From infinite to finite programs

Explicit error bounds with applications to approximate dynamic programming

Mohajerin Esfahani, Peyman; Sutter, Tobias; Kuhn, Daniel; Lygeros, John

DOI
10.1137/17M1133087

Publication date
2018

Document Version
Final published version

Published in
SIAM Journal on Optimization

SIAM J. OPTIM.
Vol. 28, No. 3, pp. 1968–1998

FROM INFINITE TO FINITE PROGRAMS: EXPLICIT ERROR BOUNDS WITH APPLICATIONS TO APPROXIMATE DYNAMIC PROGRAMMING*

PEYMAN MOHAJERIN ESFAHANI†, TOBIAS SUTTER‡, DANIEL KUHN§, AND JOHN LYGEROS‡

Abstract. We consider linear programming (LP) problems in infinite dimensional spaces that are in general computationally intractable. Under suitable assumptions, we develop an approximation bridge from the infinite dimensional LP to tractable finite convex programs in which the performance of the approximation is quantified explicitly. To this end, we adopt the recent developments in two areas of randomized optimization and first-order methods, leading to a priori as well as a posteriori performance guarantees. We illustrate the generality and implications of our theoretical results in the special case of the long-run average cost and discounted cost optimal control problems in the context of Markov decision processes on Borel spaces. The applicability of the theoretical results is demonstrated through a fisheries management problem.

Key words. infinite dimensional linear programming, Markov decision processes, approximate dynamic programming, randomized and convex optimization

AMS subject classifications. 90C05, 90C39, 90C34, 93E20, 90C40, 68W20

DOI. 10.1137/17M1133087

1. Introduction.

Linear programming (LP) problems in infinite dimensional spaces appear in, among other areas, engineering, economics, operations research, and probability theory [1]. Infinite LPs offer remarkable modeling power, subsuming general finite dimensional optimization problems and the generalized moment problem as special cases. They are, however, often computationally formidable, motivating the study of approximations schemes.

A particularly rich class of problems that can be modeled as infinite LPs involves Markov decision processes (MDP) and optimal control problems defined in this context. The history beyond this link dates back to the seventies, when the connection between multistage stochastic programs and infinite LPs was discovered [20, 39, 40]. More often than not, it is impossible to obtain explicit solutions to MDP problems, making it necessary to resort to approximation techniques. Such approximations are the core of a methodology known as *approximate dynamic programming* [6, 8]. Interestingly, a wide range of optimal control problems involving MDP can be equivalently expressed as *static* optimization problems over a closed convex set of measures, more specifically, as infinite LPs [25, 27]. This LP reformulation is particularly appealing for dealing with unconventional settings involving additional constraints [3], secondary costs [18], information-theoretic considerations [44], and reachability problems [32]. In addition, the infinite LP reformulation allows one to leverage the developments in the optimization literature, in particular convex approximation techniques, to develop

*Received by the editors June 5, 2017; accepted for publication (in revised form) April 26, 2018; published electronically July 3, 2018.
http://www.siam.org/journals/siopt/28-3/M113308.html
†Delft Center for Systems & Control, TU Delft, Delft, Netherlands (P.MohajerinEsfahani@tudelft.nl).
‡Automatic Control Lab, ETH Zurich, Zurich, Switzerland (sutter@control.ee.ethz.ch, lygeros@control.ee.ethz.ch).
§Risk Analytics and Optimization Chair, EPFL, Lausanne, Switzerland (daniel.kuhn@epfl.ch).

approximation schemes for MDP problems. This will also be the perspective adopted in the present article.

Approximation schemes to tackle infinite LPs have historically been developed for special classes of problems, e.g., the general capacity problem [30], or the generalized moment problem [31]. The literature on MDPs with infinite state or action spaces mostly concentrates on approximation schemes with asymptotic performance guarantees [26, 27]; see also the comprehensive book [29] for controlled stochastic differential equations and [33] for reachability problems in the similar setting. From a practical viewpoint, a challenge using these schemes is that the convergence analysis is not constructive and does not lead to explicit error bounds. A wealth of approximation schemes have been proposed in the literature under the names of approximate dynamic programming [5], neuro-dynamic programming [8], reinforcement learning [28, 47], and value and/or policy iteration [6, 42]. Most, however, deal with discrete (finite or at most countable) state and action spaces, while approximation over uncountable spaces remains largely unexplored.

The MDP literature on explicit approximation errors in uncountable settings can, roughly speaking, be divided into two groups in terms of the performance criteria considered: discounted cost and average cost (AC). Of the two, the discounted cost setting has received more attention as the corresponding dynamic programming operator is a contraction, a useful property to obtain a convergence rate for the approximation error. Examples include the LP approach [13, 14] and also a recent series of works [11, 17, 18] on approximating a probability measure that underlies the random transitions of the dynamics of the system using different discretization procedures. Long-run AC problems introduce new challenges due to losing the contraction property. The authors in [19] develop approximation schemes leading to finite but nonconvex optimization problems, while [43] investigates the convergence rate of the finite-state approximation to the original (uncountable) MDP problem.

The approach presented in this article tackles a class of general infinite LPs that, as a special case, cover both long-run discounted and AC performance criteria. The resulting approximation is based on finite convex programs that are different from the existing schemes. Closest in spirit to our proposed approximation is the LP approach based on constraint sampling in [13, 14, 46]. Unlike these works, however, we introduce an additional norm constraint that effectively acts as a *regularizer*. We study in detail the conditions under which this regularizer can be exploited to bound the optimizers of the primal and dual programs and hence provide an explicit approximation error for the proposed solution.

The proposed approximation scheme involves a restriction of the decision variables from an infinite dimensional space to a finite dimensional subspace, followed by the approximation of the infinite number of constraints by a finite subset; we develop two complementary methods for performing the latter step. The structure of the article is illustrated in Figure 1, where the contributions are summarized as follows:

- We introduce a subclass of infinite LPs whose *regularized* semi-infinite restriction enjoys analytical bounds for both primal and dual optimizers (Proposition 3.2). The implications for MDP with AC (Lemma 3.7) and with discounted cost (Lemma A.2) are also investigated.
- We derive an explicit error bound between the original infinite LP and the regularized semi-infinite counterpart, providing insights on the impact of the underlying norm structure as well as on how the choice of basis functions contributes to the approximation error (Theorem 3.3, Corollary 3.5). In the MDP setting, we recover an existing result as a special case (Corollary 3.9).

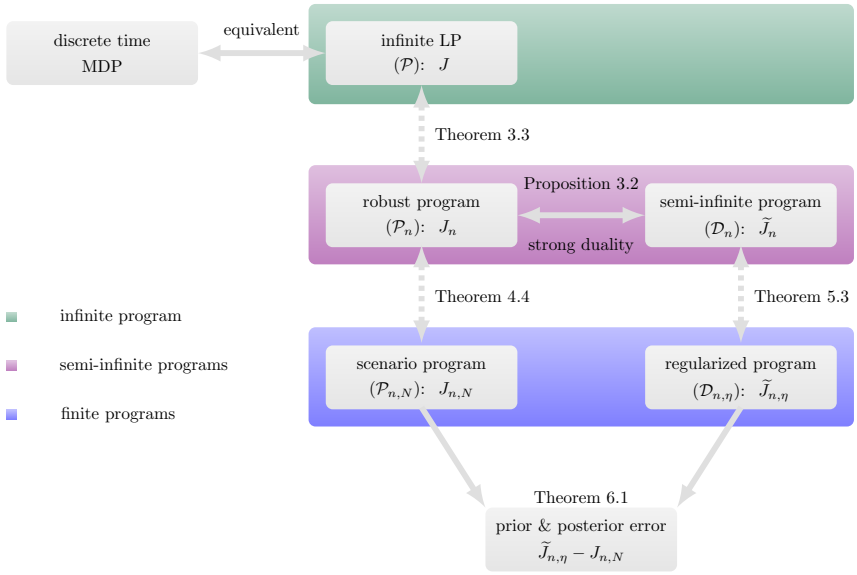


FIG. 1. *Graphical representation of the article structure and its contributions.*

- We adopt the recent developments from the randomized optimization literature to propose a finite convex program whose solution enjoys a priori probabilistic performance bounds (Theorem 4.4). We extend the existing results to also offer an a posteriori bound under a generic underlying norm structure. The required conditions and theoretical assertions are validated in the MDP setting (Corollary 4.12).
- In parallel to the randomized approach, we also utilize the recent developments in the structural convex optimization literature to propose an iterative algorithm for approximating the semi-infinite program. For this purpose, we extend the setting to incorporate unbounded prox-terms with a certain growth rate (Theorem 5.3). We illustrate how this extension allows us to deploy the entropy prox-term in the MDP setting (Lemma 5.10, Corollary 5.8).

