An information-based approximation scheme for stochastic optimization problems in continuous time

Daniel Kuhn

dkuhn@doc.ic.ac.uk

Department of Computing

Imperial College of Science, Technology, and Medicine

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

October 30, 2008

Abstract. Dynamic stochastic optimization problems with a large (possibly infinite) number of decision stages and high-dimensional state vector are inherently difficult to solve. In fact, scenario tree based algorithms are unsuitable for problems with many stages, while dynamic programming type techniques are unsuitable for problems with many state variables. This article proposes a stage aggregation scheme for stochastic optimization problems in continuous time, thus having an extremely large (i.e., uncountable) number of decision stages. By perturbing the underlying data and information processes, we construct two approximate problems that provide bounds on the optimal value of the original problem. Moreover, we prove that the gap between the bounds converges to zero as the stage-aggregation is refined. If massive aggregation of stages is possible without sacrificing too much accuracy, the aggregate approximate problems can be addressed by means of scenario tree based methods. The suggested approach applies to problems that exhibit randomness in the objective and the constraints, while the constraint functions are required to be additively separable in the decision variables and the random parameters.

Key words. Stochastic optimization, stochastic control, bounds, time discretization, stage-aggregation

1 Introduction

We study a class of continuous-time stochastic optimization problems which may have a high-dimensional state space. For this problem class, traditional solution methods based on dynamic programming are impractical. Although stochastic programming based approaches can cope with large numbers of decision variables, they only apply to problems with few decision stages. To bypass these difficulties, we extend the stage-aggregation scheme developed in [19] for discrete-time stochastic programs to the continuous-time case. This will enable us to approximate problems with a continuum of decision stages by aggregated problems with finitely many (few) decision stages, at each of which only a finite number of random variables is observed. The resulting aggregated problems are thus promising candidates for numerical solution via scenario tree based methods [4].

In stochastic programming, stage-aggregation was previously investigated in [3, 19, 27] for problems with a finite and in [12, 13, 17] for problems with a (countably) infinite planning horizon. The method in [19], which constitutes the basis of our analysis, constructs two stage-aggregated problems by perturbing the stochastic data process and the information structure of the original problem. This essentially amounts to manipulating the timing of data revelation. If the observation of new data is artificially delayed, then the aggregated problems provide upper bounds on the minimum of the original problem. In contrast, if future observations are foreseen some time ahead, we obtain lower bounds.

The continuous-time setting considered here allows us to study the asymptotic properties of the aggregated problems when their temporal granularity is refined. Our main contribution is to show that the optimal values of the aggregated problems converge to the optimal value of the original problem when the number of stages tends to infinity. This result may also provide insights on how to approximate discrete-time models with a large but finite number of stages. Our method applies to continuous-time versions of linear stochastic programs with random objective and right hand side coefficients but deterministic constraint matrices.

The structure of this paper is as follows. Section 2 introduces the notations and terminology, while Section 3 gives the problem specification. The development of a universal stage-aggregation scheme then proceeds in several steps. First, Section 4 introduces a penalty based reformulation of the original problem. Subsequently, Section 5 embeds the penalized problem of Section 4 into a family of perturbed stochastic optimization problems, the perturbation parameters being the underlying data and information processes. Section 6 identifies certain problems in the parametric family of Section 5 which have finitely many stages and provide bounds on the unperturbed problem. Finally, Section 7 demonstrates that the bounds of Section 6 can be made arbitrarily tight if the number of stages in the aggregated problems tends to infinity.

2 Preliminaries

All random objects in this article are defined on a complete probability space $(\Omega, \mathcal{A}, P)$, which is referred to as the *sample space*. Moreover, we let $\mathbb{T} := [0, T]$ be a finite time interval, $\mathcal{T} := \mathcal{B}(\mathbb{T})$ the Borel field on $\mathbb{T}$, and λ the Lebesgue measure on $\mathcal{T}$. We say that $\boldsymbol{x}$ is a stochastic process with state space $\mathbb{R}^n$ if $\boldsymbol{x} = \{\boldsymbol{x}_t\}_{t\in\mathbb{T}}$, and each random variable $\boldsymbol{x}_t$ maps $(\Omega, \mathcal{A})$ to the Borel space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. All stochastic processes considered below are assumed to be of this kind, i.e., they share the same index set $\mathbb{T}$ and have finite-dimensional Euclidean state spaces. By convention, random objects (i.e., random variables or stochastic processes) appear in boldface, while their realizations are denoted by the same symbols in normal face. The stochastic process $\boldsymbol{x}$ is called $\mathcal{S}$-measurable if $\mathcal{S}$ is a σ-algebra on $\mathbb{T} \times \Omega$ and the assignment $(t, \omega) \mapsto \boldsymbol{x}_t(\omega)$ represents a measurable mapping from $(\mathbb{T} \times \Omega, \mathcal{S})$ to $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. We say that two stochastic processes $\boldsymbol{x}$ and $\boldsymbol{x}'$ with the same state space $\mathbb{R}^n$ are almost everywhere identical if $\boldsymbol{x}_t(\omega) = \boldsymbol{x}'_t(\omega)$ for $\lambda \otimes P$-almost all (t, ω). Sometimes, we use the short-hand notation $\boldsymbol{x} \sim \boldsymbol{x}'$

to describe this relationship. For $1 \le p \le \infty$ we denote by $\mathcal{L}_n^p(\mathbb{T} \times \Omega, \mathcal{S}, \lambda \otimes P)$ the Banach space of equivalence classes of almost everywhere identical stochastic processes valued in $\mathbb{R}^n$ that are $\mathcal{S}$-measurable and bounded with respect to the p-norm on $(\mathbb{T} \times \Omega, \mathcal{S}, \lambda \otimes P)$. For an arbitrary filtration $\mathbb{G} := \{\mathcal{G}_t\}_{t \in \mathbb{T}}$ on the probability space $(\Omega, \mathcal{A}, P)$, we define[1]

$$\mathcal{S}(\mathbb{G}) := \{A \in \mathcal{T} \otimes \mathcal{A} \,|\, \{\omega \in \Omega | (t, \omega) \in A\} \in \mathcal{G}_t \; \forall \, t \in \mathbb{T}\} .$$

It can easily be verified that $\mathcal{S}(\mathbb{G})$ represents a sub-σ-algebra of $\mathcal{T} \otimes \mathcal{A}$. Moreover, one can show that $\mathcal{S}(\mathbb{G})$ is generated by the class of bounded real-valued stochastic processes which are $\mathcal{T} \otimes \mathcal{A}$-measurable and $\mathbb{G}$-adapted. Consequently, every $\mathcal{S}(\mathbb{G})$-measurable stochastic process is $\mathcal{T} \otimes \mathcal{A}$-measurable and $\mathbb{G}$-adapted. Let $\mathcal{M}$ be the completion of $\mathcal{T} \otimes \mathcal{A}$ and denote by $\mathcal{M}(\mathbb{G})$ the completion of $\mathcal{S}(\mathbb{G})$ with respect to $\lambda \otimes P$. For any $\boldsymbol{x} \in \mathcal{L}_n^1(\mathbb{T} \times \Omega, \mathcal{M}, \lambda \otimes P)$ there exists an $\mathcal{M}$-measurable process $\boldsymbol{x}'$ such that $\boldsymbol{x}'_t = \mathrm{E}(\boldsymbol{x}_t | \mathcal{G}_t)$ almost everywhere with respect to $\lambda \otimes P$ and for suitable versions of the conditional expectations. In fact, $\boldsymbol{x}'$ can be taken to be the $\mathbb{G}$-predictable or $\mathbb{G}$-optional projection [8] of any $\mathcal{T} \otimes \mathcal{A}$-measurable process in the equivalence class of $\boldsymbol{x}$. By using Fubini's theorem, one verifies that $\boldsymbol{x}'$ is unique up to a set of $\lambda \otimes P$ measure zero. Moreover, one can show $\boldsymbol{x}'$ to be $\mathcal{M}(\mathbb{G})$-measurable and integrable with respect to $\lambda \otimes P$. If the filtration $\mathbb{G}$ is constant in the sense that $\mathcal{G}_t \equiv \mathcal{G}$ for all $t \in \mathbb{T}$, we denote $\boldsymbol{x}'$ by $\mathrm{E}(\boldsymbol{x} | \mathcal{G})$ and refer to it as the conditional expectation process of $\boldsymbol{x}$ given $\mathcal{G}$.

If x is a real number, then $[x]^+$ denotes $\max\{0, x\}$, while $\lceil x \rceil$ stands for the smallest integer in $[x, +\infty)$ and $\lfloor x \rfloor$ for the largest integer in $(-\infty, x]$.

3 Problem formulation

Let $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ be two $\mathcal{M}$-measurable stochastic processes with state spaces $\mathbb{R}^k$ and $\mathbb{R}^l$, respectively. We also introduce the combined data process $\boldsymbol{\zeta}$ which is defined

[1] We use the same symbol '$\otimes$' to denote product measures and product σ-fields, see [7].

through $\boldsymbol{\zeta}_t := (\boldsymbol{\eta}_t, \boldsymbol{\xi}_t)$, thus having state space $\mathbb{R}^{k+l}$. We will restrict attention to data processes with rcll sample paths ('rcll' stands for 'right-continuous with left limits'). By convention, such processes are termed rcll. At any time t, the decision maker is assumed to have complete information about the past realizations of the process $\boldsymbol{\zeta}$, but only distributional formation about its future. The information $\mathcal{F}_t$ available at time t is therefore represented as the σ-algebra generated by all observations up to time t, that is, $\mathcal{F}_t := \sigma(\boldsymbol{\zeta}_s | 0 \le s \le t)$. We use the convention $\mathcal{F} := \mathcal{F}_T$ and let $\mathbb{F} := \{\mathcal{F}_t\}_{t\in\mathbb{T}}$ be the filtration induced by the data process. Without loss of generality, we may identify the σ-field $\mathcal{A}$ with the P-completion of $\mathcal{F}$. It represents the maximal σ-algebra on Ω that is needed throughout this article. As $\boldsymbol{\zeta}$ is rcll, $\mathcal{F}$ is countably generated, see e.g. [22, § I.4.21]. Hence, $\mathcal{M}$ is countably generated up to $\lambda \otimes P$-null sets, which implies via a classical result of Doob [10, p. 92] that $\mathcal{L}^p_n(\mathbb{T} \times \Omega, \mathcal{M}, \lambda \otimes P)$ is separable for $1 \le p < \infty$.

The decision-maker's policy[2] will be characterized by a stochastic process $\boldsymbol{x}$ with state space $\mathbb{R}^n$. It assigns to every time point $t \in \mathbb{T}$ and scenario $\omega \in \Omega$ a vector of n decisions $\boldsymbol{x}_t(\omega)$. By convention, we let $X := \mathcal{L}^\infty_n(\mathbb{T} \times \Omega, \mathcal{M}, \lambda \otimes P)$ denote the basic policy space. However, policies may not be chosen freely from X but are subject to different constraints. We will always assume below that those constraints which are independent of $\boldsymbol{\zeta}$ can be brought to the form[3]

$$\left.\begin{array}{l} \boldsymbol{x}_t \in \Gamma(t) \\ W_1(t)\,\boldsymbol{x}_t + \int_0^t W_2(s)\,\boldsymbol{x}_s\,\lambda(\mathrm{d}s) = h(t) \end{array}\right\} \quad \lambda \otimes P\text{-a.e.} \tag{3.1}$$

This representation involves a multifunction $\Gamma : \mathbb{T} \rightrightarrows \mathbb{R}^n$, two matrix-valued mappings $W_1, W_2 : \mathbb{T} \to \mathbb{R}^{n\times n}$, and a vector-valued mapping $h : \mathbb{T} \to \mathbb{R}^n$. We now introduce a family of subsets of the basic policy space X which will be relevant for our subsequent analysis.

Definition 3.1. *For any filtration $\mathbb{G} := \{\mathcal{G}_t\}_{t\in\mathbb{T}}$ on the sample space, we set*

[2] Equivalently, policies will be referred to as *strategies*, *decision rules*, or *decision processes*.

[3] Most applications require less than n dynamic constraints. In these cases, we assume simply that some rows of W_1, W_2, and h vanish identically.

$X(\mathbb{G}) := \mathcal{L}_n^\infty(\mathbb{T}\times\Omega, \mathcal{M}(\mathbb{G}), \lambda\otimes P)$. *Moreover, we let* $X_\mathrm{c}(\mathbb{G})$ *be the collection of all* $\boldsymbol{x} \in X(\mathbb{G})$ *that satisfy the constraints* (3.1).

Thus, $X_\mathrm{c}(\mathbb{G})$ contains all (equivalence classes of) bounded, measurable, and $\mathbb{G}$-adapted stochastic processes subject to a set of dynamic and pointwise constraints independent of the data process. We say that $\mathcal{M}(\mathbb{G})$-measurable decision processes are *non-anticipative* with respect to $\mathbb{G}$.

This article studies nonlinear stochastic optimization problems of the form

$$\begin{aligned} &\underset{\boldsymbol{x}\in X_\mathrm{c}(\mathbb{F})}{\text{minimize}} \ \mathrm{E}\int_0^T c(\boldsymbol{x}_t, \boldsymbol{\eta}_t, t)\,\lambda(\mathrm{d}t) \\ &\text{s.t.} \quad f(\boldsymbol{x}_t, \boldsymbol{\xi}_t, t) \le 0 \quad \lambda\otimes P\text{-a.e.} \end{aligned} \tag{$\mathcal{P}_\infty$}$$

Note that the set of admissible strategies is given by $X_\mathrm{c}(\mathbb{F})$, expressing the natural requirement that decisions must be non-anticipative with respect to the filtration induced by the data process. The scalar function $c:\mathbb{R}^n\times\mathbb{R}^k\times\mathbb{T}\to\mathbb{R}$ characterizes the instantaneous cost rate, while the vector-valued function $f:\mathbb{R}^n\times\mathbb{R}^l\times\mathbb{T}\to\mathbb{R}^m$ is used to express a set of pointwise constraints depending on the data process.

Unless otherwise noted, the following conditions are assumed to hold below:

(C1) c is Borel measurable, $c(\cdot,\eta,t)$ is convex, and $c(x,\cdot,t)$ is concave; f is Borel measurable and additively separable, $f(x,\xi,t) = f_1(x,t) + f_2(\xi,t)$, where $f_1(\cdot,t)$ and $f_2(\cdot,t)$ are componentwise convex;

(C2) there are convex nondecreasing functions ψ_x, ψ_η, $\psi_\xi : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\left.\begin{aligned} |c(x,\eta,t)| &\le \psi_x(|x|)\,\psi_\eta(|\eta|) \\ |f(x,\xi,t)| &\le \psi_x(|x|)\,\psi_\xi(|\xi|) \end{aligned}\right\} \quad \forall\, x\in\mathbb{R}^n,\ \eta\in\mathbb{R}^k,\ \xi\in\mathbb{R}^l,\ t\in\mathbb{T};$$

(C3) the data process is a Markov martingale with rcll sample paths, while the random variables $\sup_t \psi_\eta(|\boldsymbol{\eta}_t|)$ and $\sup_t \psi_\xi(|\boldsymbol{\xi}_t|)$ are P-integrable;[4]

[4]Convexity and monotonicity imply that the growth functions ψ_η and ψ_ξ are continuous. Hence, as the data process is rcll, $\sup_t \psi_\eta(|\boldsymbol{\eta}_t|)$ and $\sup_t \psi_\xi(|\boldsymbol{\xi}_t|)$ are in fact random variables.

