{u})$  defined on  $U_0(\mathbb{F}_J^l)$ . Remember that  $\varphi$  stands for the optimal value function associated with the original problem and is defined on  $U_0(\mathbb{F})$ . Exploiting the specific properties of the original and approximate value functions, we can now prove uniform boundedness of the approximate dual solutions.

**Lemma 6.5.** *If the conditions (C1)–(C4) hold, then  $\|\mathbf{y}_J^u\|_1$  is uniformly bounded for  $J$  large enough.*

*Proof.* Let  $\mathbf{x}_s$  be a strictly feasible policy for  $\mathcal{P}$ . By Lemma 6.3 there exist  $\varepsilon > 0$  and  $J_1 \in \mathbb{N}$  such that  $\mathbf{E}(\mathbf{x}_s | \mathcal{F}_J^u)$  is feasible in  $\mathcal{P}_J^u(-\varepsilon \mathbf{1})$  for all  $J \geq J_1$ . Notice that  $\|\mathbf{E}(\boldsymbol{\eta}_J^u | \mathcal{F}) - \boldsymbol{\eta}\|_\infty$  converges to zero as  $J$  tends to infinity. This follows from Lemma 6.2 and convergence of  $\|\boldsymbol{\eta}_J^u - \boldsymbol{\eta}\|_\infty$  to zero. Thus, by uniform continuity of  $c$  on its compact domain, there is some  $J_2 \in \mathbb{N}$  with

$$c(\mathbf{x}_s, \mathbf{E}(\boldsymbol{\eta} | \mathcal{F}_J^u)) \leq c(\mathbf{x}_s, \boldsymbol{\eta}) + 1 \quad P\text{-a.s. for } J \geq J_2. \quad (6.10)$$

The solution set of problem  $\mathcal{D}_J^u$  coincides with  $-\partial\varphi_J^u(0)$ , that is, the negative of the subdifferential of  $\varphi_J^u$  evaluated at the origin. Thus, we have

$$\varphi_J^u(\mathbf{u}) - \varphi_J^u(0) \geq -\langle \mathbf{y}_J^u, \mathbf{u} \rangle \quad \text{for all } \mathbf{u} \in U_0(\mathbb{F}_J^u).$$

Taking the supremum over all  $\mathbf{u} \in U_0(\mathbb{F}_J^u)$  with  $\|\mathbf{u}\|_\infty = \varepsilon$  on both sides and exploiting monotonicity of  $\varphi_J^u$ , we obtain

$$\varphi_J^u(-\varepsilon \mathbf{1}) - \varphi_J^u(0) \geq \varepsilon \|\mathbf{y}_J^u\|_1.$$

With this estimate in mind, we may conclude that for all  $J \geq \max\{J_1, J_2\}$

$$\begin{aligned} \varepsilon \|\mathbf{y}_J^u\|_1 &\leq \varphi_J^u(-\varepsilon \mathbf{1}) - \varphi_J^u(0) \\ &\leq \varphi_J^u(-\varepsilon \mathbf{1}) - \varphi(0) \\ &\leq \mathbf{E}(c(\mathbf{E}(\mathbf{x}_s | \mathcal{F}_J^u), \boldsymbol{\eta}_J^u)) - \varphi(0) \\ &\leq \mathbf{E}(c(\mathbf{x}_s, \mathbf{E}(\boldsymbol{\eta}_J^u | \mathcal{F}))) - \varphi(0) \\ &\leq \mathbf{E}(c(\mathbf{x}_s, \boldsymbol{\eta})) - \varphi(0) + 1. \end{aligned} \quad (6.11)$$

The second inequality is a direct consequence of Theorem 4.2, and the third inequality uses feasibility of  $\mathbf{E}(\mathbf{x}_s | \mathcal{F}_J^u)$  in problem  $\mathcal{P}_J^u(-\varepsilon \mathbf{1})$  for  $J \geq J_1$ . In the fourth line, we use  $\mathcal{F}_J^u$ -measurability of  $\boldsymbol{\eta}_J^u$  and the conditional version of Jensen's inequality for moving the conditional expectation with respect to  $\mathcal{F}_J^u$  out of the cost function.<sup>7</sup> Subsequently, the conditional Jensen inequality is applied once again. This time,  $\mathcal{F}$ -measurability of  $\mathbf{x}_s$  ensures that the conditional expectation with respect to  $\mathcal{F}$  can be brought inside the cost function. The fifth inequality in (6.11) follows directly from the estimate (6.10).