Section 2 introduces the main motivation for the work, namely, the control of discrete-time MDP and their LP characterization. Using standard results in the literature we embed these MDP in the more general framework of infinite LPs. Section 3 studies the link from infinite LPs to semi-infinite programs. Section 4 presents the approximation of semi-infinite programs based on randomization, while section 5 approaches the same objective using first-order convex optimization methods. Section 6 summarizes the results in the preceding sections, establishing the approximation error from the original infinite LP to the finite convex counterparts. Section 7 illustrates the theoretical results through a truncated linear quadratic Gaussian (LQG) example and a fisheries management problem. The proof of a few technical lemmas and an additional numerical simulation are given in an extended online version [34].

Notation. The set $\mathbb{R}_+$ denotes the set of nonnegative reals and $\|\cdot\|_{\ell_p}$ for $p \in [1,\infty]$ the standard p-norm in $\mathbb{R}^n$. Given a function $u : S \to \mathbb{R}$, we denote the infinity norm of the function by $\|u\|_\infty := \sup_{s\in S} |u(s)|$ and the Lipschitz norm by $\|u\|_{\mathrm{L}} := \sup_{s,s'\in S} \left\{|u(s)|, \frac{|u(s)-u(s')|}{\|s-s'\|_{\ell_\infty}}\right\}$. The space of Lipschitz functions on a set S is denoted by $\mathscr{L}(S)$; define the function $\mathbb{1}(s) \equiv 1 \ \forall s \in S$. We denote the Borel

σ-algebra on the (topological) space S by $\mathfrak{B}(S)$. Measurability is always understood in the sense of Borel. Products of topological spaces are assumed to be endowed with the product topology and the corresponding product σ-algebra. The space of finite signed measures (resp., probability measures) on S is denoted by $\mathcal{M}(S)$ (resp., $\mathcal{P}(S)$). The Wasserstein norm on the space of signed measures $\mathcal{M}(S)$ is defined by $\|\mu\|_{\mathrm{W}} := \sup_{\|u\|_{\mathrm{L}}\leq 1} \int_S u(s)\mu(\mathrm{d}s)$ and can be shown to be the dual of the Lipschitz norm. The set of extreme points of a set A is denoted by $\mathcal{E}\{A\}$. Given a bilinear form $\langle\cdot,\cdot\rangle$, the support function of A is defined by $\sigma_A(y) = \sup_{x\in A}\langle y, x\rangle$. The standard bilinear form in $\mathbb{R}^n$ (i.e., the inner product) is denoted by $y\cdot x$.

2. Motivation: Control of MDP and LP characterization.

2.1. MDP setting. We briefly recall some standard definitions and refer interested readers to [2, 24, 25] for further details. Consider a *Markov control model* $\big(S, A, \{A(s) : s \in S\}, Q, \psi\big)$, where S (resp., A) is a metric space called the *state space* (resp., *action space*) and for each $s \in S$ the measurable set $A(s) \subseteq A$ denotes the set of *feasible actions* when the system is in state $s \in S$. The *transition law* is a stochastic kernel Q on S given the feasible state-action pairs in $K := \{(s, a) : s \in S, a \in A(s)\}$. A stochastic kernel acts on real valued measurable functions u from the left as $Qu(s, a) := \int_S u(s')Q(\mathrm{d}s'|s, a)\ \forall (s, a) \in K$ and on probability measures μ on K from the right as $\mu Q(B) := \int_K Q(B|s, a)\mu\big(\mathrm{d}(s, a)\big)\ \forall B \in \mathfrak{B}(S)$. Finally $\psi : K \to \mathbb{R}_+$ denotes a measurable function called the *one-stage cost function*. The *admissible history spaces* are defined recursively as $H_0 := S$ and $H_t := H_{t-1} \times K$ for $t \in \mathbb{N}$ and the canonical sample space is defined as $\Omega := (S \times A)^\infty$. All random variables will be defined on the measurable space $(\Omega, \mathcal{G})$, where $\mathcal{G}$ denotes the corresponding product σ-algebra. A generic element $\omega \in \Omega$ is of the form $\omega = (s_0, a_0, s_1, a_1, \ldots)$, where $s_i \in S$ are the states and $a_i \in A$ the action variables. An *admissible policy* is a sequence $\pi = (\pi_t)_{t\in\mathbb{N}_0}$ of stochastic kernels π_t on A given $h_t \in H_t$, satisfying the constraints $\pi_t(A(s_t)|h_t) = 1$. The set of admissible policies will be denoted by Π. Given a probability measure $\nu \in \mathcal{P}(S)$ and policy $\pi \in \Pi$, by the Ionescu Tulcea theorem [7, pp. 140–141] there exists a unique probability measure $\mathbb{P}^\pi_\nu$ on $(\Omega, \mathcal{G})$ such that for all measurable sets $B \subset S$, $C \subset A$, $h_t \in H_t$, and $t \in \mathbb{N}_0$

$$\mathbb{P}^\pi_\nu\big(s_0 \in B\big) = \nu(B),\ \mathbb{P}^\pi_\nu\big(a_t \in C|h_t\big) = \pi_t(C|h_t),\ \mathbb{P}^\pi_\nu\big(s_{t+1} \in B|h_t, a_t\big) = Q(B|s_t, a_t).$$

The expectation operator with respect to $\mathbb{P}^\pi_\nu$ is denoted by $\mathbb{E}^\pi_\nu$. The stochastic process $\big(\Omega, \mathcal{G}, \mathbb{P}^\pi_\nu, (s_t)_{t\in\mathbb{N}_0}\big)$ is called a *discrete-time MDP*. For most of the article we consider optimal control problems where the aim is to minimize a long-term AC over the set of admissible policies and initial state measures. We definite the optimal value of the optimal control problem by

$$J^{\mathrm{AC}} := \inf_{(\pi,\nu)\in\Pi\times\mathcal{P}(S)} \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}^\pi_\nu\left[\sum_{t=0}^{T-1} \psi(s_t, a_t)\right]. \tag{2.1}$$

We emphasize, however, that the results also apply to other performance objectives, including the long-run *discounted cost* problem as shown in Appendix A.

2.2. Infinite LP characterization. The problem in (2.1) admits an alternative LP characterization under some mild assumptions.

Assumption 2.1 (control model). We stipulate that

(i) the set of feasible state-action pairs is the unit hypercube $K = [0, 1]^{\dim(S\times A)}$;
(ii) the transition law Q is Lipschitz continuous, i.e., there exists $L_Q > 0$ such that $|Qu(k) - Qu(k')| \leq L_Q\|u\|_\infty\|k - k'\|_{\ell_\infty}\ \forall k,\ k' \in K$;

(iii) the cost function ψ is nonnegative and Lipschitz continuous on K with respect to the ℓ_∞-norm.

Assumption 2.1(i) may seem restrictive; however, essentially it simply requires that the state-action set K is compact. We refer the reader to Example 7.2, where a nonrectangular K is transferred to a hypercube, and to [27, Chapter 12.3] for further information about the LP characterization in more general settings.