(C4) the multifunction Γ is Borel measurable, bounded, and nonempty convex closed-valued; W_1, W_2, and h are Borel measurable and bounded;

(C5) problem $\mathcal{P}_\infty$ is feasible.

Condition (C1) requires the constraints to be additively separable with respect to decision variables and random parameters. Note that, by choosing f appropriately, certain components of $\boldsymbol{x}$ and $\boldsymbol{\xi}$ can be tied together. In this way, the data process $\boldsymbol{\xi}$ can impact the dynamic integral constraints in (3.1). This dependence is left implicit in order to simplify our exposition. The requirement that the data process $\boldsymbol{\zeta}$ be Markovian will be crucial in Sections 6 and 7. However, the somewhat restrictive martingale requirement is only introduced to enhance transparency and will later be relaxed (see Remark 7.9).

Problem $\mathcal{P}_\infty$ can be seen as a continuous-time generalization of a linear multistage stochastic program with deterministic constraint matrices and random objective and/or right hand side coefficients. Problems of this type arise e.g. in inventory control [16], power system operation and scheduling [21], the management of interest-bearing bank deposits [14], the valuation of electricity derivatives [15], supply chain management [2], airline revenue management [9], etc.

Remark 3.2 (Path-dependent constraints)**.** *Problem $\mathcal{P}_\infty$ and the subsequent results can be readily generalized to allow for path-dependent constraints of the form*

$$f(\boldsymbol{x}_{\tau_1(t)}, \ldots, \boldsymbol{x}_{\tau_d(t)}, \boldsymbol{\xi}_t, t) \leq 0 \quad \lambda \otimes P\text{-a.e.}, \tag{3.2}$$

where each $\tau_i : \mathbb{T} \to \mathbb{T}$ is Borel measurable and satisfies $\tau_i(t) \leq t$ for all $t \in \mathbb{T}$, $i = 1, \ldots, d$. In analogy to (C1), one must require the Borel-measurable constraint function $f : \mathbb{R}^{n\times d} \times \mathbb{R}^l \times \mathbb{T} \to \mathbb{R}^m$ to be additively separable and convex in the decision variables and the random parameters. For simplicity of presentation, we do not consider this more general problem here.

In the remainder of this section we study the structural properties of $X_\mathrm{c}(\mathbb{G})$, where $\mathbb{G} := \{\mathcal{G}_t\}_{t\in\mathbb{T}}$ denotes an arbitrary filtration on the sample space. To this end, we first establish an auxiliary result about the linear space $X(\mathbb{G})$.

Lemma 3.3. *If $\boldsymbol{x} \in X(\mathbb{G})$ and $\boldsymbol{z}$ is defined via $\boldsymbol{z}_t = \int_0^t \boldsymbol{x}_s \,\lambda(\mathrm{d}s)$, then $\boldsymbol{z} \in X(\mathbb{G})$.*

The proof of this general result about stochastic processes is relegated to Appendix A. Next, we introduce the linear space $V := \mathcal{L}^1_n(\mathbb{T} \times \Omega, \mathcal{M}, \lambda \otimes P)$ together with a pairing $\langle \cdot, \cdot \rangle : X \times V \to \mathbb{R}$ defined through

$$\langle \boldsymbol{x}, \boldsymbol{v} \rangle := \mathrm{E} \int_0^T \boldsymbol{x}_t \cdot \boldsymbol{v}_t \,\lambda(\mathrm{d}t) \quad \text{for all} \quad \boldsymbol{x} \in X,\ \boldsymbol{v} \in V. \tag{3.3}$$

It is well-known that the pairing (3.3) leads to a natural identification of continuous linear functionals on V and elements of X. In this sense, X is dual to V. Moreover, every $\boldsymbol{v} \in V$ defines a continuous linear functional $\boldsymbol{x} \mapsto \langle \boldsymbol{x}, \boldsymbol{v} \rangle$ on X (but not every continuous linear functional on X is representable in this way). The weak* topology on X is the coarsest topology preserving continuity of all linear functionals associated with elements of V, see e.g. [11, § V]. By the remarks at the beginning of Section 3, V is separable, which implies that the weak* topology on X is metrizable on bounded sets [11, Theorem V.5.1]. This is a crucial prerequisite for the applicability of certain epi-convergence results needed below.

Proposition 3.4. *$X_\mathrm{c}(\mathbb{G})$ is a weak* compact subset of X.*

Proof. Set $\delta_\Gamma(t,x) = 0$ for $x \in \Gamma(t)$, and $\delta_\Gamma(t,x) = +\infty$ otherwise. By (C4) the indicator function $\delta_\Gamma : \mathbb{T} \times \mathbb{R}^n \to (-\infty, +\infty]$ constitutes a normal convex integrand, see e.g. [26, Example 14.32]. The corresponding integral functional

$$I_{\delta_\Gamma} : X \to (-\infty, +\infty]\,, \quad I_{\delta_\Gamma}(\boldsymbol{x}) := \mathrm{E} \int_0^T \delta_\Gamma(t, \boldsymbol{x}_t) \,\lambda(\mathrm{d}t)\,,$$

is well-defined by [26, Proposition 14.58]. Moreover, I_{δ_Γ} is proper and convex since Γ is convex-valued and since any $\boldsymbol{x}$ feasible in $\mathcal{P}_\infty$ satisfies $I_{\delta_\Gamma}(\boldsymbol{x}) = 0$; note that at least one such $\boldsymbol{x}$ exists by (C5). Next, introduce the conjugate integrand

$$\delta^*_\Gamma : \mathbb{T} \times \mathbb{R}^n \to (-\infty, +\infty]\,, \quad \delta^*_\Gamma(t, v) := \sup_{x \in \mathbb{R}^n} \{x \cdot v - \delta_\Gamma(t, x)\}\,,$$

which is convex and normal, see [26, Theorem 14.50]. Note that $\delta^*_\Gamma(t,0)=0$ for all t since Γ is nonempty-valued. Let us then define $I_{\delta^*_\Gamma}$ in the obvious way as an integral functional over V. It is easy to see that $I_{\delta^*_\Gamma}$ is well-defined, proper, and convex. Properness holds since $I_{\delta^*_\Gamma}(0)=0$. By [23, Theorem 2], I_{δ_Γ} on X and $I_{\delta^*_\Gamma}$ on V are conjugate to each other with respect to the pairing (3.3). This implies via [25, Theorem 5] that I_{δ_Γ} is lower semicontinuous (lsc) with respect to the weak* topology on X, and thus $X_\Gamma := \{\boldsymbol{x}\in X | I_{\delta_\Gamma}(\boldsymbol{x})\le 0\}$ is weak* closed.

Let us now introduce a deterministic process $\boldsymbol{h}\in X$ with $\boldsymbol{h}_t = h(t)$ and a bounded linear operator $W : X\to X$ defined through

$$[W(\boldsymbol{x})]_t = W_1(t)\,\boldsymbol{x}_t + \textstyle\int_0^t W_2(s)\,\boldsymbol{x}_s\,\lambda(\mathrm{d}s) \quad \text{for} \quad \boldsymbol{x}\in X\,.$$

Since W_1 and W_2 are bounded deterministic functions valued in $\mathbb{R}^{n\times n}$, we can invoke Lemma 3.3 to show that the process $W(\boldsymbol{x})$ is well-defined and contained in X. Define $X_W := \{\boldsymbol{x}\in X \,|\, W(\boldsymbol{x})\sim \boldsymbol{h}\}$ and note that

$$X_W = \bigcap_{\boldsymbol{v}\in V}\{\boldsymbol{x}\in X \,|\, \langle W(\boldsymbol{x}),\boldsymbol{v}\rangle = \langle \boldsymbol{h},\boldsymbol{v}\rangle\}\,.$$

For every $\boldsymbol{v}\in V$ the assignment $\boldsymbol{x}\mapsto\langle W(\boldsymbol{x}),\boldsymbol{v}\rangle$ represents a bounded linear functional on X and is easily shown to be weak* continuous. Hence, X_W is weak* closed as an intersection of (uncountably many) weak* closed hyperplanes.

Next, let $P_{\mathbb{G}} : X\to X$ be the projection which maps $\boldsymbol{x}$ to the process $\boldsymbol{x}'$ with the property that $\boldsymbol{x}'_t = \mathrm{E}(\boldsymbol{x}_t|\mathcal{G}_t)$ almost everywhere with respect to $\lambda\otimes P$. By using the law of iterated conditional expectations and Fubini's theorem, one can show that $X(\mathbb{G}) := \{\boldsymbol{x}\in X | P_{\mathbb{G}}(\boldsymbol{x})\sim\boldsymbol{x}\}$ is a weak* closed subspace of X.

In summary, $X_{\mathrm{c}}(\mathbb{G}) = X_\Gamma\cap X_W\cap X(\mathbb{G})$ is bounded (X_Γ is bounded by (C4)) and weak* closed. By Alaoglu's theorem, $X_{\mathrm{c}}(\mathbb{G})$ is thus weak* compact. □

Note that $X_{\mathrm{c}}(\mathbb{G})$ is nonempty for any choice of $\mathbb{G}$. By assumption (C5), it is nonempty at least for $\mathbb{G}=\mathbb{F}$. Moreover, any $\boldsymbol{x}\in X_{\mathrm{c}}(\mathbb{F})$ can be projected to a 'deterministic' process $\boldsymbol{x}^{\mathrm{d}}$ defined through $\boldsymbol{x}^{\mathrm{d}}_t := \mathrm{E}(\boldsymbol{x}_t)$. It is easily verified that

$\boldsymbol{x}^{\mathrm{d}} \in X_{\mathrm{c}}(\mathbb{G})$, where $\mathbb{G}$ may be any filtration on the sample space (a more general statement will be proved in Proposition 6.2). We conclude this section with a simple example problem of the type $\mathcal{P}_\infty$ that satisfies the conditions (C1)–(C5).

Example 3.5. *Consider a hydropower producer who operates a single pumped storage power plant over a finite planning horizon $\mathbb{T}$. At any time $t \in \mathbb{T}$, electric energy can be bought or sold on the spot market at price $\boldsymbol{\eta}_t$, while new (potential) energy becomes freely available at rate $\boldsymbol{\xi}_t$ in the form of precipitation runoffs and melt water. The portion $\boldsymbol{x}_{\mathrm{infl},t} \leq \boldsymbol{\xi}_t$ is fed into the reservoir, and the residual amount $\boldsymbol{\xi}_t - \boldsymbol{x}_{\mathrm{infl},t}$ is immediately spilt. We denote by $\boldsymbol{x}_{\mathrm{sell},t}$ the rate of energy sold, and by $\boldsymbol{x}_{\mathrm{buy},t}$ the rate of energy bought on the spot market. Only a fraction $\epsilon_{\mathrm{p}} \leq 1$ of all purchased energy quantities can effectively be stored in the reservoir for later resale; the rest is dissipated in the pumps and generators of the power plant. Reservoir content at time t is denoted by $\boldsymbol{x}_{\mathrm{store},t}$. We use the standard notation $\boldsymbol{\zeta} = (\boldsymbol{\eta}, \boldsymbol{\xi})$ to denote the data process and let $\boldsymbol{x} = (\boldsymbol{x}_{\mathrm{buy}}, \boldsymbol{x}_{\mathrm{sell}}, \boldsymbol{x}_{\mathrm{infl}}, \boldsymbol{x}_{\mathrm{store}})$ be the $(n = 4)$-dimensional decision process. Moreover, $\mathbb{F}$ denotes the filtration generated by $\boldsymbol{\zeta}$. Assuming risk-neutrality, the objective is to minimize expected costs from spot market transactions, i.e.,*

$$\mathrm{E} \int_0^T (c_{\mathrm{buy}} + \boldsymbol{\eta}_t)\, \boldsymbol{x}_{\mathrm{buy},t} + (c_{\mathrm{sell}} - \boldsymbol{\eta}_t)\, \boldsymbol{x}_{\mathrm{sell},t}\, \lambda(\mathrm{d}t)\,, \tag{3.4}$$

where c_{buy} and c_{sell} stand for the variable operating costs. The continuity equation for reservoir storage gives rise to a dynamic constraint of the form

$$\boldsymbol{x}_{\mathrm{store},t} - \boldsymbol{x}_{\mathrm{store},0} = \int_0^t \boldsymbol{x}_{\mathrm{infl},s} + \epsilon_{\mathrm{p}}\, \boldsymbol{x}_{\mathrm{buy},s} - \boldsymbol{x}_{\mathrm{sell},s}\, \lambda(\mathrm{d}s) \quad \lambda \otimes P\text{-a.e.}\,, \tag{3.5a}$$

while some physical capacity restrictions apply to the generation, pumping, and inflow rates as well as to reservoir content:

$$\left.\begin{array}{ll} 0 \leq \boldsymbol{x}_{\mathrm{buy},t} \leq \overline{x}_{\mathrm{buy}} & 0 \leq \boldsymbol{x}_{\mathrm{sell},t} \leq \overline{x}_{\mathrm{sell}} \\ 0 \leq \boldsymbol{x}_{\mathrm{infl},t} \leq \overline{x}_{\mathrm{infl}} & 0 \leq \boldsymbol{x}_{\mathrm{store},t} \leq \overline{x}_{\mathrm{store}} \end{array}\right\} \quad \lambda \otimes P\text{-a.e.} \tag{3.5b}$$

We denote by $X_{\mathrm{c}}(\mathbb{F})$ the usual set of essentially bounded decision processes which are non-anticipative with respect to $\mathbb{F}$ and satisfy the constraints (3.5)*. Only one* *additional constraint depends explicitly on the data process:*

$$\boldsymbol{x}_{\mathrm{infl},t} \leq \boldsymbol{\xi}_t \quad \lambda \otimes P\text{-}a.e. \tag{3.6}$$

Problem $\mathcal{P}_\infty$ *thus consists in minimizing* (3.4) *over all* $\boldsymbol{x} \in X_{\mathrm{c}}(\mathbb{F})$ *subject to* (3.6)*. The price and inflow rate processes are driven by two independent standard Brownian motions* $\boldsymbol{w}_\eta$ *and* $\boldsymbol{w}_\xi$ *on* $(\Omega, \mathcal{A}, P)$*, that is,* $\boldsymbol{\eta}$ *and* $\boldsymbol{\xi}$ *are defined through*

$$\left.\begin{array}{rcl} \boldsymbol{\eta}_t &=& \boldsymbol{\eta}_0 \exp\left(-\frac{1}{2}\sigma_\eta^2 t + \sigma_\eta \boldsymbol{w}_{\eta,t}\right) \\ \boldsymbol{\xi}_t &=& \boldsymbol{\xi}_0 \exp\left(-\frac{1}{2}\sigma_\xi^2 t + \sigma_\xi \boldsymbol{w}_{\xi,t}\right) \end{array}\right\} \quad \forall\, t \in \mathbb{T}$$

with deterministic initial data $\boldsymbol{\eta}_0$ *and* $\boldsymbol{\xi}_0$*.*