The last line of (6.11) is manifestly independent of  $J$ . Moreover, it is finite. Divided by  $\varepsilon > 0$  it thus provides a uniform upper bound on  $\|\mathbf{y}_J^u\|_1$  over all  $J$  large enough. This observation completes the proof.  $\square$

After having proved the above preliminary lemmas, we are ready to establish the main result of this article.

**Theorem 6.6.** *If the conditions (C1)–(C4) hold, then for  $J \rightarrow \infty$  we have*

$$\inf \mathcal{P}_J^u \rightarrow \inf \mathcal{P} \quad \text{and} \quad \inf \mathcal{P}_J^l \rightarrow \inf \mathcal{P}.$$

*Proof.* The claim can be established by using our previous results. In particular, uniform boundedness of the dual solutions corresponding to the discretized problems is crucial. We start with the following chain of inequalities.

$$\begin{aligned} 0 &\leq \inf \mathcal{P}_J^u - \inf \mathcal{P}_J^l \\ &= \mathbf{E} (L(\mathbf{x}_J^u, \mathbf{y}_J^u; \boldsymbol{\eta}_J^u, \boldsymbol{\xi}_J^u) - L(\mathbf{x}_J^l, \mathbf{y}_J^l; \boldsymbol{\eta}_J^l, \boldsymbol{\xi}_J^l)) \\ &\leq \mathbf{E} (L(\mathbf{E}(\mathbf{x}_J^l | \mathcal{F}_J^u), \mathbf{y}_J^u; \boldsymbol{\eta}_J^u, \boldsymbol{\xi}_J^u) - L(\mathbf{x}_J^l, \mathbf{E}(\mathbf{y}_J^u | \mathcal{F}_J^l); \boldsymbol{\eta}_J^l, \boldsymbol{\xi}_J^l)) \quad (6.12) \\ &\leq \mathbf{E} (L(\mathbf{x}_J^l, \mathbf{y}_J^u; \boldsymbol{\eta}_J^u, \boldsymbol{\xi}_J^u) - L(\mathbf{x}_J^l, \mathbf{y}_J^l; \boldsymbol{\eta}_J^l, \boldsymbol{\xi}_J^l)) \\ &\leq \|c(\mathbf{x}_J^l, \boldsymbol{\eta}_J^u) - c(\mathbf{x}_J^l, \boldsymbol{\eta}_J^l)\|_\infty + \sum_{t=1}^T \|\mathbf{y}_J^u\|_1 \|f_t(\mathbf{x}_J^l, \boldsymbol{\xi}_J^u) - f_t(\mathbf{x}_J^l, \boldsymbol{\xi}_J^l)\|_\infty \end{aligned}$$

The first inequality is a direct consequence of Theorem 4.2, while the equality in the second line uses Theorem 3.1 and Lemma 6.4. The second inequality (in

---

<sup>7</sup>Recall that  $c$  is convex in its first and concave in its second argument.

the third line) follows, again, from Theorem 3.1 and the fact that  $E(\mathbf{x}_J^l | \mathcal{F}_J^u)$  is contained in  $X(\mathbb{F}_J^u)$ , while  $E(\mathbf{y}_J^u | \mathcal{F}_J^l)$  is contained in  $Y(\mathbb{F}_J^l)$ , see Lemma 6.1. The third inequality is due to the conditional Jensen inequality, which applies since the Lagrangian density is convex in the primal and concave in the dual decisions, and the last inequality in (6.12) is based on the triangle and Hölder inequalities. Since  $c$  is uniformly continuous on its (compact) domain, and since  $\|\boldsymbol{\eta}_J^u - \boldsymbol{\eta}_J^l\|_\infty$  converges to zero as  $J$  tends to infinity, we may conclude that

$$\lim_{J \rightarrow \infty} \|c(\mathbf{x}_J^l, \boldsymbol{\eta}_J^u) - c(\mathbf{x}_J^l, \boldsymbol{\eta}_J^l)\|_\infty = 0.$$