THEOREM 2.2 (LP characterization [19, Proposition 2.4]). *Under Assumption* 2.1*,*

$$-J^{\mathrm{AC}} = \left\{ \begin{array}{ll} \inf\limits_{\rho,u} & -\rho \\ \text{s.t.} & \rho + u(s) - Qu(s,a) \le \psi(s,a) \quad \forall (s,a) \in K, \\ & \rho \in \mathbb{R}, \quad u \in \mathscr{L}(S). \end{array} \right. \tag{2.2}$$

The LP (2.2) can be expressed in the standard conic form $\inf_{x\in\mathbb{X}} \left\{ \langle x, c\rangle : \mathcal{A}x - b \in \mathbb{K} \right\}$ by introducing

$$\left\{ \begin{array}{ll} \mathbb{X} = \mathbb{R} \times \mathscr{L}(S), & b(s,a) = -\psi(s,a), \\ x = (\rho, u) \in \mathbb{X}, & c = (c_1, c_2) = (-1, 0), \\ \mathbb{C} = \mathbb{R} \times \mathcal{M}(S), & \langle x, c\rangle = c_1\rho + \int_S u(s) c_2(\mathrm{d}s), \\ \mathbb{K} = \mathscr{L}_+(K), & \mathcal{A}x(s,a) = -\rho - u(s) + Qu(s,a), \end{array} \right. \tag{2.3}$$

where $\mathcal{M}(S)$ is the set of finite signed measures supported on S, and $\mathscr{L}_+(K)$ is the cone of Lipschitz functions taking nonnegative values. It should be noted that the choice of the positive cone $\mathbb{K} = \mathscr{L}_+(K)$ is justified since, thanks to Assumption 2.1(ii), the linear operator $\mathcal{A}$ maps the elements of $\mathbb{X}$ into $\mathscr{L}(K)$.

Our aim is to derive an approximation scheme for a class of such infinite dimensional LPs, including problems of the form (2.2), that comes with an explicit bound on the approximation error.

3. Infinite to semi-infinite programs.

3.1. Dual pairs of normed vector spaces. The triple $\big(\mathbb{X}, \mathbb{C}, \|\cdot\|\big)$ is called a *dual pair* of normed vector spaces if

- $\mathbb{X}$ and $\mathbb{C}$ are vector spaces;
- $\langle \cdot, \cdot \rangle$ is a bilinear form on $\mathbb{X} \times \mathbb{C}$ that "separates points," i.e.,
  - for each nonzero $x \in \mathbb{X}$ there is some $c \in \mathbb{C}$ such that $\langle x, c\rangle \neq 0$,
  - for each nonzero $c \in \mathbb{C}$ there is some $x \in \mathbb{X}$ such that $\langle x, c\rangle \neq 0$;
- $\mathbb{X}$ is equipped with the norm $\|\cdot\|$, which together with the bilinear form induces a *dual* norm in $\mathbb{C}$ defined through $\|c\|_* := \sup_{\|x\|\le 1} \langle x, c\rangle$.

The norm in the vector spaces is used as a means to quantify the performance of the approximation schemes. In particular, we emphasize that the vector spaces are not necessarily complete with respect to these norms.

Let $\big(\mathbb{B}, \mathbb{Y}, \|\cdot\|\big)$ be another dual pair of normed vector spaces. As there is no danger of confusion, we use the same notation for the potentially different norm and bilinear form for each pair. Let $\mathcal{A} : \mathbb{X} \to \mathbb{B}$ be a linear operator and $\mathbb{K}$ be a convex cone in $\mathbb{B}$. Given the fixed elements $c \in \mathbb{C}$ and $b \in \mathbb{B}$, we define a linear program, hereafter called the *primal* program $\mathcal{P}$, as

$$(\mathcal{P}) \qquad J := \left\{ \begin{array}{ll} \inf\limits_{x\in\mathbb{X}} & \langle x, c\rangle \\ \text{s.t.} & \mathcal{A}x \succeq_{\mathbb{K}} b, \end{array} \right.$$

where the conic inequality $\mathcal{A}x \succeq_{\mathbb{K}} b$ is understood in the sense of $\mathcal{A}x - b \in \mathbb{K}$. Throughout this study we assume that the program $\mathcal{P}$ has an optimizer (i.e., the infimum is indeed a minimum), the cone $\mathbb{K}$ is closed, and the operator $\mathcal{A}$ is continuous where the corresponding topology is the weakest in which the topological duals of $\mathbb{X}$ and $\mathbb{B}$ are $\mathbb{C}$ and $\mathbb{Y}$, respectively. Let $\mathcal{A}^* : \mathbb{Y} \to \mathbb{C}$ be the adjoint operator of $\mathcal{A}$ defined by $\langle \mathcal{A}x, y\rangle = \langle x, \mathcal{A}^* y\rangle \ \forall x \in \mathbb{X},\ \forall y \in \mathbb{Y}$. Recall that if $\mathcal{A}$ is weakly continuous, then the adjoint operator $\mathcal{A}^*$ is well defined as its image is a subset of $\mathbb{C}$ [27, Proposition 12.2.5]. The *dual* program of $\mathcal{P}$ is denoted by $\mathcal{D}$ and is given by

$$(\mathcal{D}) \qquad \widetilde{J} := \left\{ \begin{array}{ll} \sup\limits_{y \in \mathbb{Y}} & \langle b, y\rangle \\ \text{s.t.} & \mathcal{A}^* y = c, \\ & y \in \mathbb{K}^*, \end{array} \right.$$

where $\mathbb{K}^*$ is the dual cone of $\mathbb{K}$ defined as $\mathbb{K}^* := \{y \in \mathbb{Y} : \langle b, y\rangle \geq 0 \ \forall b \in \mathbb{K}\}$. It is not hard to see that *weak duality* holds, as

$$J = \inf_{x \in \mathbb{X}} \sup_{y \in \mathbb{K}^*} \langle x, c\rangle - \langle \mathcal{A}x - b, y\rangle \geq \sup_{y \in \mathbb{K}^*} \inf_{x \in \mathbb{X}} \langle x, c\rangle - \langle \mathcal{A}x - b, y\rangle = \widetilde{J}.$$

An interesting question is when the above assertion holds as an equality. This is known as *zero duality gap*, also referred to as *strong duality* particularly when both $\mathcal{P}$ and $\mathcal{D}$ admit an optimizer [1, p. 52]. Our study is not directly concerned with conditions under which strong duality between $\mathcal{P}$ and $\mathcal{D}$ holds; see [1, section 3.6] for a comprehensive discussion of such conditions. The programs $\mathcal{P}$ and $\mathcal{D}$ are assumed to be *infinite*, in the sense that the dimensions of the decision spaces ($\mathbb{X}$ in $\mathcal{P}$, and $\mathbb{Y}$ in $\mathcal{D}$) as well as the number of constraints are both infinite.

3.2. Semi-infinite approximation. Consider a family of linearly independent elements $\{x_n\}_{n \in \mathbb{N}} \subset \mathbb{X}$, and let $\mathbb{X}_n$ be the finite dimensional subspace generated by the first n elements $\{x_i\}_{i \leq n}$. Without loss of generality, we assume that x_i are normalized, i.e., $\|x_i\| = 1$. Restricting the decision space $\mathbb{X}$ of $\mathcal{P}$ to $\mathbb{X}_n$, along with an additional norm constraint, yields the program

$$(3.1) \qquad J_n := \left\{ \begin{array}{ll} \inf\limits_{\alpha \in \mathbb{R}^n} & \sum_{i=1}^n \alpha_i \langle x_i, c\rangle \\ \text{s.t.} & \sum_{i=1}^n \alpha_i \mathcal{A}x_i \succeq_{\mathbb{K}} b, \\ & \|\alpha\|_{\mathfrak{R}} \leq \theta_{\mathcal{P}}, \end{array} \right.$$

where $\|\cdot\|_{\mathfrak{R}}$ is a given norm on $\mathbb{R}^n$ and $\theta_{\mathcal{P}}$ determines the size of the feasible set. In the spirit of dual-paired normed vector spaces, one can approximate $(\mathbb{X}, \mathbb{C}, \|\cdot\|)$ by the finite dimensional counterpart $(\mathbb{R}^n, \mathbb{R}^n, \|\cdot\|_{\mathfrak{R}})$ where the bilinear form is the standard inner product. In this view, the linear operator $\mathcal{A} : \mathbb{X} \to \mathbb{B}$ may also be approximated by the linear operator $\mathcal{A}_n : \mathbb{R}^n \to \mathbb{B}$ with the respective adjoint $\mathcal{A}_n^* : \mathbb{Y} \to \mathbb{R}^n$ defined as

$$(3.2) \qquad \mathcal{A}_n \alpha := \sum_{i=1}^n \alpha_i \mathcal{A}x_i, \qquad \mathcal{A}_n^* y := \big[\langle \mathcal{A}x_1, y\rangle, \ldots, \langle \mathcal{A}x_n, y\rangle\big].$$

It is straightforward to verify the definitions (3.2) by noting that $\langle \mathcal{A}_n \alpha, y\rangle = \alpha \bullet \mathcal{A}_n^* y$ $\forall\, \alpha \in \mathbb{R}^n$ and $y \in \mathbb{Y}$. Defining the vector $\mathbf{c} := [\langle x_1, c\rangle, \ldots, \langle x_n, c\rangle]$, we can rewrite the program (3.1) as

$$(\mathcal{P}_n) \qquad J_n := \left\{ \begin{array}{ll} \inf\limits_{\alpha \in \mathbb{R}^n} & \alpha \bullet \mathbf{c} \\ \text{s.t.} & \mathcal{A}_n \alpha \succeq_{\mathbb{K}} b, \\ & \|\alpha\|_{\mathfrak{R}} \leq \theta_{\mathcal{P}}. \end{array} \right.$$

We call $\mathcal{P}_n$ a *semi-infinite* program, as the decision variable is a finite dimensional vector $\alpha \in \mathbb{R}^n$, but the number of constraints is still in general infinite due to the conic inequality. The additional constraint on the norm of α in $\mathcal{P}_n$ acts as a *regularizer* and is a key difference between the proposed approximation schemes and existing schemes in the literature. Methods for choosing the parameter $\theta_{\mathcal{P}}$ will be discussed later.