In this simple example, assumption (C1) trivially holds, while (C2) holds with $\psi_x(u) := \max\{1, 2u\}$ *and* $\psi_\eta(u) := \psi_\xi(u) := u$ *for* $u \in \mathbb{R}_+$*. By definition,* $\boldsymbol{\zeta}$ *represents a Markov martingale with continuous sample paths. Standard results on Brownian motion [22] can be used to show that the random variables* $\sup_t \psi_\eta(|\boldsymbol{\eta}_t|)$ *and* $\sup_t \psi_\xi(|\boldsymbol{\xi}_t|)$ *are integrable with respect to* P*. Thus, condition (C3) follows. Bearing in mind the specific structure of the dynamic and pointwise constraints* (3.5)*, assumption (C4) is evident if the mappings* Γ*,* W_1*,* W_2*, and* h *are defined in the obvious way. Finally, the feasibility condition (C5) holds since the deterministic 'spill-everything' strategy (*$\boldsymbol{x}_{\mathrm{buy},t} \equiv \boldsymbol{x}_{\mathrm{sell},t} \equiv \boldsymbol{x}_{\mathrm{infl},t} \equiv 0$, $\boldsymbol{x}_{\mathrm{store},t} \equiv \boldsymbol{x}_{\mathrm{store},0}$*) satisfies the constraints* (3.5) *and* (3.6)*, while having zero objective value.*

4 Relaxation

For every nonnegative $\pi \in \mathbb{R}^m$ we introduce an auxiliary stochastic program $\mathcal{P}_\pi$.

$$\underset{\boldsymbol{x} \in X_{\mathrm{c}}(\mathbb{F})}{\text{minimize}} \; \mathrm{E} \int_0^T c(\boldsymbol{x}_t, \boldsymbol{\eta}_t, t) + \pi \cdot [f(\boldsymbol{x}_t, \boldsymbol{\xi}_t, t)]^+ \, \lambda(\mathrm{d}t) \tag{$\mathcal{P}_\pi$}$$

By construction, $\mathcal{P}_\pi$ represents a relaxation of $\mathcal{P}_\infty$, in which violations of the explicit constraints are subject to proportional penalty costs. Observe that π is conveniently interpreted as a vector of constraint violation prices. The notation

'$\mathcal{P}_\pi$' is in line with our convention to denote the original problem by '$\mathcal{P}_\infty$'. In fact, $\mathcal{P}_\infty$ is recovered from $\mathcal{P}_\pi$ by letting all components of π tend to infinity.

Proposition 4.1. *For every $\varepsilon > 0$ there exists a nonnegative $\pi \in \mathbb{R}^m$ such that*

$$\inf \mathcal{P}_\infty \geq \inf \mathcal{P}_\pi \geq \inf \mathcal{P}_\infty - \varepsilon\,.$$

Proof. Let e be the m-vector whose components are all 1's. Moreover, for every $\nu \in \mathbb{N}$, let g_ν be the objective function of problem $\mathcal{P}_{\nu e}$ (corresponding to the penalty vector $\pi = \nu\, e$). This functional is finite, convex, and weak* lsc on X. Finiteness and weak* lower semicontinuity follow from [24, Corollary 2A], which applies due to the assumptions (C2) and (C3).[5] Next, define the extended-real-valued functional γ_ν through $\gamma_\nu(\boldsymbol{x}) = g_\nu(\boldsymbol{x})$ for $\boldsymbol{x} \in X_{\mathrm{c}}(\mathbb{F})$; $= +\infty$ otherwise. Proposition 3.4 implies that γ_ν is proper, convex, and weak* lsc on X. Finally, introduce another extended-real-valued functional γ by setting

$$\gamma(\boldsymbol{x}) := \begin{cases} \mathrm{E} \int_0^T c(\boldsymbol{x}_t, \boldsymbol{\eta}_t, t)\, \lambda(\mathrm{d}t) & \text{for } \boldsymbol{x} \in X_{\mathrm{c}}(\mathbb{F}),\ f(\boldsymbol{x}_t, \boldsymbol{\xi}_t, t) \leq 0\ \lambda \otimes P\text{-a.e.}, \\ +\infty & \text{else.} \end{cases}$$

By construction, $\{\mathrm{epi}\, \gamma_\nu\}_{\nu \in \mathbb{N}}$ constitutes a decreasing sequence of convex closed[6] epigraphs with

$$\lim_{\nu \to \infty} \mathrm{epi}\, \gamma_\nu = \bigcap_{\nu=1}^{\infty} \mathrm{epi}\, \gamma_\nu = \mathrm{epi}\, \gamma\,,$$

the set limit being taken in the sense of Painlevé-Kuratowski. Thus, γ_ν epi-converges to γ. Let then $\boldsymbol{x}_\nu$ be a minimizer of γ_ν. Note that $\boldsymbol{x}_\nu$ exists because the effective domain of γ_ν is weak* compact and since γ_ν is proper and weak* lsc. As the sequence $\{\boldsymbol{x}_\nu\}_{\nu \in \mathbb{N}} \subset X_{\mathrm{c}}(\mathbb{F})$ has a weak* accumulation point, we conclude via [1, Theorem 2.5] that

$$\lim_{\nu \to \infty} \inf_{\boldsymbol{x} \in X} \gamma_\nu(\boldsymbol{x}) = \inf_{\boldsymbol{x} \in X} \gamma(\boldsymbol{x})\,.$$

[5] Note that the assignment $(t, \omega; x) \mapsto c(x, \boldsymbol{\eta}_t(\omega), t) + \pi \cdot [f(x, \boldsymbol{\xi}_t(\omega), t)]^+$ defines a normal convex integrand that is $\mathcal{M}$-measurable in (t, ω) and convex continuous in x [26, Example 14.92].

[6] We assume in this proof that $X \times \mathbb{R}$ is endowed with the product of the weak* topology on X and the Euclidean topology on $\mathbb{R}$.

The above reasoning implies that the sequence $\{\inf \mathcal{P}_{\nu e}\}_{\nu \in \mathbb{N}}$ converges from below to $\inf \mathcal{P}_\infty$, and thus the claim follows. □

As a byproduct, the proof of Proposition 4.1 reveals that the functional γ is weak* lsc. In fact, its epigraph is closed since it is representable as an intersection of closed sets. Being covered by $X_c(\mathbb{F})$, the level sets of γ are therefore weak* compact. Since γ is also proper by assumption (C5), we may conclude that problem $\mathcal{P}_\infty$ is solvable and has a finite minimum.

In the remainder, we will develop an approximation scheme for the relaxed optimization problem $\mathcal{P}_\pi$, assuming that the penalty vector π is held constant and that $\mathcal{P}_\pi$ represents an acceptable approximation for $\mathcal{P}_\infty$ (as we have seen, this can always be enforced by choosing large penalties). It is worthwhile to remark that, in reality, constraints are not always required to hold strictly — especially if they depend on uncontrollable random parameters. Instead, some constraints may be violated at a certain cost. In these cases, the penalty formulation $\mathcal{P}_\pi$ arises naturally, and we may consider $\mathcal{P}_\pi$ as the 'true' problem to be solved.

As the penalty vector π is kept fixed, below, we will henceforth refer to $\mathcal{P}_\pi$ simply as $\mathcal{P}$, that is, we drop the subscript 'π' for notational simplicity. Working with $\mathcal{P}$ instead of $\mathcal{P}_\infty$ has distinct technical advantages. In fact, the next section introduces a reformulation of $\mathcal{P}$ as a min-max problem, where minimization is over a primal and maximization over a certain dual feasible set. For problem $\mathcal{P}$ both the primal and dual feasible sets turn out to be bounded, which will facilitate the convergence proof in Section 7 (see Proposition 7.3).

Example 4.2. *In Example 3.5, the approximation of $\mathcal{P}_\infty$ by a relaxed problem $\mathcal{P} \cong \mathcal{P}_\pi$, in which violations of the constraint* (3.6) *are possible at finite cost, requires selection of a suitable penalty vector π. If the spot price $\boldsymbol{\eta}_t$ has low probability of exceeding some threshold level η^+ in the interval $\mathbb{T}$, then the choice $\pi = \eta^+/\epsilon_{\mathrm{P}}$ guarantees that* $\inf \mathcal{P}$ *is only slightly smaller than* $\inf \mathcal{P}_\infty$*; the intuition behind this argument is outlined in [18, § 6.4] in a slightly different context.*

5 Lagrangian reformulation

The first step towards a flexible approximation scheme consists in embedding $\mathcal{P}$ into a larger family of perturbed stochastic optimization problems. To this end, let $\mathbb{G} = \{\mathcal{G}_t\}_{t\in\mathbb{T}}$ and $\mathbb{H} = \{\mathcal{H}_t\}_{t\in\mathbb{T}}$ be arbitrary filtrations on the sample space. As in the discrete-time case considered in [19, § 2], we define a family of optimization problems, which depend parametrically on the two filtrations $\mathbb{G}$ and $\mathbb{H}$ as well as on the data processes $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$.

$$\underset{\boldsymbol{x}\in X_{\mathrm{c}}(\mathbb{G})}{\text{minimize}} \; \mathrm{E}\textstyle\int_0^T c(\boldsymbol{x}_t, \boldsymbol{\eta}_t, t) + \pi \cdot [\mathrm{E}\,(f(\boldsymbol{x}_t, \boldsymbol{\xi}_t, t) \mid \mathcal{H}_t)]^+ \, \lambda(\mathrm{d}t) \qquad (\mathcal{P}(\mathbb{G}, \mathbb{H}; \boldsymbol{\eta}, \boldsymbol{\xi}))$$

Note that problem $\mathcal{P}$ is equivalent to $\mathcal{P}(\mathbb{F}, \mathbb{F}; \boldsymbol{\eta}, \boldsymbol{\xi})$ since $\mathbb{F}$ represents the filtration generated by $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$. The approximation scheme to be developed below will be based on a Lagrangian reformulation of the parametric optimization problems under consideration. To this end, we introduce the Lagrangian density $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^k \times \mathbb{R}^l \times \mathbb{T} \to \mathbb{R}$ associated with the problem data. It is defined through

$$L(x, y; \eta, \xi; t) := c(x, \eta, t) + y \cdot f(x, \xi, t)\,.$$

By the regularity condition (C1), L is convex in (x, ξ) and concave in (y, η). Next, introduce a linear space of *dual* decision processes $Y := \mathcal{L}^\infty_m(\mathbb{T} \times \Omega, \mathcal{M}, \lambda \otimes P)$. The following definition specifies a family of subsets of Y which will be relevant for the main result of this section (see Proposition 5.2).

Definition 5.1. *For any filtration $\mathbb{H} = \{\mathcal{H}_t\}_{t\in\mathbb{T}}$ on the sample space, let $Y_{\mathrm{c}}(\mathbb{H})$ be the collection of all $\mathcal{M}(\mathbb{H})$-measurable $\boldsymbol{y} \in Y$ which satisfy the box constraint $0 \le \boldsymbol{y}_t(\omega) \le \pi$ almost everywhere with respect to $\lambda \otimes P$.*

Proposition 5.2. *Let $\mathbb{G}$ and $\mathbb{H}$ be arbitrary filtrations on the sample space and assume that $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are arbitrary processes satisfying (C3). Then, we find*

$$\inf \mathcal{P}(\mathbb{G}, \mathbb{H}; \boldsymbol{\eta}, \boldsymbol{\xi}) \;=\; \inf_{\boldsymbol{x}\in X_{\mathrm{c}}(\mathbb{G})} \; \sup_{\boldsymbol{y}\in Y_{\mathrm{c}}(\mathbb{H})} \; \mathrm{E}\textstyle\int_0^T L(\boldsymbol{x}_t, \boldsymbol{y}_t; \boldsymbol{\eta}_t, \boldsymbol{\xi}_t; t)\, \lambda(\mathrm{d}t)\,.$$

Proof. The proof is a straightforward generalization of the argument in [27, § 4], which applies to the linear case in discrete-time. □

6 Stage-aggregation

As in the discrete time case considered in [19], we introduce two *aggregation operators* $\uparrow$ and $\downarrow$ mapping the index set $\mathbb{T}$ to itself. We refer to $\uparrow$ and $\downarrow$ as a pair of upper and lower aggregation operators if the following axioms are satisfied:

(i) monotonicity: both $\uparrow$ and $\downarrow$ are monotonically increasing;

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

(iii) ordering: $\downarrow \leq \mathrm{id} \leq \uparrow$.

Here, 'id' stands for the identity mapping on $\mathbb{T}$. As follows from the above axioms, 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). The basic axioms further imply that $\uparrow$ is left-continuous, while $\downarrow$ is right-continuous. A pair of aggregation operators with $2^J + 1$ evenly spaced fixed points ($J \in \mathbb{N}$), for instance, is obtained by setting

$$\uparrow(t) := \lceil t\, 2^J/T \rceil\, T/2^J \quad \text{and} \quad \downarrow(t) := \lfloor t\, 2^J/T \rfloor\, T/2^J \quad \forall\, t \in \mathbb{T}\,. \tag{6.1}$$

Next, let us introduce time-changed information sets $\mathcal{G}_t^{\uparrow} := \mathcal{F}_{\uparrow(t)}$ and $\mathcal{G}_t^{\downarrow} := \mathcal{F}_{\downarrow(t)}$ for all $t \in \mathbb{T}$, and define $\mathbb{G}^{\uparrow} := \{\mathcal{G}_t^{\uparrow}\}_{t \in \mathbb{T}}$ and $\mathbb{G}^{\downarrow} := \{\mathcal{G}_t^{\downarrow}\}_{t \in \mathbb{T}}$. Monotonicity of the aggregation operators implies that $\mathbb{G}^{\uparrow}$ and $\mathbb{G}^{\downarrow}$ are filtrations. By the ordering property (iii), the filtration $\mathbb{F}$ generated by the data process is a subfiltration of $\mathbb{G}^{\uparrow}$ in the sense that $\mathcal{F}_t \subset \mathcal{G}_t^{\uparrow}$ for each t. Moreover, $\mathbb{F}$ is a superfiltration of $\mathbb{G}^{\downarrow}$ in the sense that $\mathcal{F}_t \supset \mathcal{G}_t^{\downarrow}$ for each t. Intuitively speaking, $\mathbb{G}^{\downarrow}$ delays the information flow conveyed by $\mathbb{F}$, whereas $\mathbb{G}^{\uparrow}$ accelerates it: at any time t, the filtrations $\mathbb{G}^{\uparrow}$ and $\mathbb{G}^{\downarrow}$ provide full information about the path of the data process up to time $\uparrow(t)$ and $\downarrow(t)$, respectively. Thus, $\mathbb{G}^{\uparrow}$ implies that future events are foreseen to a limited extent since $\uparrow(t)$ is typically strictly larger than t. In contrast, $\mathbb{G}^{\downarrow}$ implies that past events are ignored to a certain extent since $\downarrow(t)$ is typically strictly smaller than t. The information flow associated with the time changed filtrations $\mathbb{G}^{\uparrow}$ and $\mathbb{G}^{\downarrow}$ is not continuous but bursty. It conveys information in shocks that occur at the (typically isolated) fixed points of the aggregation operators.