A similar reasoning shows that

$$\lim_{J \rightarrow \infty} \|f_t(\mathbf{x}_J^l, \boldsymbol{\xi}_J^u) - f_t(\mathbf{x}_J^l, \boldsymbol{\xi}_J^l)\|_\infty = 0 \quad \text{for all } t = 1, \dots, T.$$

By (6.12) and uniform boundedness of  $\|\mathbf{y}_J^u\|_1$  for  $J$  large enough, see Lemma 6.5, we thus have

$$\max\{\inf \mathcal{P}_J^u - \inf \mathcal{P}, \inf \mathcal{P} - \inf \mathcal{P}_J^l\} \leq \inf \mathcal{P}_J^u - \inf \mathcal{P}_J^l \rightarrow 0 \quad \text{for } J \rightarrow \infty.$$

This observation completes the proof.  $\square$

## 7 Example

We formulate a simple stochastic program to determine the cost-minimal production schedule of a hydropower plant over a two-period planning horizon. The plant comprises a single reservoir. At any time, water can be released through a turbine to produce electric energy. For simplicity, we assume that reservoir content is measured in energy units. The initial content is set to 1. In period  $t$ , the plant produces  $\mathbf{x}_t$  units of energy, which are sold at price  $\boldsymbol{\eta}_t$  per unit. Meanwhile, the reservoir content increases by  $\boldsymbol{\xi}_t$  units due to natural water inflows. The reservoir thus contains  $1 - \mathbf{x}_1 + \boldsymbol{\xi}_1$  units after the first and  $1 - \mathbf{x}_1 - \mathbf{x}_2 + \boldsymbol{\xi}_1 + \boldsymbol{\xi}_2$  units after the second period. We require that the reservoir volume never exceeds 2.



Moreover, it must not drop below 0 (0.8) at the end of the first (second) period. The production decision  $\mathbf{x}_t$  may depend only on information about  $\boldsymbol{\eta}^t$  and  $\boldsymbol{\xi}^t$ . Capacity constraints confine  $\mathbf{x}_t$  to the interval  $[0, 1]$ , and the expected reservoir volume at the end of the second stage (conditional on first stage information) must not be smaller than 1. This is rather a regulatory than a physical restriction which takes account of the fact that the model has a finite horizon, whereas real operation of the plant has an indefinite horizon.

Let us assume that  $\boldsymbol{\eta}_1 = 1.9$  and  $\boldsymbol{\xi}_1 = 0$  are deterministic, while  $\boldsymbol{\eta}_2$  and  $\boldsymbol{\xi}_2$  are independent and follow uniform distributions over  $\Theta := [\frac{1}{2}, 2]$  and  $\Xi := [0, \frac{1}{2}]$ , respectively. Using the notation of Section 2, the above decision problem translates to a two-stage stochastic program as follows.

$$\begin{array}{rcl}
& \underset{\mathbf{x} \in X(\mathbb{F})}{\text{minimize}} & \mathbb{E}(-\boldsymbol{\eta}_1 \mathbf{x}_1 - \boldsymbol{\eta}_2 \mathbf{x}_2) \\
\text{s.t.} & \left. \begin{array}{l}
-\mathbf{x}_1 \leq 1 - \boldsymbol{\xi}_1 \\
\mathbf{x}_1 \leq 1 + \boldsymbol{\xi}_1 \\
-\mathbf{x}_1 - \mathbf{x}_2 \leq 1 - \boldsymbol{\xi}_1 - \boldsymbol{\xi}_2 \\
\mathbf{x}_1 + \mathbf{x}_2 \leq \frac{1}{5} + \boldsymbol{\xi}_1 + \boldsymbol{\xi}_2 \\
\mathbb{E}(\mathbf{x}_1 + \mathbf{x}_2 | \mathcal{F}^1) \leq \boldsymbol{\xi}_1 + \mathbb{E}(\boldsymbol{\xi}_2 | \mathcal{F}^1)
\end{array} \right\} & P\text{-a.s.} \quad (\mathcal{P})
\end{array}$$