Dualizing the conic inequality constraint in $\mathcal{P}_n$ and using the dual norm definition leads to a dual counterpart

$$(\mathcal{D}_n) \qquad \widetilde{J}_n := \left\{ \begin{array}{ll} \sup\limits_{y \in \mathbb{Y}} & \big\langle b, y \big\rangle - \theta_{\mathcal{P}} \|\mathcal{A}_n^* y - \mathbf{c}\|_{\mathfrak{R}^*} \\ \text{s.t.} & y \in \mathbb{K}^*, \end{array} \right.$$

where $\|\cdot\|_{\mathfrak{R}^*}$ denotes the dual norm of $\|\cdot\|_{\mathfrak{R}}$. Note that setting $\theta_{\mathcal{P}} = \infty$ effectively implies that the second term of the objective in $\mathcal{D}_n$ introduces n hard constraints $\mathcal{A}_n^* y = \mathbf{c}$ (cf. (3.2)). We study further the connection between $\mathcal{P}_n$ and $\mathcal{D}_n$ under the following regularity assumption.

Assumption 3.1 (semi-infinite regularity). We stipulate that

(i) the program $\mathcal{P}_n$ is feasible;

(ii) there exists a positive constant γ such that $\|\mathcal{A}_n^* y\|_{\mathfrak{R}^*} \geq \gamma \|y\|_*$ for every $y \in \mathbb{K}^*$, and $\theta_{\mathcal{P}}$ is large enough so that $\gamma \theta_{\mathcal{P}} > \|b\|$.

Assumption 3.1(ii) is closely related to the condition

$$\inf_{y \in \mathbb{K}^*} \sup_{x \in \mathbb{X}_n} \frac{\big\langle \mathcal{A} x, y \big\rangle}{\|x\| \|y\|_*} \geq \gamma,$$

which in the literature of numerical algorithms in infinite dimensional spaces, in particular the Galerkin discretization methods for partial differential equations, is often referred to as the *inf-sup* condition; see [21] for a comprehensive survey. To see this, note that for every $x \in \mathbb{X}_n$ the definitions in (3.2) imply that $\big\langle \mathcal{A} x, y \big\rangle = \big\langle \mathcal{A}_n \alpha, y \big\rangle = \alpha \bullet \mathcal{A}_n^* y$, $x = \sum_{i=1}^n \alpha_i x_i$. These conditions are in fact equivalent if the norm $\|\cdot\|_{\mathfrak{R}}$ is induced by the original norm on $\mathbb{X}$, i.e., $\|\alpha\|_{\mathfrak{R}} := \|\sum_{i=1}^n \alpha_i x_i\|$. We note that $\mathcal{A}_n^*$ maps an infinite dimensional space to a finite dimensional one, and as such Assumption 3.1(ii) effectively necessitates that the null-space of $\mathcal{A}_n^*$ intersects the positive cone $\mathbb{K}^*$ only at 0. In the following we show that this regularity condition leads to a zero duality gap between $\mathcal{P}_n$ and $\mathcal{D}_n$, as well as an upper bound for the dual optimizers. The latter turns out to be a critical quantity for the performance bounds of this study.

Proposition 3.2 (duality gap and bounded dual optimizers). *Under Assumption* 3.1(i)*, the duality gap between the programs $\mathcal{P}_n$ and $\mathcal{D}_n$ is zero, i.e., $J_n = \widetilde{J}_n$. If in addition Assumption* 3.1(ii) *holds, then for any optimizer $y_n^\star$ of the program $\mathcal{D}_n$ and any lower bound $J_n^{\mathrm{LB}} \leq J_n$ we have*

$$(3.3) \qquad \|y_n^\star\|_* \leq \theta_{\mathcal{D}} := \frac{\theta_{\mathcal{P}} \|\mathbf{c}\|_{\mathfrak{R}^*} - J_n^{\mathrm{LB}}}{\gamma \theta_{\mathcal{P}} - \|b\|} \leq \frac{2 \theta_{\mathcal{P}} \|\mathbf{c}\|_{\mathfrak{R}^*}}{\gamma \theta_{\mathcal{P}} - \|b\|}.$$

Proof. Since the elements $\{x_i\}_{i \leq n}$ are linearly independent, the feasible set of the decision variable α in program $\mathcal{P}_n$ is a bounded closed subset of a finite dimensional

space and hence compact. Thus, thanks to the feasibility Assumption 3.1(i) and compactness of the feasible set, the zero duality gap follows because

$$J_n = \inf_{\|\alpha\|_{\mathfrak{R}} \le \theta_{\mathcal{P}}} \Big\{ \alpha \bullet \mathbf{c} + \sup_{y \in \mathbb{K}^*} \big\langle b - \mathcal{A}_n \alpha, y \big\rangle \Big\} = \sup_{y \in \mathbb{K}^*} \inf_{\|\alpha\|_{\mathfrak{R}} \le \theta_{\mathcal{P}}} \Big\{ \big\langle b, y \big\rangle - \alpha \bullet (\mathcal{A}_n^* y - \mathbf{c}) \Big\} = \widetilde{J}_n,$$

where the first equality holds by the definition of the dual cone $\mathbb{K}^*$, and the second equality follows from Sion's minimax theorem [45, Theorem 4.2]. Thanks to the zero duality gap above, we have

$$J_n^{\mathrm{LB}} \le J_n = \widetilde{J}_n = \big\langle b, y_n^\star \big\rangle - \theta_{\mathcal{P}} \|\mathcal{A}_n^* y_n^\star - \mathbf{c}\|_{\mathfrak{R}^*} \le \big\langle b, y_n^\star \big\rangle - \theta_{\mathcal{P}} \|\mathcal{A}_n^* y_n^\star\|_{\mathfrak{R}^*} + \theta_{\mathcal{P}} \|\mathbf{c}\|_{\mathfrak{R}^*}.$$

By Assumption 3.1(ii), we then have

$$J_n \le \|b\| \|y_n^\star\|_* - \gamma \theta_{\mathcal{P}} \|y_n^\star\|_* + \theta_{\mathcal{P}} \|\mathbf{c}\|_{\mathfrak{R}^*} = \theta_{\mathcal{P}} \|\mathbf{c}\|_{\mathfrak{R}^*} - \big(\gamma \theta_{\mathcal{P}} - \|b\|\big) \|y_n^\star\|_*,$$

which together with the lower bound $J_n^{\mathrm{LB}} := -\theta_{\mathcal{P}} \|\mathbf{c}\|_{\mathfrak{R}^*} \le J_n$ concludes the proof. ∎

Proposition 3.2 effectively implies that in the program $\mathcal{D}_n$ one can add a norm constraint $\|y\|_* \le \theta_{\mathcal{D}}$ without changing the optimal value. The parameter $\theta_{\mathcal{D}}$ depends on J_n^{LB}, a lower bound for the optimal value of J_n. A simple choice for such a lower bound is $-\theta_{\mathcal{P}} \|\mathbf{c}\|_{\mathfrak{R}^*}$, but in particular problem instances one may be able to obtain a less conservative bound. We validate the assertions of Proposition 3.2 for long-run AC problems in the next section and for long-run discounted cost problems in Appendix A.