As in the case of the information processes, it proves useful to work with time-changed data processes. In fact, we let $\boldsymbol{\zeta}^{\uparrow}$ and $\boldsymbol{\zeta}^{\downarrow}$ be determined through $\boldsymbol{\zeta}^{\uparrow}_t := \boldsymbol{\zeta}_{\uparrow(t)}$ and $\boldsymbol{\zeta}^{\downarrow}_t := \boldsymbol{\zeta}_{\downarrow(t)}$, $t \in \mathbb{T}$, while the subordinate processes $\boldsymbol{\eta}^{\uparrow}$, $\boldsymbol{\xi}^{\uparrow}$, $\boldsymbol{\eta}^{\downarrow}$, and $\boldsymbol{\xi}^{\downarrow}$ are defined accordingly in the obvious way. Notice that the time-changed data processes $\boldsymbol{\zeta}^{\uparrow}$ and $\boldsymbol{\zeta}^{\downarrow}$ can be viewed as approximations for $\boldsymbol{\zeta}$, whose sample paths are constant except at the fixed points of the aggregation operators. The corresponding induced filtrations are constructed as usual, that is,

$$\mathcal{F}^{\uparrow}_t := \sigma(\boldsymbol{\zeta}^{\uparrow}_s | 0 \leq s \leq t) \quad \text{for} \quad t \in \mathbb{T}, \quad \mathcal{F}^{\uparrow} := \mathcal{F}^{\uparrow}_T, \quad \mathbb{F}^{\uparrow} := \{\mathcal{F}^{\uparrow}_t\}_{t\in\mathbb{T}}\,,$$
$$\mathcal{F}^{\downarrow}_t := \sigma(\boldsymbol{\zeta}^{\downarrow}_s | 0 \leq s \leq t) \quad \text{for} \quad t \in \mathbb{T}, \quad \mathcal{F}^{\downarrow} := \mathcal{F}^{\downarrow}_T, \quad \mathbb{F}^{\downarrow} := \{\mathcal{F}^{\downarrow}_t\}_{t\in\mathbb{T}}\,.$$

By construction, at any time t, the filtrations $\mathbb{F}^{\uparrow}$ and $\mathbb{F}^{\downarrow}$ provide full information about the data process at all fixed points of the aggregation operators up to time $\uparrow(t)$ and $\downarrow(t)$, respectively. The restriction to these fixed points implies that $\mathbb{F}^{\uparrow}$ is a subfiltration of $\mathbb{G}^{\uparrow}$, while $\mathbb{F}^{\downarrow}$ is a subfiltration of $\mathbb{G}^{\downarrow}$. We emphasize that, unlike $\mathbb{G}^{\uparrow}$ and $\mathbb{G}^{\downarrow}$, the filtrations $\mathbb{F}^{\uparrow}$ and $\mathbb{F}^{\downarrow}$ provide *no* information about the path of the data process between the fixed points of the aggregation operators.

Let us now elaborate a useful connection between the newly introduced filtrations, which relies on the Markov property of the data process. To motivate this result, fix some $t \in \mathbb{T}$, and assume that we want to predict the value of a random variable $\boldsymbol{\varphi}$ which depends on the path of the data process $\boldsymbol{\zeta}$ up to time $\downarrow(t)$. We also assume that we only have incomplete information about $\boldsymbol{\zeta}$, that is, we can observe $\boldsymbol{\zeta}$ only at times $\{\downarrow(s)\}_{s\leq t}$. Our information set thus corresponds to $\mathcal{F}^{\downarrow}_t$, and our best prediction of $\boldsymbol{\varphi}$ is $\mathrm{E}(\boldsymbol{\varphi}|\mathcal{F}^{\downarrow}_t)$. Lemma 6.1 below asserts that we cannot improve this prediction if $\boldsymbol{\zeta}$ becomes observable at all time points $\{\downarrow(s)\}_{s\leq T}$ and our information set increases to $\mathcal{F}^{\downarrow}$. This result follows from the Markov property of the data process (see assumption (C3)), which ensures that past and future values of $\boldsymbol{\zeta}$ are conditionally independent given the present value.

Lemma 6.1. *The following hold for all $t \in \mathbb{T}$:*

(i) $\mathrm{E}(\boldsymbol{\varphi}|\mathcal{F}_t^{\downarrow}) = \mathrm{E}(\boldsymbol{\varphi}|\mathcal{F}^{\downarrow})$ *P-a.s. for all* $\boldsymbol{\varphi} \in \mathcal{L}_1^{\infty}(\Omega, \mathcal{G}_t^{\downarrow}, P)$,

(ii) $\mathrm{E}(\boldsymbol{\varphi}|\mathcal{F}_t^{\uparrow}) = \mathrm{E}(\boldsymbol{\varphi}|\mathcal{F}^{\uparrow})$ *P-a.s. for all* $\boldsymbol{\varphi} \in \mathcal{L}_1^{\infty}(\Omega, \mathcal{G}_t^{\uparrow}, P)$.

Proof. Select an arbitrary $t \in \mathbb{T}$ and set

$$H := \{\boldsymbol{\varphi} \in \mathcal{L}_1^{\infty}(\Omega, \mathcal{G}_t^{\downarrow}, P) \,|\, \mathrm{E}(\boldsymbol{\varphi}|\mathcal{F}_t^{\downarrow}) = \mathrm{E}(\boldsymbol{\varphi}|\mathcal{F}^{\downarrow}) \ P\text{-a.s.}\}\,.$$

It is easily verified that H is a real vector space of bounded functions which contains all constant functions and is closed under uniform convergence. Moreover, if $\{\boldsymbol{\varphi}_i\}_{i\in\mathbb{N}}$ is a sequence of nonnegative functions in H which converges pointwise from below to some bounded $\boldsymbol{\varphi}$, then $\boldsymbol{\varphi} \in H$ (use the monotone convergence theorem for conditional expectations). Next, define

$$\begin{aligned} M := \{\boldsymbol{\varphi} = \textstyle\prod_{i=1}^{I} \boldsymbol{\varphi}_i \,|\, I \in \mathbb{N} \text{ and } \boldsymbol{\varphi}_i \in \mathcal{L}_1^{\infty}(\Omega, \sigma(\boldsymbol{\zeta}_{t_i}), P) \\ \text{with } 0 \le t_i \le \downarrow\!(t) \text{ for all } i = 1, \ldots, I\}\,. \end{aligned}$$

Note that M is closed under multiplication. Moreover, the σ-algebra generated by M coincides with $\mathcal{G}_t^{\downarrow} = \sigma(\boldsymbol{\zeta}_s | 0 \le s \le \downarrow\!(t))$. Let us next show that $M \subset H$. To this end, choose an arbitrary $\boldsymbol{\varphi} = \prod_{i=1}^{I} \boldsymbol{\varphi}_i \in M$. It is straightforward to verify that $\boldsymbol{\varphi}$ is bounded and $\mathcal{G}_t^{\downarrow}$-measurable since $\boldsymbol{\varphi}_i$ is bounded and $0 \le t_i \le \downarrow\!(t)$ for all $i = 1, \ldots, I$. Besides that, we have $\mathrm{E}(\boldsymbol{\varphi}|\mathcal{F}_t^{\downarrow}) = \mathrm{E}(\boldsymbol{\varphi}|\mathcal{F}^{\downarrow})$ P-a.s. since the sets of random vectors $\{\boldsymbol{\zeta}_{t_i}\}_{i=1}^{I}$ and $\{\boldsymbol{\zeta}_{\downarrow(\tau)}\}_{\downarrow(\tau)>\downarrow(t)}$ are conditionally independent given $\{\boldsymbol{\zeta}_{\downarrow(\tau)}\}_{\downarrow(\tau)\le\downarrow(t)}$. Conditional independence is a direct consequence of the Markovian nature of the data process $\boldsymbol{\zeta}$, see e.g. [19]. Therefore, $\boldsymbol{\varphi} \in H$, which implies that $M \subset H$. By the monotone class theorem [7, § I.21], we may thus conclude that H contains all bounded $\mathcal{G}_t^{\downarrow}$-measurable functions, that is, $H = \mathcal{L}_1^{\infty}(\Omega, \mathcal{G}_t^{\downarrow}, P)$. This observation establishes (i). The proof of (ii) is analogous. □

We can use Lemma 6.1 to show that conditional expectation with respect to $\mathcal{F}^{\downarrow}$ ($\mathcal{F}^{\uparrow}$) projects $\mathbb{G}^{\downarrow}$-adapted ($\mathbb{G}^{\uparrow}$-adapted) primal and dual strategies to $\mathbb{F}^{\downarrow}$-adapted ($\mathbb{F}^{\uparrow}$-adapted) primal and dual strategies, respectively. This result constitutes a main ingredient to establish aggregation bounds in Theorem 6.3 below.

Proposition 6.2. *If the data process is Markovian, then*

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

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

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

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

Proof. Choose $\boldsymbol{x} \in X_{\mathrm{c}}(\mathbb{G}^{\downarrow})$. Without loss of generality we assume that $\boldsymbol{x}$ is $\mathbb{G}^{\downarrow}$-adapted and set $\boldsymbol{x}^{\downarrow} := \mathrm{E}(\boldsymbol{x}|\mathcal{F}^{\downarrow})$. By the defining properties of conditional expectation processes (see Section 2) and by applying Lemma 6.1(i) to each $\boldsymbol{x}_t$, $t \in \mathbb{T}$, one easily shows that $\boldsymbol{x}^{\downarrow}$ is $\mathcal{M}(\mathbb{F}^{\downarrow})$-measurable.

Next, set $\delta_\Gamma(t,x) = 0$ for $x \in \Gamma(t)$, and $\delta_\Gamma(t,x) = +\infty$ otherwise. Since $\boldsymbol{x}_t \in \Gamma(t)$ except on a set of $\lambda \otimes P$ measure zero, we obtain

$$\begin{aligned} 0 &= \int_0^T \mathrm{E}(\delta_\Gamma(t,\boldsymbol{x}_t))\,\lambda(\mathrm{d}t) = \int_0^T \mathrm{E}(\mathrm{E}(\delta_\Gamma(t,\boldsymbol{x}_t)|\mathcal{F}^{\downarrow}))\,\lambda(\mathrm{d}t) \\ &\geq \int_0^T \mathrm{E}(\delta_\Gamma(t,\mathrm{E}(\boldsymbol{x}_t|\mathcal{F}^{\downarrow})))\,\lambda(\mathrm{d}t) = \int_0^T \mathrm{E}(\delta_\Gamma(t,\boldsymbol{x}_t^{\downarrow})) \geq 0\,. \end{aligned}$$

All integrals are well-defined, as follows from the argumentation in the proof of Proposition 3.4. The first inequality follows from the conditional Jensen inequality, which applies since Γ is convex-valued. As all inequalities must be binding, we conclude that $\boldsymbol{x}_t^{\downarrow} \in \Gamma(t)$ except on a set of $\lambda \otimes P$ measure zero. This also implies that $\boldsymbol{x}^{\downarrow}$ is $\lambda \otimes P$-essentially bounded. Furthermore, by using Fubini's theorem, one verifies that $\boldsymbol{x}_t^{\downarrow}$ satisfies the dynamic constraints specified in (3.1). This establishes (i). The assertions (ii)–(iv) are proved in a similar fashion. □

We are now ready to construct bounds on $\inf \mathcal{P}$. To this end, we rewrite $\mathcal{P}$ as a min-max problem (see Proposition 5.2) in which competing primal and a dual decision makers attempt to minimize and maximize the problem Lagrangian, respectively. Lower bounds are obtained by allowing the primal decision maker to foresee future observations some time ahead and by delaying the dual decision maker's information inflow. In doing so, we enlarge the primal feasible set from $X_{\rm c}(\mathbb{F})$ to $X_{\rm c}(\mathbb{G}^{\uparrow})$ and reduce the dual feasible set from $Y_{\rm c}(\mathbb{F})$ to $Y_{\rm c}(\mathbb{G}^{\downarrow})$. At the same time, the data processes $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are replaced by $\boldsymbol{\eta}^{\uparrow}$ and $\boldsymbol{\xi}^{\downarrow}$, respectively. By using Jensen's inequality, one can show that this substitution preserves the lower bounding property of the approximate problem. In a last step, we replace $X_{\rm c}(\mathbb{G}^{\uparrow})$ by $X_{\rm c}(\mathbb{F}^{\uparrow})$ and $Y_{\rm c}(\mathbb{G}^{\downarrow})$ by $Y_{\rm c}(\mathbb{F}^{\downarrow})$. Since the Markovian data processes are memoryless, this reduction of the primal and dual feasible sets has no effect on the optimal value of the lower bounding problem. Upper bounds are obtained analogously. The following theorem formalizes the above reasoning.

Theorem 6.3. *If the data process represents a Markov martingale, then*

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

Proof. With Proposition 6.2 at our disposal, this theorem becomes a simple corollary of the Theorems 3 and 4 in [19], which cover a similar result in discrete time. The only difference is that here we have to invoke Fubini's theorem to interchange conditional expectations and time integrals. For the sake of brevity, we merely outline the proof idea and refer to [19] for the details.

By using the relation $\mathbb{G}^{\downarrow} \subset \mathbb{F} \subset \mathbb{G}^{\uparrow}$, the martingale property of the data process, and the conditional Jensen inequality, one first shows that

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

Next, one employs Proposition 6.2 to derive the equalities

$$\inf \mathcal{P}(\mathbb{G}^{\downarrow}, \mathbb{G}^{\uparrow}; \boldsymbol{\eta}^{\downarrow}, \boldsymbol{\xi}^{\uparrow}) = \inf \mathcal{P}(\mathbb{F}^{\downarrow}, \mathbb{F}^{\uparrow}; \boldsymbol{\eta}^{\downarrow}, \boldsymbol{\xi}^{\uparrow}) , \tag{6.2a}$$

$$\inf \mathcal{P}(\mathbb{G}^{\uparrow}, \mathbb{G}^{\downarrow}; \boldsymbol{\eta}^{\uparrow}, \boldsymbol{\xi}^{\downarrow}) = \inf \mathcal{P}(\mathbb{F}^{\uparrow}, \mathbb{F}^{\downarrow}; \boldsymbol{\eta}^{\uparrow}, \boldsymbol{\xi}^{\downarrow}) , \tag{6.2b}$$

and thus the claim follows. ☐

Remark 6.4. *A crucial idea of our aggregation scheme is to distinguish primal and dual information processes, which are either coarsened or refined when constructing bounds. In a discrete-time setting, Wright has studied situations in which aggregation bounds can be obtained by only coarsening the information processes, while the data processes are not (explicitly) modified [27]. A related aggregation scheme was suggested by Casey and Sen [5]. Both approaches are likely to carry over to the continuous-time setting, but their applicability seems to be restricted to special problem structures with random parameters only appearing either in the objective or in the constraints. Moreover, the resulting fully aggregated problems only provide either lower or upper bounds.*

If we select aggregation operators with a finite number of fixed points, then the approximate optimization problems of Theorem 6.3 involve finitely many different (scalar) random parameters. Since only the distribution of these parameters has practical relevance, the approximate problems can principally be formulated on a finite-dimensional probability space. Note, however, that the original problem $\mathcal{P}$ is generically built on an infinite-dimensional probability space.