Here, the state space  $X$  of the primal decision process is set to the two-dimensional unit square  $[0, 1]^2$ . The almost sure lower bound on terminal reservoir volume (the fourth constraint in  $\mathcal{P}$ ) generates an induced constraint on  $\mathbf{x}_1$ , see [44, 50]. Thus,  $\mathcal{P}$  fails to have relatively complete recourse. However, it satisfies the basic regularity conditions (C1)–(C4), which ensure applicability of the approximation scheme advocated in this article.

As the random parameters of the first stage are deterministic, we have to discretize only  $\boldsymbol{\zeta}_2 = (\boldsymbol{\eta}_2, \boldsymbol{\xi}_2)$ . In order to keep notation simple, we suppress the fixed stage index in the subsequent discussion. We construct a sequence of discrete random vectors  $\{\boldsymbol{\zeta}_J^u\}_{J \in \mathbb{N}}$  in the following way (detailed background information is provided in [31, § 4]). For each  $J \in \mathbb{N}$ , let  $\boldsymbol{\lambda}_J$  be a discrete random

variable uniformly distributed over  $\Lambda_J := \{1, \dots, J\}^2$ , i.e., the event  $\{\boldsymbol{\lambda}_J = (i, k)\}$  has probability  $J^{-2}$  for  $(i, k) \in \Lambda_J$ . Furthermore, for each  $J \in \mathbb{N}$  and  $(i, k) \in \Lambda_J$ , we introduce a two-dimensional random vector  $\boldsymbol{\zeta}_{J,i,k} := (\boldsymbol{\eta}_{J,i,k}, \boldsymbol{\xi}_{J,i,k})$  which is uniformly distributed over the rectangle  $Z_{J,i,k} := \Theta_{J,i} \times \Xi_{J,k}$ ,

$$\Theta_{J,i} := \left[\frac{1}{2} + \frac{3(i-1)}{2J}, \frac{1}{2} + \frac{3i}{2J}\right] \quad \text{and} \quad \Xi_{J,k} := \left[\frac{k-1}{2J}, \frac{k}{2J}\right], \quad (i, k) \in \Lambda_J.$$

Assuming mutual independence of  $\boldsymbol{\lambda}_J$  and  $\{\boldsymbol{\zeta}_{J,i,k}\}_{(i,k) \in \Lambda_J}$ , it is easily seen that

$$\boldsymbol{\zeta} \sim \sum_{(i,k) \in \Lambda_J} \mathbf{1}_{\{\boldsymbol{\lambda}_J = (i,k)\}} \boldsymbol{\zeta}_{J,i,k}. \quad (7.13)$$

Here, the symbol ‘ $\sim$ ’ stands for equivalence in distribution. Since only the distribution of  $\boldsymbol{\zeta}$  has practical relevance for the stochastic program  $\mathcal{P}$ , there is no loss of generality to assume that (7.13) holds with equality. Next, each  $\boldsymbol{\zeta}_{J,i,k}$  is approximated by a discrete random vector  $\boldsymbol{\zeta}_{J,i,k}^u$  as exemplified in [31, § 4.1]. By construction, the (regular) conditional distribution of  $\boldsymbol{\zeta}_{J,i,k}^u$  given  $\boldsymbol{\zeta}_{J,i,k}$  reads

$$P_{\boldsymbol{\zeta}_{J,i,k}^u | \boldsymbol{\zeta}_{J,i,k}}(B | \zeta) = \frac{e_{J,k} - \xi}{e_{J,k} - e_{J,k-1}} \delta_{\zeta_{\Xi, i, k-1}^J}(B) + \frac{\xi - e_{J,k-1}}{e_{J,k} - e_{J,k-1}} \delta_{\zeta_{\Xi, i, k}^J}(B)$$