Program $\mathcal{P}_n$ is a restricted version of the original program $\mathcal{P}$ (also called an *inner approximation* [27, Definition 12.2.13]), and thus $J \le J_n$. However, under Assumption 3.1, we show that the gap $J_n - J$ can be quantified explicitly. To this end, we consider the projection mapping $\Pi_{\mathbb{A}}(x) := \arg\min_{x' \in \mathbb{A}} \|x' - x\|$ and the operator norm $\|\mathcal{A}\| := \sup_{\|x\| \le 1} \|\mathcal{A}x\|$ and define the set

$$\mathsf{B}_n := \left\{ \sum_{i=1}^n \alpha_i x_i \in \mathbb{X}_n \ : \ \|\alpha\|_{\mathfrak{R}} \le \theta_{\mathcal{P}} \right\}. \tag{3.4}$$

THEOREM 3.3 (semi-infinite approximation). *Let $x^\star$ and $y_n^\star$ be optimizers for the programs $\mathcal{P}$ and $\mathcal{D}_n$, respectively, and let $r_n := x^\star - \Pi_{\mathsf{B}_n}(x^\star)$ be the projection residual of the optimizer $x^\star$ onto the set B_n as defined in* (3.4). *Under Assumption* 3.1(i), *we have $0 \le J_n - J \le \big\langle r_n, \mathcal{A}^* y_n^\star - c \big\rangle$, where J_n and J are the optimal value of the programs $\mathcal{P}_n$ and $\mathcal{P}$. In addition, if Assumption* 3.1(ii) *holds, then*

$$0 \le J_n - J \le \big(\|c\|_* + \theta_{\mathcal{D}} \|\mathcal{A}\|\big) \|r_n\|, \tag{3.5}$$

where $\theta_{\mathcal{D}}$ is the dual optimizer bound introduced in (3.3).

Proof. The lower bound $0 \le J_n - J$ is trivial, and we only need to prove the upper bound. Note that since the optimizer $x^\star \in \mathbb{X}$ is a feasible solution of $\mathcal{P}$, then $\mathcal{A}x^\star - b \in \mathbb{K}$. By the definition of the dual cone $\mathbb{K}^*$, this implies that $\big\langle \mathcal{A}x^\star - b, y \big\rangle \ge 0$ $\forall\, y \in \mathbb{K}^*$. Since the dual optimizer $y_n^\star$ belongs to the dual cone $\mathbb{K}^*$, then

$$\begin{aligned} J_n - J &\le J_n - J + \big\langle \mathcal{A}x^\star - b, y_n^\star \big\rangle = J_n - \big\langle x^\star, c \big\rangle + \big\langle \mathcal{A}x^\star, y_n^\star \big\rangle - \big\langle b, y_n^\star \big\rangle \\ &= J_n + \big\langle x^\star, \mathcal{A}^* y_n^\star - c \big\rangle - \big\langle b, y_n^\star \big\rangle \\ &= J_n + \big\langle r_n, \mathcal{A}^* y_n^\star - c \big\rangle + \big\langle \Pi_{\mathsf{B}_n}(x^\star), \mathcal{A}^* y_n^\star - c \big\rangle - \big\langle b, y_n^\star \big\rangle \\ &= J_n + \big\langle r_n, \mathcal{A}^* y_n^\star - c \big\rangle + \widetilde{\alpha} \bullet \big(\mathcal{A}_n^* y_n^\star - \mathbf{c}\big) - \big\langle b, y_n^\star \big\rangle \end{aligned}$$

for some $\widetilde{\alpha} \in \mathbb{R}^n$ with norm $\|\widetilde{\alpha}\|_{\mathfrak{R}} \leq \theta_{\mathcal{P}}$; for the last line, see the definition of the operator $\mathcal{A}_n$ in (3.2) as well as the vector $\mathbf{c}$ in the program $\mathcal{P}_n$. Using the definition of the dual norm and the operators (3.2), one can deduce from above that

$$J_n - J \leq J_n + \big\langle r_n, \mathcal{A}^* y_n^\star - c \big\rangle + \theta_{\mathcal{P}} \|\mathcal{A}_n^* y_n^\star - \mathbf{c}\|_{\mathfrak{R}^*} - \big\langle b, y_n^\star \big\rangle = J_n + \big\langle r_n, \mathcal{A}^* y_n^\star - c \big\rangle - \widetilde{J}_n,$$

which in conjunction with the zero duality gap ($J_n = \widetilde{J}_n$) establishes the first assertion of the proposition. The second assertion is simply the consequence of the first part and the norm definitions, i.e.,

$$\big\langle r_n, \mathcal{A}^* y_n^\star - c \big\rangle \leq \|r_n\| \|c\|_* + \|\mathcal{A} r_n\| \|y_n^\star\|_* \leq \|r_n\| \Big(\|c\|_* + \|\mathcal{A}\| \|y_n^\star\|_* \Big).$$

Invoking the bound on the dual optimizer $y_n^\star$ from Proposition 3.2 completes the proof. ∎

Remark 3.4 (impact of norms on semi-infinite approximation). We note the following concerning the impact of the choice of norms on the approximation error:

(i) The only norm that influences the semi-infinite program $\mathcal{P}_n$ is $\|\cdot\|_{\mathfrak{R}}$ on $\mathbb{R}^n$. When it comes to the approximation error (3.5), the norm $\|\cdot\|_{\mathfrak{R}}$ may have an impact on the residual r_n only if the set B_n in (3.4) does not contain $\Pi_{\mathbb{X}_n}(x^\star)$, the projection $x^\star$ on the subspace $\mathbb{X}_n$, where $x^\star$ is an optimizer of the infinite program $\mathcal{P}$.

(ii) The norms of the dual pairs of vector spaces only appear in Theorem 3.3 to quantify the approximation error. Note that in (3.5) the stronger the norm on $\mathbb{X}$, the higher $\|r_n\|$, and the lower $\|c\|_*$ and $\|\mathcal{A}\|$. On the other hand, the stronger the norm on $\mathbb{B}$, the higher $\|b\|$ and $\|\mathcal{A}\|$ and the lower γ (cf. Assumption 3.1(ii)).

The error bound (3.5) can be further improved when $\mathbb{X}$ is a Hilbert space. In this case, let $\overline{\mathbb{X}}_n$ denote the orthogonal complement of $\mathbb{X}_n$. We define the *restricted* norms by

$$\|c\|_{*n} := \sup_{x \in \overline{\mathbb{X}}_n} \frac{\langle x, c \rangle}{\|x\|}, \qquad \|\mathcal{A}\|_n := \sup_{x \in \overline{\mathbb{X}}_n} \frac{\|\mathcal{A}x\|}{\|x\|}. \tag{3.6}$$

It is straightforward to see that by definition $\|c\|_{*n} \leq \|c\|_*$ and $\|\mathcal{A}\|_n \leq \|\mathcal{A}\|$.

COROLLARY 3.5 (Hilbert structure). *Suppose that $\mathbb{X}$ is a Hilbert space and $\|\cdot\|$ is the norm induced by the corresponding inner product. Let $\{x_i\}_{i \in \mathbb{N}}$ be an orthonormal dense family and $\|\cdot\|_{\mathfrak{R}} = \|\cdot\|_{\ell_2}$. Let $x^\star$ be an optimal solution for $\mathcal{P}$ and chose $\theta_{\mathcal{P}} \geq \|x^\star\|$. Under the assymptions of Theorem* 3.3*, we have*

$$0 \leq J_n - J \leq \big(\|c\|_n + \theta_{\mathcal{D}} \|\mathcal{A}\|_n \big) \big\| \Pi_{\overline{\mathbb{X}}_n}(x^\star) \big\|.$$

Proof. We first note that the ℓ_2-norm on $\mathbb{R}^n$ is indeed the norm induced by $\|\cdot\|$, since due to the orthonormality of $\{x_i\}_{i \in \mathbb{N}}$ we have

$$\|\alpha\|_{\mathfrak{R}} := \Big\| \sum_{i=1}^n \alpha_i x_i \Big\| = \sqrt{\sum_{i=1}^n \alpha_i^2 \|x_i\|^2} = \|\alpha\|_{\ell_2}.$$