By reducing the dimensionality of the probability space from infinite to finite, stage-aggregation greatly simplifies problem $\mathcal{P}$. The resulting approximate problems represent multistage stochastic programs with finitely many decision stages (given by the fixed points of the aggregation operators). Although the number of stages is now finite, the decision vectors in each stage and scenario are still functions of time, thus having infinite dimension. Simple heuristics, such as freezing or linearizing the decisions on short time intervals, can reduce the dimensionality of these decision spaces. However, such extra restrictions reduce the flexibility of the decision maker and thus provide upper bounds for the approximate problems. Imposing similar restrictions on certain dual versions of the approximate problems can give lower bounds. As it is computationally cheap to maintain a high number of decision variables per stage and scenario, the corresponding approximation error can usually be kept small. Under certain conditions, restricting the decisions to be piecewise constant or linear functions of time has no influence on the optimal values of the approximate problems at all (see Appendix B).

Once a finite parametrization of the decisions in each stage and scenario has been established, the approximate problems can be addressed via standard stochastic programming techniques, which rely on discretizing the uncountable (but finite-dimensional) probability space [4]. Discretization is performed, for example, by using techniques developed in [20] and amounts to retaining only a finite number of scenarios, which are arranged in a tree structure.

A systematic description of the additional approximations complementing stage-aggregation is beyond the scope of this article. It should be emphasized, however, that stage-aggregation constitutes the key approximation since it is directly aimed at breaking the exponential growth of the arising scenario tree.

7 Convergence

Consider now two sequences of aggregation operators $\{\uparrow_J\}_{J\in\mathbb{N}}$ and $\{\downarrow_J\}_{J\in\mathbb{N}}$ such that each pair $(\uparrow_J, \downarrow_J)$ satisfies the axioms of monotonicity, idempotence, and ordering. Below, we will always assume that these sequences are monotonic, that is, $\uparrow_{J+1} \leq \uparrow_J$ and $\downarrow_{J+1} \geq \downarrow_J$ for all $J \in \mathbb{N}$. Moreover, we also assume that $\{\uparrow_J\}_{J\in\mathbb{N}}$ and $\{\downarrow_J\}_{J\in\mathbb{N}}$ converge uniformly to the identity mapping on $\mathbb{T}$. Observe that uniform convergence is implied by pointwise convergence and that convergence of $\{\uparrow_J\}_{J\in\mathbb{N}}$ implies convergence of $\{\downarrow_J\}_{J\in\mathbb{N}}$ and vice versa. This follows from the defining properties of the aggregation operators.

For each $J \in \mathbb{N}$, we define the stochastic processes $\boldsymbol{\zeta}^{\uparrow_J}$ and $\boldsymbol{\zeta}^{\downarrow_J}$ as well as the filtrations $\mathbb{G}^{\uparrow_J}$, $\mathbb{G}^{\downarrow_J}$, $\mathbb{F}^{\uparrow_J}$, and $\mathbb{F}^{\downarrow_J}$ in the usual way as in Section 6. In order to avoid the frequent use of cumbersome double sub- and superscripts, we will denote the aggregated stochastic processes by $\boldsymbol{\zeta}^{\uparrow}_J := \boldsymbol{\zeta}^{\uparrow_J}$ and $\boldsymbol{\zeta}^{\downarrow}_J := \boldsymbol{\zeta}^{\downarrow_J}$, while the aggregated filtrations will be represented as $\mathbb{G}_J^{\uparrow} := \mathbb{G}^{\uparrow_J}$, $\mathbb{G}_J^{\downarrow} := \mathbb{G}^{\downarrow_J}$, $\mathbb{F}_J^{\uparrow} := \mathbb{F}^{\uparrow_J}$, and $\mathbb{F}_J^{\downarrow} := \mathbb{F}^{\downarrow_J}$. This means that $\mathbb{G}_J^{\uparrow} := \{\mathcal{G}_{J,t}^{\uparrow}\}_{t\in\mathbb{T}}$, where $\mathcal{G}_{J,t}^{\uparrow} := \mathcal{G}_t^{\uparrow_J}$, etc.

We will now show that the approximate optimization problems of Section 6 converge to the original problem $\mathcal{P}$ as the refinement parameter J tends to infinity. We split the convergence proof into a series of simpler lemmas and propositions, first investigating the asymptotic properties of the primal and dual feasible sets. This motivates us to consider the weak* topology on the dual decision space Y. Like in the case of the primal decision space X, the weak* topology on Y is the coarsest topology preserving continuity of the linear functionals

$$\boldsymbol{y} \mapsto \mathrm{E} \textstyle\int_0^T \boldsymbol{y}_t \cdot \boldsymbol{u}_t \, \lambda(\mathrm{d}t) \quad \text{where} \quad \boldsymbol{u} \in \mathcal{L}_m^1(\mathbb{T}\times\Omega, \mathcal{M}, \lambda\otimes P)\,.$$

Adapting the arguments of Section 3, one can show that the weak* topology on Y is metrizable on bounded sets and that $Y_\mathrm{c}(\mathbb{H})$ is nonempty and weak* compact for any filtration $\mathbb{H}$ on the sample space.

Proposition 7.1. *For $Z \in \{X, Y\}$ we find*

(i) $Z_\mathrm{c}(\mathbb{F}) = \bigcap_{J=1}^{\infty} Z_\mathrm{c}(\mathbb{G}_J^{\uparrow})$,

(ii) $Z_\mathrm{c}(\mathbb{F}) = \text{weak*-cl}\left(\bigcup_{J=1}^{\infty} Z_\mathrm{c}(\mathbb{G}_J^{\downarrow})\right)$,

where 'weak*-cl' *denotes closure with respect to the weak* topology on Z.*

Proof. For brevity of exposition, we consider only the case $Z = X$. The case $Z = Y$ requires no additional ideas. As for statement (i), it suffices to show that the intersection of all $X_\mathrm{c}(\mathbb{G}_J^{\uparrow})$ is a subset of $X_\mathrm{c}(\mathbb{F})$. The converse inclusion trivially holds since $X_\mathrm{c}(\mathbb{F}) \subset X_\mathrm{c}(\mathbb{G}_J^{\uparrow})$ for all $J \in \mathbb{N}$. Thus, we select some $\boldsymbol{x} \in \bigcap_{J=1}^{\infty} X_\mathrm{c}(\mathbb{G}_J^{\uparrow})$. This policy satisfies the constraints (3.1) because $\boldsymbol{x}$ is certainly contained in $X_\mathrm{c}(\mathbb{G}_1^{\uparrow})$, all of whose elements satisfy (3.1) by definition. By changing it on a set of $\lambda\otimes P$ measure zero, if necessary, we may assume that $\boldsymbol{x}$ is $\mathcal{T}\otimes\mathcal{A}$-measurable. In addition, for any fixed $J \in \mathbb{N}$ the process $\boldsymbol{x}$ is $\mathcal{M}(\mathbb{G}_J^{\uparrow})$-measurable. Put differently, for λ-almost all $t \in \mathbb{T}$ the random vector $\boldsymbol{x}_t$ is measurable with respect to the P-completion of the σ-algebra $\mathcal{G}^{\uparrow}_{J,t}$. Since this statement holds true for all $J \in \mathbb{N}$, we may conclude that $\boldsymbol{x}_t$ is $\mathcal{F}^+_t$-measurable for λ-almost all $t \in \mathbb{T}$, where $\mathcal{F}^+_t$ denotes the P-completion of the σ-algebra $\bigcap_{J=1}^{\infty} \mathcal{G}^{\uparrow}_{J,t}$. Moreover, as $\uparrow_J(t) \searrow t$ for $J \to \infty$, we can readily identify $\mathbb{F}^+ := \{\mathcal{F}^+_t\}_{t\in\mathbb{T}}$ as the right-continuous completion of the filtration $\mathbb{F}$, see [7, § IV.48]. Thus, $\boldsymbol{x}$ is $\mathbb{F}^+$-adapted on $\mathbb{T}^0 \times \Omega$, where $\mathbb{T}^0$ is some $\mathcal{T}$-measurable set with $\lambda(\mathbb{T}) = \lambda(\mathbb{T}^0)$. Since $\mathbb{T}^0 \times \Omega$ is $\mathcal{T} \otimes \mathcal{A}$-measurable, $\boldsymbol{x}$ is almost everywhere equivalent to some $\mathcal{T} \otimes \mathcal{A}$-measurable process adapted to $\mathbb{F}^+$. By [6, Theorem 3.6] there exists an $\mathbb{F}^+$-predictable process $\hat{\boldsymbol{x}} \sim \boldsymbol{x}$, and by [8, Lemma 7, p. 413] there exists an $\mathbb{F}$-predictable (and, a fortiori, $\mathbb{F}$-adapted) process $\boldsymbol{x}' \sim \hat{\boldsymbol{x}}$. In conclusion, $\boldsymbol{x}$ is almost everywhere equivalent to some $\mathcal{T} \otimes \mathcal{A}$-measurable process adapted to $\mathbb{F}$, thus being $\mathcal{M}(\mathbb{F})$-measurable. Collecting the above results, we find that $\boldsymbol{x} \in X_{\mathrm{c}}(\mathbb{F})$.

As for (ii), notice that $X_{\mathrm{c}}(\mathbb{F}) \supset \text{weak*-cl}(\bigcup_{J=1}^{\infty} X_{\mathrm{c}}(\mathbb{G}^{\downarrow}_J))$ is automatically satisfied since each $X_{\mathrm{c}}(\mathbb{G}^{\downarrow}_J)$ is contained in $X_{\mathrm{c}}(\mathbb{F})$, the latter being closed with respect to the weak* topology on X. By definition, any $\boldsymbol{x} \in X_{\mathrm{c}}(\mathbb{F})$ satisfies the constraints (3.1), and thus it remains to be shown that $X(\mathbb{F}) \subset H$, where

$$H := \text{weak*-cl}\left(\bigcup_{J=1}^{\infty} X(\mathbb{G}^{\downarrow}_J)\right).$$

We now proceed with a monotone class argument. It is easily verified that H is a real vector space of bounded functions which contains the constant functions. As strong convergence implies weak* convergence, H is closed under uniform limits. Moreover, if $\{\boldsymbol{x}^i\}_{i\in\mathbb{N}}$ is a sequence of nonnegative elements of H which converges pointwise from below to some bounded $\boldsymbol{x}$, then, by the dominated convergence theorem, $\{\boldsymbol{x}^i\}_{i\in\mathbb{N}}$ converges to $\boldsymbol{x}$ also with respect to the weak* topology on X. Thus, we have $\boldsymbol{x} \in H$. For the further argumentation, we define

$$\mathbb{T}' := \bigcup_{J=1}^{\infty} \{t \in \mathbb{T} \mid \downarrow_J(t) = t = \uparrow_J(t)\}.$$

Uniform convergence of $\{\uparrow_J\}_{J\in\mathbb{N}}$ and $\{\downarrow_J\}_{J\in\mathbb{N}}$ to the identity mapping implies that $\mathbb{T}'$ is dense in $\mathbb{T}$. Next, we define

$$M := \{\boldsymbol{x} \in X \,|\, \boldsymbol{x}_t(\omega) = 1_{[r,s)}(t)\, 1_A(\omega)\ \lambda \otimes P\text{-a.e. for } r, s \in \mathbb{T}';\ A \in \mathcal{F}_r\}\,,$$

which is closed under multiplication. In order to prove $M \subset H$, we select an arbitrary $\boldsymbol{x} \in M$, which is determined by $r, s \in \mathbb{T}'$ and $A \in \mathcal{F}_r$. By construction of $\mathbb{T}'$, there are $J_r, J_s \in \mathbb{N}$ such that r is a fixed point of $\downarrow_{J_r}$ while s is a fixed point of $\downarrow_{J_s}$. As the lower aggregation operators are monotonically increasing with J, we may conclude that r and s are fixed points of $\downarrow_J$ for all $J \geq \max\{J_r, J_s\}$. It is then easy to see that $\boldsymbol{x} \in X(\mathbb{G}_J^{\downarrow})$ for all J large enough, which implies that $\boldsymbol{x} \in H$. Next, observe that the σ-algebra generated by M coincides with $\mathcal{M}(\mathbb{F})$. In fact, the sets of the form $[r, s) \times A$ with $r, s \in \mathbb{T}'$ and $A \in \mathcal{F}_r$ generate the $\mathbb{F}$-predictable σ-field, whose $\lambda \otimes P$-completion is given by $\mathcal{M}(\mathbb{F})$, see e.g. [8, § IV.67] and [6, Theorem 3.6]. By the monotone class theorem [7, § I.21], we may thus conclude that $X(\mathbb{F}) \subset H$. This observation completes the proof. $\square$

Proposition 7.1 is an important ingredient for the next Lemma, which uses epi-convergence techniques to prove asymptotic consistency of certain 'monotonic' sequences of stochastic optimization problems.

Lemma 7.2. *Let g_J, $J \in \mathbb{N}$, and g be convex weak* lsc functionals from Z to $\mathbb{R}$, where $Z \in \{X, Y\}$. Moreover, assume that $g_J \to g$ pointwise as $J \to \infty$.*

(i) If $g_{J+1} \geq g_J$ for all $J \in \mathbb{N}$, then $\displaystyle\lim_{J\to\infty} \inf_{\boldsymbol{z}\in Z_{\mathrm{c}}(\mathbb{G}_J^{\uparrow})} g_J(\boldsymbol{z}) = \inf_{\boldsymbol{z}\in Z_{\mathrm{c}}(\mathbb{F})} g(\boldsymbol{z})\,.$

(ii) If $g_{J+1} \leq g_J$ for all $J \in \mathbb{N}$, then $\displaystyle\lim_{J\to\infty} \inf_{\boldsymbol{z}\in Z_{\mathrm{c}}(\mathbb{G}_J^{\downarrow})} g_J(\boldsymbol{z}) = \inf_{\boldsymbol{z}\in Z_{\mathrm{c}}(\mathbb{F})} g(\boldsymbol{z})\,.$

Proof. Without loss of generality, we may focus on the case $Z = X$. Assume first that $\{g_J\}_{J\in\mathbb{N}}$ is monotonically increasing, and let γ_J be the extended-real-valued functional defined through $\gamma_J(\boldsymbol{x}) = g_J(\boldsymbol{x})$ for $\boldsymbol{x} \in X_{\mathrm{c}}(\mathbb{G}_J^{\uparrow})$; $= +\infty$ otherwise. Note that γ_J is convex and weak* lsc as follows from Proposition 3.4. Similarly, define a convex weak* lsc functional γ through $\gamma(\boldsymbol{x}) = g(\boldsymbol{x})$ for $\boldsymbol{x} \in X_{\mathrm{c}}(\mathbb{F})$; $= +\infty$

otherwise. By Proposition 7.1(i), $\{\operatorname{epi}\gamma_J\}_{J\in\mathbb{N}}$ represents a decreasing sequence of convex closed[7] epigraphs with

$$\lim_{J\to\infty}\operatorname{epi}\gamma_J = \bigcap_{J=1}^{\infty}\operatorname{epi}\gamma_J = \operatorname{epi}\gamma\,,$$

the set limit being taken in the sense of Painlevé-Kuratowski. Thus, γ_J epi-converges to γ. Next, let $\boldsymbol{x}_J$ be a minimizer of γ_J. Note that $\boldsymbol{x}_J$ exists because γ_J is weak* lsc, and the effective domain of γ_J is weak* compact. As $\{\boldsymbol{x}_J\}_{J\in\mathbb{N}}\subset X_{\mathrm{c}}(\mathbb{G}_1^{\uparrow})$ has a weak* accumulation point, we conclude via [1, Theorem 2.5] that

$$\lim_{J\to\infty}\inf_{\boldsymbol{x}\in X_{\mathrm{c}}(\mathbb{G}_J^{\uparrow})} g_J(\boldsymbol{x}) = \lim_{J\to\infty}\inf_{\boldsymbol{x}\in X}\gamma_J(\boldsymbol{x}) = \inf_{\boldsymbol{x}\in X}\gamma(\boldsymbol{x}) = \inf_{\boldsymbol{x}\in X_{\mathrm{c}}(\mathbb{F})} g(\boldsymbol{x})\,.$$