for all  $B \in \mathcal{B}(Z_{J,i,k})$  and  $\zeta \in Z_{J,i,k}$ , where  $e_{J,k} := \frac{k}{2J}$ ,  $\zeta_{\Xi, i, k}^J := (\mathbb{E}(\boldsymbol{\eta}_{J,i,k}), e_{J,k})$ , and  $\delta_{\zeta_{\Xi, i, k}^J}$  denotes the Dirac measure concentrated at  $\zeta_{\Xi, i, k}^J$ . The discrete random vector  $\boldsymbol{\zeta}_J^u$  is now defined through

$$\boldsymbol{\zeta}_J^u := \sum_{(i,k) \in \Lambda_J} \mathbf{1}_{\{\boldsymbol{\lambda}_J = (i,k)\}} \boldsymbol{\zeta}_{J,i,k}^u.$$

If we assume  $\boldsymbol{\lambda}_J$  and  $\{\boldsymbol{\zeta}_{J,i,k}^u\}_{(i,k) \in \Lambda_J}$  to be mutually independent, then the unconditional distribution of  $\boldsymbol{\zeta}_J^u$  is given by<sup>8</sup>

$$P_{\boldsymbol{\zeta}_J^u}(B) = \sum_{i=1}^J \sum_{k=0}^J \frac{2 - \delta_{k,0} - \delta_{k,J}}{2J^2} \delta_{\zeta_{\Xi, i, k}^J}(B) \quad \text{for} \quad B \in \mathcal{B}(Z).$$

---

<sup>8</sup>Notice that only the unconditional distribution of  $\boldsymbol{\zeta}_J^u$  has practical relevance for the approximate stochastic program  $\mathcal{P}_J^u$ .

As demonstrated in [31],  $\zeta_J^u$  and  $\zeta$  satisfy the conditions (4.2), and  $\|\zeta_J^u - \zeta\|_\infty$  approaches zero as  $J$  tends to infinity. The dual approximation  $\zeta_J^l$  is obtained in a similar way. We set

$$\zeta_J^l := \sum_{(i,k) \in \Lambda_J} 1_{\{\lambda_J=(i,k)\}} \zeta_{J,i,k}^l$$

with  $\lambda_J$  and  $\{\zeta_{J,i,k}^l\}_{(i,k) \in \Lambda_J}$  mutually independent. Each random vector  $\zeta_{J,i,k}^l$  is defined through its conditional distribution given  $\zeta_{J,i,k}$ , that is,

$$P_{\zeta_{J,i,k}^l | \zeta_{J,i,k}}(B | \zeta) = \frac{e_{J,i} - \eta}{e_{J,i} - e_{J,i-1}} \delta_{\zeta_{\Theta,i-1,k}^J}(B) + \frac{\eta - e_{J,i-1}}{e_{J,i} - e_{J,i-1}} \delta_{\zeta_{\Theta,i,k}^J}(B)$$

for  $B \in \mathcal{B}(Z_{J,i,k})$  and  $\zeta \in Z_{J,i,k}$ , where  $e_{J,i} := \frac{1}{2} + \frac{3i}{2J}$  and  $\zeta_{\Theta,i,k}^J := (e_{J,i}, \mathbb{E}(\xi_{J,i,k}))$ .

Then, the unconditional distribution of  $\zeta_J^l$  amounts to

$$P_{\zeta_J^l}(B) = \sum_{i=0}^J \sum_{k=1}^J \frac{2 - \delta_{i,0} - \delta_{i,J}}{2J^2} \delta_{\zeta_{\Theta,i,k}^J}(B) \quad \text{for } B \in \mathcal{B}(Z).$$

By construction,  $\zeta_J^l$  is conditionally independent of  $\zeta_J^u$  given  $\zeta$ . Moreover,  $\zeta_J^l$  and  $\zeta$  satisfy the conditions (4.3), while  $\|\eta_J^l - \zeta\|_\infty$  converges to zero as  $J$  tends to infinity; see [31] for details. Replacing the original data process in  $\mathcal{P}$  by its discrete approximations  $\zeta_J^u$  and  $\zeta_J^l$ , we obtain finite-dimensional problems  $\mathcal{P}_J^u$  and  $\mathcal{P}_J^l$ , whose minima provide upper and lower bounds on  $\inf \mathcal{P}$ , respectively. Convergence of the bounds upon increase of  $J$  is visualized in Figure 1. Notice that both problems  $\mathcal{P}_J^u$  and  $\mathcal{P}_J^l$  involve  $J(J+1)$  different scenarios.