If $\theta_{\mathcal{P}} \geq \|x^\star\|$, then $\Pi_{\mathsf{B}_n}(x^\star) = \Pi_{\mathbb{X}_n}(x^\star)$, i.e., the projection of the optimizer $x^\star$ on the ball B_n is in fact the projection onto the subspace $\mathbb{X}_n$. Therefore, thanks to the orthonormality, the projection residual $r_n = x^\star - \Pi_{\mathbb{X}_n}(x^\star)$ belongs to the orthogonal complement $\overline{\mathbb{X}}_n$. Thus, following the same reasoning as in the proof of Theorem 3.3, one arrives at a bound similar to (3.5) but using the restricted norms (3.6); recall that the norm in a Hilbert space is self-dual. ∎

3.3. Semi-infinite results in the MDP setting. We now return to the MDP setting in section 2, and in particular the AC problem (2.2), to investigate the application of the proposed approximation scheme. Recall that the AC problem (2.1) can be recast in an LP framework in the form of $\mathcal{P}$; see (2.3). To complete this transition to the dual pairs, we introduce the spaces

$$\text{(3.7)} \qquad \begin{cases} \mathbb{X} = \mathbb{R} \times \mathscr{L}(S), & \mathbb{C} = \mathbb{R} \times \mathcal{M}(S), \\ \mathbb{B} = \mathscr{L}(K), & \mathbb{Y} = \mathcal{M}(K), \\ \mathbb{K} = \mathscr{L}_+(K), & \mathbb{K}^* = \mathcal{M}_+(K). \end{cases}$$

The bilinear form between each pair $(\mathbb{X}, \mathbb{C})$ and $(\mathbb{B}, \mathbb{Y})$ is defined in an obvious way (cf. (2.3)). The linear operator $\mathcal{A} : \mathbb{X} \to \mathbb{B}$ is defined as $\mathcal{A}(\rho, u)(s, a) := -\rho - u(s) + Qu(s, a)$, and it can be shown to be weakly continuous [27, p. 220]. On the pair $(\mathbb{X}, \mathbb{C})$ we consider the norms

$$\text{(3.8a)} \qquad \begin{cases} \|x\| = \|(\rho, u)\| = \max\big\{|\rho|, \|u\|_{\mathrm{L}}\big\} = \max\big\{|\rho|, \|u\|_\infty, \sup\limits_{s,s' \in S} \frac{u(s)-u(s')}{\|s-s'\|_{\ell_\infty}}\big\}, \\ \|c\|_* := \sup\limits_{\|x\| \le 1} \big\langle x, c \big\rangle = |c_1| + \sup\limits_{\|u\|_{\mathrm{L}} \le 1} \int_S u(s) c_2(\mathrm{d}s) = |c_1| + \|c_2\|_{\mathrm{W}}. \end{cases}$$

Recall that $\|\cdot\|_{\mathrm{L}}$ is the Lipschitz norm on $\mathscr{L}(S)$ whose dual norm $\|\cdot\|_{\mathrm{W}}$ in $\mathcal{M}(S)$ is known as the Wasserstein norm [48, p. 105]. The adjoint operator $\mathcal{A}^* : \mathbb{Y} \to \mathbb{C}$ is given by $\mathcal{A}^* y(\cdot) := \big(-\big\langle \mathbb{1}, y \big\rangle, -y(\cdot \times A) + yQ(\cdot)\big)$, where $\mathbb{1}$ is the constant function in $\mathscr{L}(S)$ with value 1. In the second pair $(\mathbb{B}, \mathbb{Y})$, we consider the norms

$$\text{(3.8b)} \qquad \begin{cases} \|b\| = \|b\|_{\mathrm{L}} := \max\big\{\|b\|_\infty, \sup\limits_{k,k' \in K} \frac{b(k)-b(k')}{\|k-k'\|_{\ell_\infty}}\big\}, \\ \|y\|_* := \sup\limits_{\|b\|_{\mathrm{L}} \le 1} \big\langle b, y \big\rangle = \|y\|_{\mathrm{W}}. \end{cases}$$

A commonly used norm on the set of measures is the total variation whose dual (variational) characterization is associated with $\|\cdot\|_\infty$ in the space of continuous functions [27, p. 2]. We note that in the positive cone $\mathbb{K}^* = \mathcal{M}_+(K)$ the total variation and Wasserstein norms indeed coincide.

Following the construction in $\mathcal{P}_n$, we consider a collection of n-linearly independent, normalized functions $\{u_i\}_{i \le n}$, $\|u_i\|_{\mathrm{L}} = 1$, and define the semi-infinite approximation of the AC problem (2.2) by

$$\text{(3.9)} \qquad -J_n^{\mathrm{AC}} = \begin{cases} \inf\limits_{(\rho,\alpha) \in \mathbb{R} \times \mathbb{R}^n} & -\rho \\ \text{s.t.} & \rho + \sum\limits_{i=1}^n \alpha_i \big(u_i(s) - Qu_i(s,a)\big) \le \psi(s,a) \quad \forall (s,a) \in K, \\ & \|\alpha\|_{\mathfrak{R}} \le \theta_{\mathcal{P}}. \end{cases}$$

Comparing with the program $\mathcal{P}_n$, we note that the finite dimensional subspace $\mathbb{X}_n \subset \mathbb{R} \times \mathscr{L}(S)$ is the subspace spanned by the basis elements $x_0 = (1, 0)$ and $x_i = (0, u_i)$ $\forall\, i \in \{1, \ldots, n\}$, i.e., the subspace $\mathbb{X}_n$ is in fact $n + 1$ dimensional. Moreover, the norm constraint in (3.9) is only imposed on the second coordinate of the decision variables (ρ, α) (i.e., $\|\alpha\|_{\mathfrak{R}} \le \theta_{\mathcal{P}}$). The following lemmas address the operator norm and the respective regularity requirements of Assumption 3.1 for the program (3.9).

LEMMA 3.6 (MDP operator norm). *In the AC problem* (2.2) *under Assumption* 2.1(ii) *with the specific norms defined in* (3.8)*, the linear operator norm satisfies* $\|I - Q\| := \sup_{\|u\|_{\mathrm{L}} \le 1} \|u - Qu\|_{\mathrm{L}} \le 1 + \max\{L_Q, 1\}$.

Proof. Using the triangle inequality it is straightforward to see that

$$\|I-Q\| = \sup_{u\in\mathscr{L}(S)} \frac{\|u-Qu\|_{\mathrm{L}}}{\|u\|_{\mathrm{L}}} \le 1 + \sup_{u\in\mathscr{L}(S)} \frac{\|Qu\|_{\mathrm{L}}}{\|u\|_{\mathrm{L}}} \le 1 + \sup_{u\in\mathscr{L}(S)} \frac{\|Qu\|_{\mathrm{L}}}{\|u\|_{\infty}}$$
$$\le 1 + \max\Big\{L_Q, \sup_{u\in\mathscr{L}(S)} \frac{\|Qu\|_{\infty}}{\|u\|_{\infty}}\Big\} \le 1 + \max\{L_Q, 1\},$$

where the second line is an immediate consequence of Assumption 2.1(ii) and the fact that the operator Q is a stochastic kernel. Hence, $|Qu(s,a)| = |\int_S u(y)Q(\mathrm{d}y|s,a)| \le \|u\|_\infty(\int_S Q(\mathrm{d}y|s,a)) = \|u\|_\infty$. ∎

LEMMA 3.7 (MDP semi-infinite regularity). *Consider the AC program* (2.2) *under Assumption* 2.1. *Then, Assumption* 3.1 *holds for the semi-infinite counterpart in* (3.9) *for any positive* $\theta_{\mathcal{P}}$ *and all sufficiently large* γ. *In particular, the dual optimizer bound in Proposition* 3.2 *simplifies to* $\|y_n^\star\|_{\mathrm{W}} \le \theta_{\mathcal{D}} = 1$.