This establishes (i). Assume next that $\{g_J\}_{J\in\mathbb{N}}$ is decreasing. While leaving γ unchanged, we redefine γ_J as follows: set $\gamma_J(\boldsymbol{x}) = g_J(\boldsymbol{x})$ for $\boldsymbol{x}\in X_{\mathrm{c}}(\mathbb{G}_J^{\downarrow})$; $=+\infty$ otherwise. Note that γ_J is still convex and weak* lsc. By Proposition 7.1(ii), $\{\operatorname{epi}\gamma_J\}_{J\in\mathbb{N}}$ now represents an increasing sequence of convex closed epigraphs with

$$\lim_{J\to\infty}\operatorname{epi}\gamma_J = \operatorname{cl}\bigcup_{J=1}^{\infty}\operatorname{epi}\gamma_J = \operatorname{epi}\gamma\,.$$

As before, γ_J epi-converges to γ. We let $\boldsymbol{x}_J$ again denote a minimizer of γ_J. As the sequence $\{\boldsymbol{x}_J\}_{J\in\mathbb{N}}\subset X_{\mathrm{c}}(\mathbb{F})$ has a weak* accumulation point, we conclude via [1, Theorem 2.5] that

$$\lim_{J\to\infty}\inf_{\boldsymbol{x}\in X_{\mathrm{c}}(\mathbb{G}_J^{\downarrow})} g_J(\boldsymbol{x}) = \lim_{J\to\infty}\inf_{\boldsymbol{x}\in X}\gamma_J(\boldsymbol{x}) = \inf_{\boldsymbol{x}\in X}\gamma(\boldsymbol{x}) = \inf_{\boldsymbol{x}\in X_{\mathrm{c}}(\mathbb{F})} g(\boldsymbol{x})\,.$$

This establishes assertion (ii). □

Armed with the convergence results of Lemma 7.2, we are now in a favorable position to study the asymptotic properties of some approximate optimization problems which preserve the original data process, but whose primal and dual information processes are given by $\mathbb{G}_J^{\downarrow}$ and $\mathbb{G}_J^{\uparrow}$, respectively (or vice versa).

[7] As in the proof of Proposition 4.1, we assume that $X\times\mathbb{R}$ is endowed with the product of the weak* topology on X and the Euclidean topology on $\mathbb{R}$.

Proposition 7.3. *If J tends to infinity, then*

(i) $\inf \mathcal{P}(\mathbb{G}_J^{\downarrow}, \mathbb{G}_J^{\uparrow}; \boldsymbol{\eta}, \boldsymbol{\xi}) \to \inf \mathcal{P}$*;*

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

Proof. Consider the mapping

$$\ell : X \times Y \to \mathbb{R}\,, \quad \ell(\boldsymbol{x}, \boldsymbol{y}) := \mathrm{E} \int_0^T L(\boldsymbol{x}_t, \boldsymbol{y}_t; \boldsymbol{\eta}_t, \boldsymbol{\xi}_t; t)\, \lambda(\mathrm{d}t)\,.$$

By the special structure of the Lagrangian density, $\ell(\boldsymbol{x}, \cdot)$ is the sum of a bounded linear functional in $\boldsymbol{y}$ and a constant for every $\boldsymbol{x} \in X$, and it is easily shown to be weak* continuous. Moreover, $\ell(\cdot, \boldsymbol{y})$ is convex and weak* lsc in $\boldsymbol{x}$ for every nonnegative $\boldsymbol{y} \in Y$. Weak* lower semicontinuity follows from [24, Corollary 2A], which applies due to (C2), (C3), and the essential boundedness of $\boldsymbol{y}$. The aforementioned regularity conditions also imply that ℓ is in fact finite on its entire domain.

For every $J \in \mathbb{N}$, introduce the optimal value functionals

$$g_J : X \to \mathbb{R}\,, \quad g_J(\boldsymbol{x}) := \sup_{\boldsymbol{y} \in Y_{\mathrm{c}}(\mathbb{G}_J^{\uparrow})} \ell(\boldsymbol{x}, \boldsymbol{y})\,, \tag{7.1}$$

and define

$$g : X \to \mathbb{R}\,, \quad g(\boldsymbol{x}) := \sup_{\boldsymbol{y} \in Y_{\mathrm{c}}(\mathbb{F})} \ell(\boldsymbol{x}, \boldsymbol{y})\,. \tag{7.2}$$

As $\ell(\boldsymbol{x}, \cdot)$ is weak* continuous and $Y_{\mathrm{c}}(\mathbb{G}_J^{\uparrow})$ is weak* compact, the maximum in (7.1) is attained. Finiteness of ℓ then implies that g_J is finite on its domain. Observe also that g_J is convex and weak* lsc since it is the supremum of (uncountably many) convex weak* lsc functions. A similar reasoning shows that g is convex, weak* lsc, and finite on its domain.

The inclusion $Y_{\mathrm{c}}(\mathbb{G}_{J+1}^{\uparrow}) \subset Y_{\mathrm{c}}(\mathbb{G}_J^{\uparrow})$ implies that $g_{J+1} \le g_J$ for all $J \in \mathbb{N}$, that is, $\{g_J\}_{J \in \mathbb{N}}$ constitutes a decreasing sequence. By Lemma 7.2(i), this sequence converges pointwise to g. Collecting the above results, we conclude that

$$\lim_{J\to\infty} \inf \mathcal{P}(\mathbb{G}_J^{\downarrow}, \mathbb{G}_J^{\uparrow}; \boldsymbol{\eta}, \boldsymbol{\xi}) = \lim_{J\to\infty} \inf_{\boldsymbol{x} \in X_{\mathrm{c}}(\mathbb{G}_J^{\downarrow})} g_J(\boldsymbol{x}) = \inf_{\boldsymbol{x} \in X_{\mathrm{c}}(\mathbb{F})} g(\boldsymbol{x}) = \inf \mathcal{P}\,,$$

where the second equality follows from Lemma 7.2(ii). This establishes assertion (i). The proof of (ii) is widely parallel and will thus be omitted. □

Even though Proposition 7.3 provides valuable insights into the convergence behavior of optimization problems with time-changed information processes, we need additional results that include also the effects of the time-changed data processes. The following exposition works towards this goal (see Lemmas 7.4 and 7.5 as well as Proposition 7.6). In a first step, we investigate the continuity properties of the data process.

Lemma 7.4. *Under the standard assumption that the sample paths of $\boldsymbol{\zeta}$ are rcll, we have $\boldsymbol{\zeta}^- \sim \boldsymbol{\zeta} \sim \boldsymbol{\zeta}^+$, where $\boldsymbol{\zeta}^\pm$ is defined through $\boldsymbol{\zeta}^\pm_t := \lim_{\varepsilon \downarrow 0} \boldsymbol{\zeta}_{t\pm\varepsilon}$, respectively.*

Proof. Observe first that $\boldsymbol{\zeta}^-$ inherits $\mathcal{M}$-measurability from $\boldsymbol{\zeta}$. For the further argumentation, we need the $\mathcal{M}$-measurable sets

$$C_i := \{(t,\omega) \in \mathbb{T} \times \Omega \,|\, |\boldsymbol{\zeta}_t(\omega) - \boldsymbol{\zeta}^-_t(\omega)| > i^{-1}\}, \quad i \in \mathbb{N}.$$

Fix $\omega \in \Omega$ and assume that the section $C_i(\omega) := \{t \in \mathbb{T} | (t,\omega) \in C_i\}$ has an accumulation point at t^*. As the path of $\boldsymbol{\zeta}$ corresponding to ω is rcll, there exists for each $\varepsilon > 0$ a $\delta > 0$ such that

$$\begin{aligned} \boldsymbol{\zeta}_t(\omega), \boldsymbol{\zeta}^-_t(\omega) &\in B_\varepsilon(\boldsymbol{\zeta}^-_{t^*}(\omega)) \ \text{ for all } \ t \in (t^* - \delta, t^*)\,, \\ \boldsymbol{\zeta}_t(\omega), \boldsymbol{\zeta}^-_t(\omega) &\in B_\varepsilon(\boldsymbol{\zeta}_{t^*}(\omega)) \ \text{ for all } \ t \in (t^*, t^* + \delta)\,. \end{aligned}$$

For $\varepsilon < \frac{1}{2i}$ this contradicts the assumption that $(t^* - \delta, t^*) \cup (t^*, t^* + \delta)$ contains infinitely many points in $C_i(\omega)$. We conclude that $C_i(\omega)$ has no accumulation point at all, thus being finite. Fubini's theorem then implies that C_i has $\lambda \otimes P$ measure zero for each $i \in \mathbb{N}$. Hence, $\bigcup_{i=1}^\infty C_i$ is a $\lambda \otimes P$ null set, as a consequence of which we find $\boldsymbol{\zeta} \sim \boldsymbol{\zeta}^-$. As the sample paths of $\boldsymbol{\zeta}$ are right-continuous, the process $\boldsymbol{\zeta}^+$ is in fact equal to $\boldsymbol{\zeta}$. Obviously, this implies $\boldsymbol{\zeta}^+ \sim \boldsymbol{\zeta}$. □

Lemma 7.5. *For any filtration $\mathbb{G}$ on the sample space let $\{\boldsymbol{x}_J\}_{J\in\mathbb{N}}$ and $\{\boldsymbol{y}_J\}_{J\in\mathbb{N}}$ be arbitrary sequences in $X_\mathrm{c}(\mathbb{G})$ and $Y_\mathrm{c}(\mathbb{G})$, respectively. Then, we have* *(i)* $\lim_{J\to\infty} |c(\boldsymbol{x}_{J,t}, \boldsymbol{\eta}^{\downarrow}_{J,t}, t) - c(\boldsymbol{x}_{J,t}, \boldsymbol{\eta}_t, t)| = 0 \quad \lambda \otimes P\text{-}a.e.,$

(ii) $\lim_{J\to\infty} |\boldsymbol{y}_{J,t}|\, |f_2(\boldsymbol{\xi}^{\uparrow}_{J,t}, t) - f_2(\boldsymbol{\xi}_t, t)| = 0 \quad \lambda \otimes P\text{-}a.e.$

Proof. By Lemma 7.4 and since $\boldsymbol{x}_J$ satisfies (3.1), the $\mathcal{M}$-measurable set

$$A := \{(t,\omega) \in \mathbb{T} \times \Omega \,|\, \lim_{J\to\infty} \boldsymbol{\eta}^{\downarrow}_{J,t}(\omega) = \boldsymbol{\eta}^{-}_t(\omega) = \boldsymbol{\eta}_t(\omega)\,,\ \boldsymbol{x}_{J,t}(\omega) \in \Gamma_t \;\forall\, J \in \mathbb{N}\}$$

differs from $\mathbb{T} \times \Omega$ at most by a set of $\lambda \otimes P$-measure zero. Fix some $(t,\omega) \in A$ and choose a tolerance $\varepsilon > 0$. Since $c(x,\cdot,t)$ is concave and thus continuous, there exists for each $x \in \mathbb{R}^n$ some $J_x \in \mathbb{N}$ with

$$|c(x, \boldsymbol{\eta}^{\downarrow}_{J,t}(\omega), t) - c(x, \boldsymbol{\eta}_t(\omega), t)| < \varepsilon \tag{7.3}$$

for all $J \geq J_x$. Furthermore, since $c(\cdot,\eta,t)$ is convex and thus continuous, there exists for each $x' \in \mathbb{R}^n$ an open set $N_{x'} \ni x'$ such that (7.3) holds for all $J \geq J_{x'}$ and $x \in N_{x'}$. Recall now that Γ_t is compact. Hence, there is a finite collection of points $x^i \in \Gamma_t$, $i = 1, \ldots, I$, with the property that $\Gamma_t \subset \bigcup_{i=1}^{I} N_{x^i}$. We may thus conclude that (7.3) holds for all $J \geq \max\{J_{x^i} | i = 1, \ldots, I\}$ and $x \in \Gamma_t$. As $\varepsilon > 0$ was chosen freely, the above argument implies that

$$\lim_{J\to\infty} |c(\boldsymbol{x}_{J,t}(\omega), \boldsymbol{\eta}^{\downarrow}_{J,t}(\omega), t) - c(\boldsymbol{x}_{J,t}(\omega), \boldsymbol{\eta}_t(\omega), t)| = 0$$

for each $(t,\omega) \in A$. This establishes (i). The proof of assertion (ii) is straightforward if one recalls that $f_2(\cdot,t)$ is convex continuous and that $0 \leq \boldsymbol{y}_{J,t} \leq \pi$ almost everywhere with respect to $\lambda \otimes P$ and uniformly over all $J \in \mathbb{N}$. □

The subsidiary Lemma 7.5 finally enables us to prove the following key result.