## 8 Conclusions

Bounding approximation schemes for MSPs have very appealing properties. They provide bounds on the optimal objective value which are explicitly available after solving two finite-dimensional approximate problems. These bounds are deterministic (they represent no confidence interval), and they are not dependent on unknown constants, which typically arise in other quantitative approaches. As argued in Section 5, the presented method even enables us to obtain a near-optimal

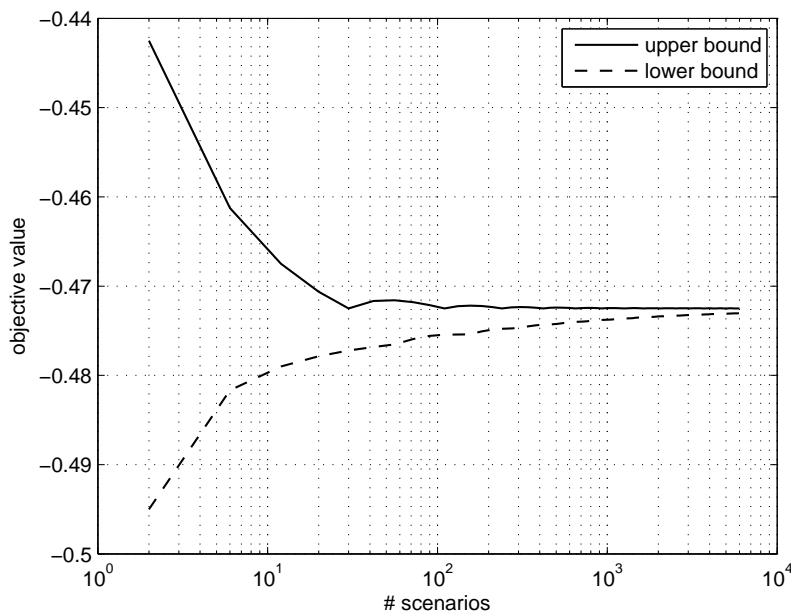


Figure 1: Convergence of bounds

policy for the original MSP. This policy is implementable for every scenario in the (potentially uncountable) support of the original data process, and not just for scenarios in the support of the approximate processes. Moreover, the cost associated with this policy differs from the minimal cost at most by the difference between the bounds. Convergence of the bounds thus implies that there is a sequence of feasible policies whose objective values approach the true minimal cost. No epi-convergence arguments are employed to derive this result. In addition, the presented approximation scheme applies to MSPs *with* expected value constraints and *without* relatively complete recourse. Certain problems with integrated chance constraints and conditional value-at-risk constraints fall into this category.

These strong results are not for free. We require the state spaces of the data and (primal) decision processes to be compact, which is a standard assumption. However, we also require the constraint functions to be jointly convex in the

decisions and the random parameters, which rules out stochastic recourse and technology matrices in linear stochastic programs. In addition, the proposed discretization scheme seems to apply only to admissible data processes in the sense of Definition 4.1.

Extensions to problems with stochastic technology matrices might be possible under certain circumstances and are subject of ongoing research. Other extensions to more general stochastic processes with complex interstage dependencies have been discussed by the author in a dynamic programming framework [30].

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

## References

- [1] BIRGE, J., AND LOUVEAUX, F. *Introduction to Stochastic Programming*. Springer-Verlag, New York, 1997.
- [2] BIRGE, J., AND WETS, R.-B. Computing bounds for stochastic programming problems by means of a generalized moment problem. *Math. Oper. Res.* 12 (1987), 149–162.
- [3] CASEY, M., AND SEN, S. The scenario generation algorithm for multistage stochastic linear programming. *Math. Oper. Res.* 30, 3 (2005), 615–631.
- [4] CHARNES, A., AND COOPER, W. W. Chance-constrained programming. *Manage. Sci.* 6 (1959/1960), 73–79.
- [5] DANTZIG, G., AND INFANGER, G. Large-scale stochastic linear programs—importance sampling and Benders decomposition. *Comput. Appl. Math. I* (1992), 111–120.