Proof. Since K is compact, for any nonnegative $\theta_{\mathcal{P}}$, the program (3.9) is feasible and the optimal value is bounded; recall that $\|(Q-I)u_i\|_{\mathrm{L}} \le 1 + \max\{L_Q, 1\}$ from Lemma 3.6 and $\|\psi\|_\infty < \infty$ thanks to Assumption 2.1(iii). Hence, the optimal value of (3.9) is bounded and, without loss of generality, one can add a redundant constraint $|\rho| \le \omega^{-1}\theta_{\mathcal{P}}$, where ω is a sufficiently small positive constant. In this view, the last constraint $\|\alpha\|_{\mathfrak{R}} \le \theta_{\mathcal{P}}$ may be replaced with

$$\|(\rho,\alpha)\|_\omega := \max\{\omega|\rho|, \|\alpha\|_{\mathfrak{R}}\} \le \theta_{\mathcal{P}}, \tag{3.10}$$

where $\|\cdot\|_\omega$ can be cast as the norm on the pair $(\rho,\alpha) \in \mathbb{R}\times\mathbb{R}^{n+1}$. Using the ω-norm as defined in (3.10), we can now directly translate the program (3.9) into the semi-infinite framework of $\mathcal{P}_n$. As mentioned above, the feasibility requirement in Assumption 3.1(i) immediately holds. In addition, observe that for every $y \in \mathbb{K}^*$ we have

$$\|\mathcal{A}_n^* y\|_{\omega^*} = \sup_{\|(\rho,\alpha)\|_\omega\le 1} (\rho,\alpha)\cdot\big[-\langle \mathbb{1}, y\rangle, \langle Qu_1 - u_1, y\rangle, \ldots, \langle Qu_n - u_n, y\rangle\big]$$
$$= \sup_{\omega|\rho|\le 1} -\rho\langle \mathbb{1}, y\rangle + \sup_{\|\alpha\|_{\mathfrak{R}}\le 1} \alpha\cdot\big[\langle Qu_1 - u_1, y\rangle, \ldots, \langle Qu_n - u_n, y\rangle\big]$$
$$\ge \omega^{-1}\|y\|_{\mathrm{W}},$$

where the third line above follows from the equality $\langle \mathbb{1}, y\rangle = \|y\|_{\mathrm{W}}$ for every y in the positive cone $\mathbb{K}^*$ and the fact that the second term in the second line is nonnegative. Since ω can be arbitrarily close to 0, the inf-sup requirement Assumption 3.1(ii) holds for all sufficiently large $\gamma = \omega^{-1}$. The second assertion of the lemma follows from the bound (3.3) in Proposition 3.2. To show this, recall that in the MDP setting $c = (-1, 0) \in \mathbb{R}\times\mathcal{M}(S)$ (cf. (2.3)) with the respective vector $\mathbf{c} = [-1, 0, \ldots, 0] \in \mathbb{R}\times\mathbb{R}^n$ (cf. $\mathcal{P}_n$). Thus, $\|\mathbf{c}\|_{\omega^*} = \sup_{\|(\rho,\alpha)\|_\omega\le 1} (\rho,\alpha)\cdot[-1, 0, \ldots, 0] = \omega^{-1}$, which helps simplifying the bound (3.3) to

$$\|y_n^\star\|_{\mathrm{W}} \le \theta_{\mathcal{D}} := \frac{\theta_{\mathcal{P}}\|\mathbf{c}\|_{\mathfrak{R}^*} - J_n^{\mathrm{LB}}}{\gamma\theta_{\mathcal{P}} - \|b\|} = \frac{\theta_{\mathcal{P}}\omega^{-1} + \|\psi\|_\infty}{\omega^{-1}\theta_{\mathcal{P}} - \|\psi\|_{\mathrm{L}}},$$

which delivers the desired assertion when ω tends to 0. ∎

Remark 3.8 (AC dual optimizers bound). As opposed to the general LP in Proposition 3.2, Lemma 3.7 implies that the dual optimizers for the AC problem are not

influenced by the primal norm bound $\theta_{\mathcal{P}}$ and are uniformly bounded by 1. In fact, this result can be strengthened to $\|y_n^\star\|_{\mathrm{W}} = 1$ due to the special minimax structure of the AC program (3.9). This refinement is not needed at this stage and we postpone the discussion to section 5.2. The feature discussed in this remark, however, does not hold for the class of long-run discounted cost problems; see Lemma A.2 in Appendix A.

Now we are in a position to translate Theorem 3.3 to the MDP setting for the AC problem (2.2).

COROLLARY 3.9 (MDP semi-infinite approximation). *Let J^{AC} and $u^\star$ be the optimal value and an optimizer for the AC program* (2.2)*, respectively. Consider the semi-infinite program* (3.9) *where $\theta_{\mathcal{P}} > \|\psi\|_{\mathrm{L}}$, and let $\mathbb{U}_n := \{\sum_{i=1}^n \alpha_i u_i \ : \ \|\alpha\|_{\mathfrak{R}} \leq \theta_{\mathcal{P}}\}$. Then, the optimal value of* (3.9) *satisfies the inequality*

$$0 \leq J^{\mathrm{AC}} - J_n^{\mathrm{AC}} \leq \big(1 + \max\{L_Q, 1\}\big)\big\|u^\star - \Pi_{\mathbb{U}_n}(u^\star)\big\|_{\mathrm{L}}.$$

Proof. We first note that the existence of the optimizer $u^\star$ is guaranteed under Assumption 2.1 [27, Theorem 12.4.2]. The proof is a direct application of Theorem 3.3 under the preliminary results in Lemmas 3.7 and 3.6. Observe that the projection error is $r_n := (\rho^\star, u^\star) - \Pi_{\mathbb{U}_n}(\rho^\star, u^\star) = \big(0, u^\star - \Pi_{\mathbb{U}_n}(u^\star)\big)$, resulting in $\langle r_n, c\rangle = 0$. Thanks to this observation, Lemma 3.6, the assertion of Theorem 3.3 translates to

$$\begin{aligned} 0 \leq J^{\mathrm{AC}} - J_n^{\mathrm{AC}} = J_n - J \leq \langle r_n, \mathcal{A}^* y_n^\star - c\rangle = \langle \mathcal{A} r_n, y_n^\star\rangle \leq \|I - Q\|\, \|r_n\|_{\mathrm{L}}\, \|y_n^\star\|_{\mathrm{W}} \\ \leq (1 + \max\{L_Q, 1\})\|u^\star - \Pi_{\mathbb{U}_n}(u^\star)\|_{\mathrm{L}}. \end{aligned}$$

∎

Observe that if from the beginning we consider the norm $\|\cdot\|_\infty$ on the spaces $\mathbb{X}$ and $\mathbb{B}$, it is not difficult to see that the operator norm in Lemma 3.6 simplifies to 2 (recall that Q is a stochastic kernel). Thus, the semi-infinite bound reduces to $J^{\mathrm{AC}} - J_n^{\mathrm{AC}} \leq 2\|u^\star - \Pi_{\mathbb{U}_n}(u^\star)\|_\infty$. One may arrive at this particular observation through a more straightforward approach: Using the shorthand notation $(Q-I)u := Qu - u$, we have

$$\begin{aligned} J^{\mathrm{AC}} - J_n^{\mathrm{AC}} &\leq \min_{k\in K}\Big(\big(Q-I\big)u^\star(k) + \psi(k)\Big) - \min_{k\in K}\Big(\big(Q-I\big)\Pi_{\mathbb{U}_n}(u^\star)(k) + \psi(k)\Big) \\ &\leq \max_{k\in K}\big(Q-I\big)\big(u^\star - \Pi_{\mathbb{U}_n}(u^\star)\big)(k) \leq \big\|\big(Q-I\big)\big(u^\star - \Pi_{\mathbb{U}_n}(u^\star)\big)\big\|_\infty \\ &\leq 2\big\|u^\star - \Pi_{\mathbb{U}_n}(u^\star)\big\|_\infty. \end{aligned}$$

Theorem 3.3 is a generalization to the above observation in two respects:

- It holds for a general LP that, unlike the AC problem (2.2), may not necessarily enjoy a min-max structure.
- The result reflects how the bound on the decision space (i.e., $\theta_{\mathcal{P}}$ in $\mathcal{P}_n$) influences the dual optimizers as well as the approximation performance in generic normed spaces.

The latter feature is of particular interest as the boundedness of the decision space is often an a priori requirement for optimization algorithms; see, for instance, [37] and the results in section 5. The approximation error from the original infinite LP to the semi-infinite version is quantified in terms of the projection residual of the value function. Clearly, this is where the choice of the finite dimensional ball $\mathbb{U}_n$ plays a crucial role. We close this section with a remark on this point.