Proposition 7.6. *If J tends to infinity, then*

(i) $\inf \mathcal{P}(\mathbb{G}^{\downarrow}_J, \mathbb{G}^{\uparrow}_J; \boldsymbol{\eta}^{\downarrow}_J, \boldsymbol{\xi}^{\uparrow}_J) - \inf \mathcal{P}(\mathbb{G}^{\downarrow}_J, \mathbb{G}^{\uparrow}_J; \boldsymbol{\eta}, \boldsymbol{\xi}) \to 0;$

(ii) $\inf \mathcal{P}(\mathbb{G}^{\uparrow}_J, \mathbb{G}^{\downarrow}_J; \boldsymbol{\eta}^{\uparrow}_J, \boldsymbol{\xi}^{\downarrow}_J) - \inf \mathcal{P}(\mathbb{G}^{\uparrow}_J, \mathbb{G}^{\downarrow}_J; \boldsymbol{\eta}, \boldsymbol{\xi}) \to 0.$

Proof. For each $J \in \mathbb{N}$ set

$$\ell_J : X \times Y \to \mathbb{R}\,, \quad \ell_J(\boldsymbol{x},\boldsymbol{y}) := \mathrm{E}\int_0^T L(\boldsymbol{x}_t,\boldsymbol{y}_t;\boldsymbol{\eta}^{\downarrow}_{J,t},\boldsymbol{\xi}^{\uparrow}_{J,t};t)\,\lambda(\mathrm{d}t)\,,$$

and let ℓ be defined as in the proof of Proposition 7.3. It is easily seen that the integrability conditions put forth in (C3) remain valid if $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are replaced by $\boldsymbol{\eta}^{\downarrow}_J$ and $\boldsymbol{\xi}^{\uparrow}_J$, respectively. Therefore, ℓ_J has the same continuity properties as ℓ, that is, $\ell_J(\cdot,\boldsymbol{y})$ is weak* lsc in $\boldsymbol{x}$ for all $\boldsymbol{y} \in Y$, while $\ell_J(\boldsymbol{x},\cdot)$ is weak* continuous in $\boldsymbol{y}$ for all $\boldsymbol{x} \in X$. Assume next that $\boldsymbol{x}_J \in X_{\mathrm{c}}(\mathbb{G}^{\downarrow}_J)$ represents an optimal solution[8] of problem $\mathcal{P}(\mathbb{G}^{\downarrow}_J,\mathbb{G}^{\uparrow}_J;\boldsymbol{\eta}^{\downarrow}_J,\boldsymbol{\xi}^{\uparrow}_J)$. As $\boldsymbol{x}_J$ is feasible in $\mathcal{P}(\mathbb{G}^{\downarrow}_J,\mathbb{G}^{\uparrow}_J;\boldsymbol{\eta},\boldsymbol{\xi})$, we obtain

$$\begin{aligned} &\inf \mathcal{P}(\mathbb{G}^{\downarrow}_J,\mathbb{G}^{\uparrow}_J;\boldsymbol{\eta}^{\downarrow}_J,\boldsymbol{\xi}^{\uparrow}_J) - \inf \mathcal{P}(\mathbb{G}^{\downarrow}_J,\mathbb{G}^{\uparrow}_J;\boldsymbol{\eta},\boldsymbol{\xi}) \\ &\quad\leq \sup_{\boldsymbol{y}\in Y_{\mathrm{c}}(\mathbb{G}^{\uparrow}_J)} \ell_J(\boldsymbol{x}_J,\boldsymbol{y}) - \sup_{\boldsymbol{y}\in Y_{\mathrm{c}}(\mathbb{G}^{\uparrow}_J)} \ell(\boldsymbol{x}_J,\boldsymbol{y})\,. \end{aligned}$$

Consider now the two maximization problems in the second line of the above expression. If $\boldsymbol{y}_J \in Y_{\mathrm{c}}(\mathbb{G}^{\uparrow}_J)$ is optimal for the second problem, it is feasible for the first problem. This observation leads to the estimate

$$\inf \mathcal{P}(\mathbb{G}^{\downarrow}_J,\mathbb{G}^{\uparrow}_J;\boldsymbol{\eta}^{\downarrow}_J,\boldsymbol{\xi}^{\uparrow}_J) - \inf \mathcal{P}(\mathbb{G}^{\downarrow}_J,\mathbb{G}^{\uparrow}_J;\boldsymbol{\eta},\boldsymbol{\xi}) \leq \ell_J(\boldsymbol{x}_J,\boldsymbol{y}_J) - \ell(\boldsymbol{x}_J,\boldsymbol{y}_J)\,.$$

The above reasoning implies that

$$\begin{aligned} &\limsup_{J\to\infty}\ \inf \mathcal{P}(\mathbb{G}^{\downarrow}_J,\mathbb{G}^{\uparrow}_J;\boldsymbol{\eta}^{\downarrow}_J,\boldsymbol{\xi}^{\uparrow}_J) - \inf \mathcal{P}(\mathbb{G}^{\downarrow}_J,\mathbb{G}^{\uparrow}_J;\boldsymbol{\eta},\boldsymbol{\xi}) \\ &\quad\leq \limsup_{J\to\infty}\ \mathrm{E}\int_0^T |c(\boldsymbol{x}_{J,t},\boldsymbol{\eta}^{\downarrow}_{J,t},t) - c(\boldsymbol{x}_{J,t},\boldsymbol{\eta}_t,t)|\,\lambda(\mathrm{d}t) \\ &\qquad + \limsup_{J\to\infty}\ \mathrm{E}\int_0^T |\boldsymbol{y}_{J,t}|\,|f_2(\boldsymbol{\xi}^{\uparrow}_{J,t},t) - f_2(\boldsymbol{\xi}_t,t)|\,\lambda(\mathrm{d}t) \\ &\quad\leq \mathrm{E}\int_0^T \limsup_{J\to\infty}\ |c(\boldsymbol{x}_{J,t},\boldsymbol{\eta}^{\downarrow}_{J,t},t) - c(\boldsymbol{x}_{J,t},\boldsymbol{\eta}_t,t)|\,\lambda(\mathrm{d}t) \\ &\qquad + \mathrm{E}\int_0^T \limsup_{J\to\infty}\ |\boldsymbol{y}_{J,t}|\,|f_2(\boldsymbol{\xi}^{\uparrow}_{J,t},t) - f_2(\boldsymbol{\xi}_t,t)|\,\lambda(\mathrm{d}t) \;=\; 0\,. \end{aligned}$$

The second inequality follows from Fatou's lemma, which applies due to (C2), (C3), and since $\boldsymbol{x}_J$ and $\boldsymbol{y}_J$ are essentially bounded. The equality in the last line

[8] The infimum of $\mathcal{P}(\mathbb{G}^{\downarrow}_J,\mathbb{G}^{\uparrow}_J;\boldsymbol{\eta}^{\downarrow}_J,\boldsymbol{\xi}^{\uparrow}_J)$ is finite and attained since the set $X_{\mathrm{c}}(\mathbb{G}^{\downarrow}_J)$ is weak* compact and the objective function $\boldsymbol{x} \mapsto \sup\{\ell_J(\boldsymbol{x},\boldsymbol{y}) | \boldsymbol{y} \in Y_{\mathrm{c}}(\mathbb{G}^{\uparrow}_J)\}$ is finite and weak* lsc.

follows from Lemma 7.5. By means of an analogous argument one further shows

$$\liminf_{J\to\infty}\ \inf \mathcal{P}(\mathbb{G}_J^{\downarrow}, \mathbb{G}_J^{\uparrow}; \boldsymbol{\eta}_J^{\downarrow}, \boldsymbol{\xi}_J^{\uparrow}) - \inf \mathcal{P}(\mathbb{G}_J^{\downarrow}, \mathbb{G}_J^{\uparrow}; \boldsymbol{\eta}, \boldsymbol{\xi}) \geq 0\,,$$

and thus assertion (i) is established. As the above reasoning remains valid if the upper and lower aggregation operators are interchanged, assertion (ii) follows. □

Having in mind the particular results of Propositions 7.3 and 7.6, our main convergence theorem now adopts the character of a simple corollary.

Theorem 7.7. *If J tends to infinity, then*

(i) $\inf \mathcal{P}(\mathbb{F}_J^{\downarrow}, \mathbb{F}_J^{\uparrow}; \boldsymbol{\eta}_J^{\downarrow}, \boldsymbol{\xi}_J^{\uparrow}) \to \inf \mathcal{P}$*;*

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

Proof. By using the relation (6.2a), we obtain

$$\begin{aligned}
\inf \mathcal{P}(\mathbb{F}_J^{\downarrow}, \mathbb{F}_J^{\uparrow}; \boldsymbol{\eta}_J^{\downarrow}, \boldsymbol{\xi}_J^{\uparrow}) - \inf \mathcal{P} &= \inf \mathcal{P}(\mathbb{G}_J^{\downarrow}, \mathbb{G}_J^{\uparrow}; \boldsymbol{\eta}_J^{\downarrow}, \boldsymbol{\xi}_J^{\uparrow}) - \inf \mathcal{P} \\
&= \inf \mathcal{P}(\mathbb{G}_J^{\downarrow}, \mathbb{G}_J^{\uparrow}; \boldsymbol{\eta}_J^{\downarrow}, \boldsymbol{\xi}_J^{\uparrow}) - \inf \mathcal{P}(\mathbb{G}_J^{\downarrow}, \mathbb{G}_J^{\uparrow}; \boldsymbol{\eta}, \boldsymbol{\xi}) \\
&\qquad + \inf \mathcal{P}(\mathbb{G}_J^{\downarrow}, \mathbb{G}_J^{\uparrow}; \boldsymbol{\eta}, \boldsymbol{\xi}) - \inf \mathcal{P}\,.
\end{aligned}$$

Propositions 7.3(i) and 7.6(i) then imply that the last expression converges to zero as the parameter J becomes large. Thus, assertion (i) is established. The proof of (ii) is analogous and will therefore be omitted. □

Example 7.8. *Consider again Example 3.5. Assume that a reservoir with capacity* $\overline{x}_{\text{store}} = 0.5$ *is initially empty and may be filled until time* $T = 1$*. The rate of energy purchases must not exceed* $\overline{x}_{\text{buy}} = 1$*, while natural reservoir inflows are absent, and energy sales are forbidden, that is,* $\overline{x}_{\text{infl}} = \overline{x}_{\text{sell}} = 0$*. Assume that the initial electricity price amounts to* $\boldsymbol{\eta}_0 = 1.1$ *and that* $\sigma_\eta = 0.25$*. We further set* $\epsilon_{\text{P}} = 1$ *and* $c_{\text{buy}} = -1$*. With these specifications, the hydropower plant reduces to a 'swing put option' [15] with strike price 1 and a budget of exercise times*

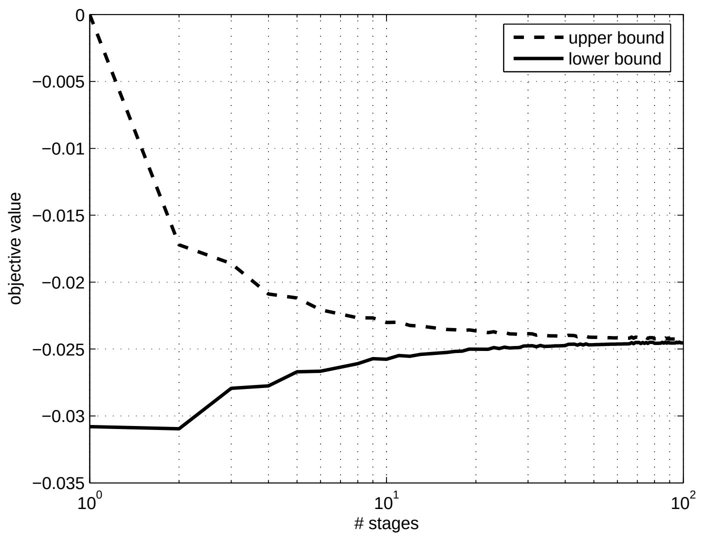

Figure 1: Convergence of bounds

that covers half of the contract period. Note that a swing put option constitutes a path-dependent American-style derivative whose valuation is nontrivial.

For each $J \in \mathbb{N}$ we introduce aggregation operators defined through $\uparrow_J(t) := \lceil t\,J \rceil / J$ and $\downarrow_J(t) := \lfloor t\,J / \rfloor / J$, $t \in \mathbb{T}$. The aggregated problems associated with $(\uparrow_J, \downarrow_J)$ are now particularly simple. By the martingale property of $\boldsymbol{\eta}$, the timing of energy purchases within the intervals $\mathbb{T}_j^J := [\frac{j-1}{J}, \frac{j}{J})$, $j = 1, \ldots, J$, is irrelevant. Without loss of generality, we may therefore append an extra constraint to the approximate problems, that is, we may require $\boldsymbol{x}_{\mathrm{buy},t}$ to be constant on the intervals $\{\mathbb{T}_j^J\}_{j=1}^J$. This has no effect on the optimal value of the aggregated problems; see Theorem B.1 in the appendix. Consequently, the aggregated problems can be solved accurately by standard dynamic programming even for large values of J, and no additional approximations based on scenario trees are necessary.[9] In agreement with the Theorems 6.3 and 7.7 we find that the optimal values of the upper/lower approximate problems converge from above/below to the optimal

[9]This follows from the fact that our simplified model has only two relevant state variables, namely the spot price of electricity and the storage level.

value of the original problem as J tends to infinity, see Figure 1.

Remark 7.9. *As in the discrete-time case [19, § 5.2], one can extend the scope of our approximation method to cases in which the data process follows a linear Markov process, that is, a Markov process $\boldsymbol{\zeta}$ which satisfies*

$$\left.\begin{aligned} \mathrm{E}(\boldsymbol{\eta}_t|\mathcal{F}_s) = H^{\mathrm{o}}_{t,s}(\boldsymbol{\eta}_s) \\ \mathrm{E}(\boldsymbol{\xi}_t|\mathcal{F}_s) = H^{\mathrm{c}}_{t,s}(\boldsymbol{\xi}_s) \end{aligned}\right\} \quad P\text{-a.s.} \quad \text{for } 0 \le s \le t \le T\,. \tag{7.1}$$

The continuous functions $(t,s,\eta) \mapsto H^{\mathrm{o}}_{t,s}(\eta)$ and $(t,s,\xi) \mapsto H^{\mathrm{c}}_{t,s}(\xi)$ are assumed to be linear affine and invertible in η and ξ, respectively. For $s > t$ it is convenient to define $H^{\mathrm{o}}_{t,s} := (H^{\mathrm{o}}_{s,t})^{-1}$ and $H^{\mathrm{c}}_{t,s} := (H^{\mathrm{c}}_{s,t})^{-1}$. Notice that the linear Markov processes cover the class of Markov martingales, for which $H^{\mathrm{o}}_{t,s}$ and $H^{\mathrm{c}}_{t,s}$ reduce to identity matrices, as well as a large class of mean-reverting processes. For the further argumentation, we update the regularity condition (C3) as follows:

(C3') $\boldsymbol{\zeta}$ is a linear Markov process with rcll sample paths, while the random variables $\sup_{t,s} \psi_\eta(|H^{\mathrm{o}}_{t,s}\boldsymbol{\eta}_s|)$ and $\sup_{t,s} \psi_\xi(|H^{\mathrm{c}}_{t,s}\boldsymbol{\xi}_s|)$ are P-integrable.

From now on, we assume that the updated condition (C3') replaces (C3). Next, we consider two sequences of aggregation operators as in Section 7, $\{\uparrow_J\}_{J\in\mathbb{N}}$ and $\{\downarrow_J\}_{J\in\mathbb{N}}$, which converge monotonically to the identity mapping on $\mathbb{T}$. For each $J \in \mathbb{N}$, we define two automorphisms $H^{\mathrm{o},\uparrow}_J$ and $H^{\mathrm{o},\downarrow}_J$ on $\mathcal{L}^1_k(\mathbb{T}\times\Omega, \mathcal{M}, \lambda\otimes P)$ via

$$H^{\mathrm{o},\uparrow}_J(\boldsymbol{\eta}) := \{H^{\mathrm{o}}_{t,\uparrow_J(t)}(\boldsymbol{\eta}_t)\}_{t\in\mathbb{T}} \quad \text{and} \quad H^{\mathrm{o},\downarrow}_J(\boldsymbol{\eta}) := \{H^{\mathrm{o}}_{t,\downarrow_J(t)}(\boldsymbol{\eta}_t)\}_{t\in\mathbb{T}}\,.$$

Two additional automorphisms $H^{\mathrm{c},\uparrow}_J$ and $H^{\mathrm{c},\downarrow}_J$ on $\mathcal{L}^1_l(\mathbb{T}\times\Omega, \mathcal{M}, \lambda\otimes P)$ are defined analogously. The following result generalizes the Theorems 6.3 and 7.7 to situations in which the data processes may fail to be martingales.