- [6] DEMPSTER, M., AND THOMPSON, R. EVPI-based importance sampling solution procedures for multistage stochastic linear programmes on parallel MIMD architectures. *Ann. Oper. Res.* 90 (1999), 161–184.
- [7] DUNFORD, N., AND SCHWARTZ, J. T. *Linear Operators: Part I*. John Wiley and Sons, New York, Chichester, Brisbane, Toronto, Singapore, 1988.
- [8] DUPAČOVÁ, J., GRÖWE-KUSKA, N., AND RÖMISCH, W. Scenario reduction in stochastic programming: an approach using probability metrics. *Math. Program., Ser. A* 95 (2003), 493–511.
- [9] DUPAČOVÁ (AS ŽÁČKOVÁ), J. On minimax solutions of stochastic linear programming problems. *Časopis pro Pěstování Matematiky* 91 (1966), 423–429.
- [10] EDIRISINGHE, N., AND ZIEMBA, W. Bounding the expectation of a saddle function with application to stochastic programming. *Math. Oper. Res.* 19 (1994), 314–340.
- [11] EDIRISINGHE, N., AND ZIEMBA, W. Bounds for two-stage stochastic programs with fixed recourse. *Math. Oper. Res.* 19 (1994), 292–313.
- [12] ERMOLIEV, Y., AND GAIVORONSKI, A. Stochastic quasigradient methods for optimization of discrete event systems. *Ann. Oper. Res.* 39 (1992), 1–39.
- [13] FRAUENDORFER, K. Solving SLP recourse problems with arbitrary multivariate distributions — the dependent case. *Math. Oper. Res.* 13 (1988), 377–394.
- [14] FRAUENDORFER, K. *Stochastic two-stage programming*, vol. 392 of *Lect. Notes Econ. Math. Syst.* Springer-Verlag, Berlin, 1992.
- [15] FRAUENDORFER, K. Multistage stochastic programming: Error analysis for the convex case. *Z. Oper. Res.* 39, 1 (1994), 93–122.

- [16] FRAUENDORFER, K. Barycentric scenario trees in convex multistage stochastic programming. *Math. Program.* 75, 2 (1996), 277–294.
- [17] GASSMANN, H., AND ZIEMBA, W. A tight upper bound for the expectation of a convex function of a multivariate random variable. *Math. Program. Study* 27 (1986), 39–53.
- [18] HANEVELD, W. K., AND VAN DER VLERK, M. Integrated chance constraints: Reduced forms and an algorithm. *Comput. Manag. Sci.* (2006). To appear.
- [19] HEITSCH, H., AND RÖMISCH, W. Scenario reduction algorithms in stochastic programming. *Comput. Optim. Appl.* 24 (2003), 187–206.
- [20] HEITSCH, H., RÖMISCH, W., AND STRUGAREK, C. Stability of multistage stochastic programs. *SIAM J. Optim.* (2006). To appear.
- [21] HIGLE, J., AND SEN, S. Stochastic decomposition: An algorithm for two-stage linear programs with recourse. *Math. Oper. Res.* 16 (1991), 650—669.
- [22] HOCHREITER, R., AND PFLUG, G. Financial scenario generation for stochastic multi-stage decision processes as facility location problems. *Ann. Oper. Res.* (2006). To appear.
- [23] HØYLAND, K., AND WALLACE, S. Generating scenario trees for multistage decision problems. *Manage. Sci.* 47, 2 (2001), 295–307.
- [24] INFANGER, G. *Planning under Uncertainty: Solving Large-Scale Stochastic Linear Programs*. Boyd and Fraser, Danvers, 1994.
- [25] KALL, P. An upper bound for SLP using first and total second moments. *Ann. Oper. Res.* 30 (1991), 267–276.
- [26] KALL, P., AND WALLACE, S. *Stochastic Programming*. John Wiley & Sons, Chichester, 1994.