Theorem 7.10. *Let $\boldsymbol{\zeta}$ be a linear Markov process. If J tends to infinity, then*

$$\begin{aligned} &\inf \mathcal{P}(\mathbb{F}^{\downarrow}_J, \mathbb{F}^{\uparrow}_J; H^{\mathrm{o},\downarrow}_J(\boldsymbol{\eta}^{\downarrow}_J), H^{\mathrm{c},\uparrow}_J(\boldsymbol{\xi}^{\uparrow}_J)) \searrow \inf \mathcal{P}\,, \\ &\inf \mathcal{P}(\mathbb{F}^{\uparrow}_J, \mathbb{F}^{\downarrow}_J; H^{\mathrm{o},\uparrow}_J(\boldsymbol{\eta}^{\uparrow}_J), H^{\mathrm{c},\downarrow}_J(\boldsymbol{\xi}^{\downarrow}_J)) \nearrow \inf \mathcal{P}\,. \end{aligned}$$

This result can be proved along the lines of Theorems 6.3 and 7.7, but the relations (7.1) *are used in replacement of the martingale property. In particular, as the mappings* $H^{\mathrm{o}} : \mathbb{T}^2 \times \mathbb{R}^k \to \mathbb{R}^k$ *and* $H^{\mathrm{c}} : \mathbb{T}^2 \times \mathbb{R}^l \to \mathbb{R}^l$ *are continuous, the reasoning in Section 7 remains valid with obvious slight modifications.*

Example 7.11. *The martingale model of the data processes in Example 3.5 is unrealistic for long planning horizons since the electricity price and inflow rate are believed to exhibit mean-reversion. To overcome this shortcoming, we establish a new model in which* $\boldsymbol{\eta}$ *and* $\boldsymbol{\xi}$ *satisfy the stochastic differential equations*

$$\mathrm{d}\boldsymbol{\eta}_t = \kappa_\eta(\theta_\eta - \boldsymbol{\eta}_t)\,\mathrm{d}t + \boldsymbol{\eta}_t\,\sigma_\eta\,\mathrm{d}\boldsymbol{w}_{\eta,t} \tag{7.2a}$$

$$\mathrm{d}\boldsymbol{\xi}_t = \kappa_\xi(\theta_\xi - \boldsymbol{\xi}_t)\,\mathrm{d}t + \boldsymbol{\xi}_t\,\sigma_\xi\,\mathrm{d}\boldsymbol{w}_{\xi,t} \tag{7.2b}$$

with deterministic initial data $\boldsymbol{\eta}_0$ *and* $\boldsymbol{\xi}_0$*. As usual,* $\boldsymbol{w}_\eta$ *and* $\boldsymbol{w}_\xi$ *denote two independent standard Brownian motions on* $(\Omega, \mathcal{A}, P)$*. Equation* (7.2a) *has a unique strong solution, which is given by*

$$\boldsymbol{\eta}_t = e^{-\left(\kappa_\eta + \frac{\sigma_\eta^2}{2}\right)t + \sigma_\eta\,\boldsymbol{w}_{\eta,t}} \left(\boldsymbol{\eta}_0 + \kappa_\eta\,\theta_\eta \int_0^t e^{+\left(\kappa_\eta + \frac{\sigma_\eta^2}{2}\right)s - \sigma_\eta\,\boldsymbol{w}_{\eta,s}}\mathrm{d}s\right),\ t \in \mathbb{T}, \tag{7.3}$$

and the solution of (7.2b) *takes a similar form. It is easily verified that* $\boldsymbol{\zeta}$ *represents a linear Markov process with continuous sample paths. In fact, we have*

$$H^{\mathrm{o}}_{t,s}(\eta) = \eta\, e^{-\kappa_\eta(t-s)} + \theta_\eta\left(1 - e^{-\kappa_\eta(t-s)}\right), \quad (t,s,\eta) \in \mathbb{T}^2 \times \mathbb{R}\,, \tag{7.4}$$

while an equivalent formula holds for $H^{\mathrm{c}}_{t,s}(\xi)$*. Standard results on Brownian motion [22] can be used to show that the random variables* $\sup_{t,s}\psi_\eta(|H^{\mathrm{o}}_{t,s}\boldsymbol{\eta}_s|)$ *and* $\sup_{t,s}\psi_\xi(|H^{\mathrm{c}}_{t,s}\boldsymbol{\xi}_s|)$ *are integrable with respect to* P*. Thus, condition (C3') holds.*

Remark 7.12. *Our approximation scheme also extends to situations in which the relationship* (7.1) *is nonlinear: if the data process is given by a smooth transformation of a linear Markov process, one can use regularization techniques as in [18, § 5] to locally enforce the saddle property of the Lagrangian density.*

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

A Proof of Lemma 3.3

Let $\mathbb{G}^+ := \{\mathcal{G}_t^+\}_{t\in\mathbb{T}}$ be the right-continuous completion of $\mathbb{G}$ [7, § IV.48]. It is easy to see that $\boldsymbol{x} \sim \tilde{\boldsymbol{x}}$ for some process $\tilde{\boldsymbol{x}}$ which is $\mathcal{T}\otimes\mathcal{A}$-measurable and $\mathbb{G}^+$-adapted. By [6, Theorem 3.6] there exists a $\mathbb{G}^+$-predictable process $\hat{\boldsymbol{x}} \sim \tilde{\boldsymbol{x}}$, and by [8, Lemma 7, p. 413] there exists a $\mathbb{G}$-predictable (and, a fortiori, $\mathbb{G}$-progressive) process $\boldsymbol{x}' \sim \hat{\boldsymbol{x}}$. By the transitivity of equivalence relations, we have $\boldsymbol{x} \sim \boldsymbol{x}'$, that is, both processes represent the same element of X.

Next, define $\boldsymbol{z}'$ through $\boldsymbol{z}'_t = \int_0^t \boldsymbol{x}'_s\,\lambda(\mathrm{d}s)$. Since the mapping $(s,\omega)\mapsto \boldsymbol{x}'_s(\omega)$ is $\mathcal{B}([0,t])\otimes\mathcal{G}_t$-measurable on $[0,t]\times\Omega$, we can use Fubini's theorem to verify that $\boldsymbol{z}'_t$ is $\mathcal{G}_t$-measurable. Moreover, since the mapping $(s,t,\omega)\mapsto 1_{(s\le t)}\,\boldsymbol{x}'_s(\omega)$ is $\mathcal{T}\otimes\mathcal{T}\otimes\mathcal{A}$-measurable on $\mathbb{T}\times\mathbb{T}\times\Omega$, Fubini's theorem implies that $\boldsymbol{z}'$ is $\mathcal{T}\otimes\mathcal{A}$-measurable. It can easily be verified that $\boldsymbol{z}'$ is essentially bounded with respect to $\lambda\otimes P$, and thus we find that $\boldsymbol{z}'$ is $\mathcal{M}(\mathbb{G})$-measurable and contained in X.

The mapping $(s,t,\omega)\mapsto 1_{(s\le t)}\,|\boldsymbol{x}_s(\omega)-\boldsymbol{x}'_s(\omega)|$ is $\mathcal{T}\otimes\mathcal{T}\otimes\mathcal{A}$-measurable and vanishes almost everywhere with respect to $\lambda\otimes\lambda\otimes P$. Another application of Fubini's theorem then implies that $|\boldsymbol{z}_t(\omega)-\boldsymbol{z}'_t(\omega)|$ vanishes almost everywhere with respect to $\lambda\otimes P$. Thus, we have $\boldsymbol{z}\sim\boldsymbol{z}'$, which completes the proof. □

B Problems with simple time dependence

Here we discuss a new regularity condition which ensures that the aggregated problems of Theorem 6.3 are equivalent to standard stochastic programs [4] with finitely many stages and finitely many decision variables per stage and scenario.

Assume that a pair of aggregation operators $\uparrow$ and $\downarrow$ with a finite number of fixed points has been selected, and denote by $\mathcal{T}^{\uparrow\downarrow}$ the σ-field on $\mathbb{T}$ generated by $\uparrow$ and $\downarrow$. Without loss of generality, we may assume that each primal decision process $\boldsymbol{x}\in X$ can be decomposed into subprocesses $\boldsymbol{u}$ and $\boldsymbol{v}$ valued in $\mathbb{R}^{n_u}$ and $\mathbb{R}^{n_v}$, respectively, where $n_u+n_v=n$. We refer to $\boldsymbol{u}$ as the *control process*, while

$\boldsymbol{v}$ is termed *state process*. With obvious definitions of the involved matrix- and vector-valued functions, the dynamic constraint in (3.1) is then representable as

$$W_1^u(t)\boldsymbol{u}_t + W_1^v(t)\boldsymbol{v}_t + \int_0^t W_2^u(s)\boldsymbol{u}_s + W_2^v(s)\boldsymbol{v}_s \, \lambda(\mathrm{d}s) = h(t) \quad \lambda \otimes P\text{-a.e.}$$

Next, we impose an additional regularity condition:

(C6) the mappings $W_1^u(t)$, $W_1^v(t)$, $W_2^u(t)$, and $h(t)$ as well as the multifunction $\Gamma(t)$ are $\mathcal{T}^{\uparrow\downarrow}$-measurable. Moreover, $W_2^v(t)$ vanishes identically, and $W_1^v(t)$ has full rank $\forall t \in \mathbb{T}$. The cost rate function $c(u, v, \eta, t)$ and the constraint function $f(u, v, \xi, t)$ are affine in u, constant in v, and $\mathcal{T}^{\uparrow\downarrow}$-measurable in t.

Assumption (C6) implies that the state process $\boldsymbol{v}$ is uniquely determined by the control process $\boldsymbol{u}$. Moreover, the evaluation of $\boldsymbol{v}$ for a given $\boldsymbol{u}$ is straightforward since $\boldsymbol{v}$ does not influence the integral in the dynamic constraint.

Let $\mathcal{M}^{\uparrow\downarrow}$ be the completion of the σ-algebra $\mathcal{T}^{\uparrow\downarrow} \otimes \mathcal{A}$ on $\mathbb{T} \times \Omega$ with respect to $\lambda \otimes P$. An $\mathcal{M}^{\uparrow\downarrow}$-measurable stochastic process has sample paths that are (essentially) constant between the fixed points of the aggregation operators. For instance, $\boldsymbol{\zeta}^{\uparrow}$ and $\boldsymbol{\zeta}^{\downarrow}$ are $\mathcal{M}^{\uparrow\downarrow}$-measurable by construction. For any filtration $\mathbb{G}$ on the sample space, we denote by $X_{\mathrm{c}}^{\uparrow\downarrow}(\mathbb{G})$ the set of all primal policies in $X_{\mathrm{c}}(\mathbb{G})$ whose control processes are $\mathcal{M}^{\uparrow\downarrow}$-measurable. Notice that by assumption (C6) the state processes corresponding to such policies have (essentially) affine sample paths between the fixed points of the aggregation operators. Next, for any filtrations $\mathbb{G}$ and $\mathbb{H}$ and for any data processes $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ satisfying (C3) we define $\mathcal{P}^{\uparrow\downarrow}(\mathbb{G}, \mathbb{H}; \boldsymbol{\eta}, \boldsymbol{\xi})$ as the stochastic optimization problem $\mathcal{P}(\mathbb{G}, \mathbb{H}; \boldsymbol{\eta}, \boldsymbol{\xi})$ with the extra constraint $\boldsymbol{x} \in X_{\mathrm{c}}^{\uparrow\downarrow}(\mathbb{G})$. By using (C6) one can now show that the approximate problems of Theorem 6.3 are equivalent to standard stochastic programs.

Theorem B.1. *If the conditions (C1)–(C6) hold, then*

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

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

The proof of Theorem B.1 is lengthy but uses only techniques familiar from Section 6. For brevity, it will thus be omitted. The theorem states that the optimal values of the aggregated problems are not affected by restricting the control variables to be piecewise constant and the state variables to be piecewise linear between the fixed points of the aggregation operators. Therefore, all function-valued decisions can be parametrized by finitely many (scalar) variables, while all continuous constraints can be replaced by finitely many (scalar) constraints in each decision stage and scenario. In this manner, the stage-aggregated approximate problems of Theorem 6.3 reduce to standard multistage stochastic programs for which many well-established solution algorithms are available, see e.g. [4].

References

[1] Attouch, H., and Wets, R.-B. A convergence theory for saddle functions. *Trans. Am. Math. Soc. 280*, 1 (1983), 1–41.

[2] Ben-Tal, A., Golany, B., Nemirovski, A., and Vial, J. Supplier-retailer flexible commitments contracts: A robust optimization approach. *Manufacturing Service Oper. Managment 7*, 3 (2005), 248–273.

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

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

[5] Casey, M., and Sen, S. The scenario generation algorithm for multistage stochastic linear programming. *Math. Oper. Res. 30*, 3 (2005), 615–631.

[6] Chung, K., and Williams, R. *Introduction to Stochastic Integration.* Birkhäuser, Boston, 1983.

[7] Dellacherie, C., and Meyer, P. *Probabilités et potentiel. Chapitres I IV.* Hermann, Paris, 1975.

[8] Dellacherie, C., and Meyer, P. *Probabilités et potentiel. Chapitres V VIII.* Hermann, Paris, 1975.

[9] DeMiguel, V., and Mishra, N. What multistage stochastic programming can do for network revenue management. London Business School working paper, 2006.

[10] Doob, J. *Measure Theory.* Springer, New York, 1994.

[11] Dunford, N., and Schwartz, J. T. *Linear Operators: Part I.* John Wiley and Sons, New York, Chichester, Brisbane, Toronto, Singapore, 1988.

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

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

[14] Frauendorfer, K., and Schürle, M. Management of non-maturing deposits by multistage stochastic programming. *Eur. J. Oper. Res. 151* (2003), 602–616.

[15] Haarbrücker, G., and Kuhn, D. Valuation of electricity swing options by multistage stochastic programming. To appear in Automatica, 2008.

[16] Karmarkar, U. S. The multiperiod multilocation inventory problem. *Oper. Res. 29*, 2 (1981), 215–228.

[17] 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.

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

[19] Kuhn, D. Aggregation and discretization in multistage stochastic programming. *Math. Program. A 113*, 1 (2008), 61–94.

[20] Kuhn, D. Convergent bounds for stochastic programs with expected value constraints. To appear in *J. Optim. Theory Appl.*, 2008.

[21] Pereira, M., and Pinto, L. Multistage stochastic optimization applied to energy planning. *Math. Program. 52* (1991), 359–375.

[22] Revuz, D., and Yor, M. *Continuous Martingales and Brownian Motion*. Springer, New York, 1991.

[23] Rockafellar, R. Integrals which are convex functionals. *Pacific J. Math. 24*, 3 (1968), 525–539.

[24] Rockafellar, R. Integrals which are convex functionals, II. *Pacific J. Math. 39*, 2 (1971), 439–469.

[25] Rockafellar, R. *Conjugate Duality and Optimization*. SIAM, Philadelphia, 1974.

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

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