- [27] KAUT, M., AND WALLACE, S. Evaluation of scenario-generation methods for stochastic programming. The Stochastic Programming E-Print Series (SPEPS), May 2003.
- [28] KLEIN HANEVELD, W. *Duality in Stochastic Linear and Dynamic Programming*, vol. 274 of *Lect. Notes Econ. Math. Syst.* Springer-Verlag, Berlin, 1985.
- [29] KOIVU, M. Variance reduction in sample approximations of stochastic programs. *Math. Program.* 103, 3, Ser. A (2005), 463–485.
- [30] KUHN, D. *Generalized Bounds for Convex Multistage Stochastic Programs*, vol. 548 of *Lect. Notes Econ. Math. Syst.* Springer-Verlag, Berlin, 2004.
- [31] KUHN, D. Aggregation and discretization in multistage stochastic programming. *Math. Program.* (2007). To appear.
- [32] MADANSKY, A. Inequalities for stochastic linear programming problems. *Manage. Sci.* 6 (1960), 197–204.
- [33] PENNANEN, T. Epi-convergent discretizations of multistage stochastic programs. *Math. Oper. Res.* 30, 1 (2005), 245–256.
- [34] PENNANEN, T. Epi-convergent discretizations of multistage stochastic programs via integration quadratures. *Math. Program.* (2006). To appear.
- [35] PFLUG, G. Scenario tree generation for multiperiod financial optimization by optimal discretization. *Math. Program., Ser. B* 89 (2001), 251–271.
- [36] PRÉKOPA, A. *Stochastic Programming*. Kluwer Academic Publishers, Dordrecht, 1995.
- [37] PRÉKOPA, A. Contributions to the theory of stochastic programming. *Math. Program.* 4 (1973), 202–221.



- [38] RACHEV, S., AND RÖMISCH, W. Quantitative stability in stochastic programming: the method of probability metrics. *Math. Oper. Res.* 27 (2002), 792–818.
- [39] ROCKAFELLAR, R. *Conjugate Duality and Optimization*. SIAM, Philadelphia, 1974.
- [40] ROCKAFELLAR, R., AND URYASEV, S. Optimization of conditional value-at-risk. *Journal of Risk* 2, 3 (2000), 21–41.
- [41] ROCKAFELLAR, R., AND URYASEV, S. Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26 (2002), 1443–1471.
- [42] ROCKAFELLAR, R., AND WETS, R.-B. Nonanticipativity and  $L^1$ -martingales in stochastic optimization problems. In *Stoch. Syst.: Model., Identif., Optim. II; Math. Program. Study 6* (1976), pp. 170–187.
- [43] ROCKAFELLAR, R., AND WETS, R.-B. Stochastic convex programming: Basic duality. *Pac. J. Math.* 62 (1976), 173–195.
- [44] ROCKAFELLAR, R., AND WETS, R.-B. Stochastic convex programming: Relatively complete recourse and induced feasibility. *SIAM J. Control Optimization* 14 (1976), 574–589.
- [45] ROCKAFELLAR, R., AND WETS, R.-B. Stochastic convex programming: Singular multipliers and extended duality. *Pac. J. Math.* 62 (1976), 507–522.
- [46] ROCKAFELLAR, R., AND WETS, R.-B. The optimal recourse problem in discrete time:  $L^1$ -multipliers for inequality constraints. *SIAM J. Control Optimization* 16 (1978), 16–36.
- [47] SHAPIRO, A. Inference of statistical bounds for multistage stochastic programming problems. *Math. Methods Oper. Res.* 58, 1 (2003), 57–68.

- [48] SHAPIRO, A. On complexity of multistage stochastic programs. *Oper. Res. Lett.* 34 (2006), 1–8.
- [49] WALLACE, S. W., AND ZIEMBA, W. T., Eds. *Applications of stochastic programming*, vol. 5 of *MPS/SIAM Series on Optimization*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2005.
- [50] WETS, R.-B. Induced constraints for stochastic optimization problems. In *Techniques of Optimization* (1972), A. Balakrishnan, Ed., Academic Press, pp. 433–443.
- [51] WRIGHT, S. Primal-dual aggregation and disaggregation for stochastic linear programs. *Math. Oper. Res.* 19, 4 (1994), 893–908.