Remark 3.10 (projection residual). The residual error $\big\|u^\star - \Pi_{\mathbb{U}_n}(u^\star)\big\|_{\mathrm{L}}$ can be approximated by leveraging results from the literature on universal function approximation. Prior information about the value function $u^\star$ may offer explicit quantitative

bounds. For instance, for MDP under Assumption 2.1 we know that $u^\star$ is Lipschitz continuous. For an appropriate choice of basis functions, we can therefore ensure a convergence rate of $n^{-1/\dim(S)}$, where $\dim(S)$ is the dimension of the state-action set S; see, for instance, [22] for polynomials and [41] for the Fourier basis functions.

4. Semi-infinite to finite programs: Randomized approach. We study conditions under which one can provide a finite approximation to the semi-infinite programs of the form $\mathcal{P}_n$ that are in general known to be computationally intractable—NP-hard [4, p. 16]. We approach this goal by deploying tools from two areas, leading to different theoretical guarantees for the proposed solutions. This section focuses on a randomized approach and the next section is dedicated to an iterative gradient-based descent method. The solution of each of these methods comes with a priori as well as a posteriori performance certificates.

4.1. Randomized approach. We start with a lemma suggesting a simple bound on the norm of the operator $\mathcal{A}_n$ in (3.2). We will use the bound to quantify the approximation error of our proposed solutions.

LEMMA 4.1 (semi-infinite operator norm). *Consider the operator $\mathcal{A}_n : \mathbb{R}^n \to \mathbb{B}$ as defined in* (3.2)*. Then,*

$$\|\mathcal{A}_n\| := \sup_{\alpha\in\mathbb{R}^n} \frac{\|\mathcal{A}_n\alpha\|}{\|\alpha\|_{\mathfrak{R}}} \le \|\mathcal{A}\|\varrho_n, \qquad \varrho_n := \sup_{\|\alpha\|_{\mathfrak{R}}\le 1} \|\alpha\|_{\ell_1}, \tag{4.1}$$

where the constant ϱ_n is the equivalence ratio between the norms $\|\cdot\|_{\mathfrak{R}}$ and $\|\cdot\|_{\ell_1}$.[1]

Proof. See [34, Lemma 4.1]. The proof follows directly from the definition of the operator norm, that is,

$$\|\mathcal{A}_n\alpha\| = \Big\|\sum_{i=1}^n \alpha_i \mathcal{A}x_i\Big\| \le \|\mathcal{A}\|\Big\|\sum_{i=1}^n \alpha_i x_i\Big\|,$$

together with the inequality $\big\|\sum_{i=1}^n \alpha_i x_i\big\| \le \|\alpha\|_{\ell_1} \max_{i\le n}\|x_i\| = \|\alpha\|_{\ell_1}$, which concludes the proof. ∎

Since $\mathbb{K}$ is a closed convex cone, then $\mathbb{K}^{**} = \mathbb{K}$ [1, p. 40], and as such the conic constraint in program $\mathcal{P}_n$ can be reformulated as

$$\mathcal{A}_n\alpha \succeq_{\mathbb{K}} b \qquad \Longleftrightarrow \qquad \big\langle \mathcal{A}_n\alpha - b, y\big\rangle \ge 0 \quad \forall y \in \mathcal{K} := \mathcal{E}\{y \in \mathbb{K}^* : \|y\|_* = 1\}, \tag{4.2}$$

where $\mathcal{E}\{B\}$ denotes the extreme points of the set B, i.e., the set of points that cannot be represented as a strict convex combination of some other elements of the set. Notice that the norm constraint as well as the restriction to the extreme points in the definition of $\mathcal{K}$ in (4.2) do not sacrifice any generality, as conic constraints are homogeneous. These restrictions are introduced to improve the approximation errors. In what follows, however, one can safely replace the set $\mathcal{K}$ with any subset of the cone $\mathbb{K}^*$ whose closure contains $\mathcal{K}$. This adjustment may be taken into consideration for computational advantages. Let $\mathbb{P}$ be a Borel probability measure supported on $\mathcal{K}$, and $\{y_j\}_{j\le N}$ be independent, identically distributed (i.i.d.) samples generated from $\mathbb{P}$. Consider the *scenario* counterpart of the program $\mathcal{P}_n$ defined as

$$(\mathcal{P}_{n,N}) \qquad J_{n,N} := \left\{ \begin{array}{ll} \min\limits_{\alpha\in\mathbb{R}^n} & \alpha\cdot\mathbf{c} \\ \text{s.t.} & \alpha\cdot\mathcal{A}_n^* y_j \ge \big\langle b, y_j\big\rangle, \quad j \in \{1,\dots,N\}, \\ & \|\alpha\|_{\mathfrak{R}} \le \theta_{\mathcal{P}}, \end{array} \right.$$

[1] The constant ϱ_n is indexed by n as it potentially depends on the dimension of $\alpha \in \mathbb{R}^n$.

where the adjoint operator $\mathcal{A}_n^* : \mathbb{B} \to \mathbb{R}^n$ is introduced in (3.2). The optimization problem $\mathcal{P}_{n,N}$ is a standard finite convex program and thus computationally tractable whenever the norm constraint $\|\alpha\|_{\mathfrak{R}} \le \theta_{\mathcal{P}}$ is tractable. Program $\mathcal{P}_{n,N}$ is a relaxation of $\mathcal{P}_n$, i.e., $J_n \ge J_{n,N}$; note that $J_{n,N}$ is a random variable, and therefore the relaxation error $J_n - J_{n,N}$ can only be interpreted in a probabilistic sense.

DEFINITION 4.2 (tail bound). *Given a probability measure* $\mathbb{P}$ *supported on* $\mathcal{K}$*, we define the function* $p : \mathbb{R}^n \times \mathbb{R}_+ \to [0,1]$ *as*

$$p(\alpha,\zeta) := \mathbb{P}\Big[y \ : \ \sigma_{\mathcal{K}}(-\mathcal{A}_n\alpha + b) < \big\langle -\mathcal{A}_n\alpha + b, y\big\rangle + \zeta\Big],$$

where $\sigma_{\mathcal{K}}(\cdot) := \sup_{y\in\mathcal{K}} \langle \cdot, y\rangle$ *is the support function of* $\mathcal{K}$*. We call* $h : \mathbb{R}^n \times [0,1] \to \mathbb{R}_+$ *a* tail bound *(TB) of the program* $\mathcal{P}_{n,N}$ *if* $\forall\, \varepsilon \in [0,1]$ *and* α *we have*

$$h(\alpha,\varepsilon) \ge \sup\big\{\zeta \ : \ p(\alpha,\zeta) \le \varepsilon\big\}.$$

The TB function in Definition 4.2 can be interpreted as a *shifted* quantile function of the mapping $y \mapsto \big\langle -\mathcal{A}_n\alpha + b, y\big\rangle$ on $\mathcal{K}$—the "shift" is referred to the maximum value of the mapping which is $\sigma_{\mathcal{K}}(-\mathcal{A}_n\alpha + b)$. TB functions depend on the probability measure $\mathbb{P}$ generating the scenarios $\{y_j\}_{j\le N}$ in the program $\mathcal{P}_{n,N}$, as well as the properties of the optimization problem. Definition 4.2 is rather abstract and not readily applicable. The following example suggests a more explicit, but not necessarily optimal, candidate for a TB.

Example 4.3 (TB candidate). Let $g : \mathbb{R}_+ \to [0,1]$ be a nondecreasing function such that for any $\kappa \in \mathcal{K}$ we have $g(\gamma) \le \mathbb{P}\big[\mathsf{B}_\gamma(\kappa)\big]$, where $\mathsf{B}_\gamma(\kappa)$ is the open ball centered at κ with radius γ; note that function g depends on the choice of the norm on $\mathbb{Y}$. Then, a candidate for a TB function of the program $\mathcal{P}_{n,N}$ is

$$h(\alpha,\varepsilon) := \|\mathcal{A}_n\alpha - b\|g^{-1}(\varepsilon) \le \big(\varrho_n\|\mathcal{A}\|\|\alpha\|_{\mathfrak{R}} + \|b\|\big)g^{-1}(\varepsilon),$$

where the inverse function is understood as $g^{-1}(\varepsilon) := \sup\{\gamma \in \mathbb{R}_+ \ : g(\gamma) \le \varepsilon\}$, and ϱ_n is the constant ratio defined in (4.1).

To see this note that according to Definition 4.2 we have

$$\begin{aligned} p(\alpha,\zeta) &= \mathbb{P}\left[y \ : \ \sup_{\kappa\in\mathcal{K}} \big\langle -\mathcal{A}_n\alpha + b, \kappa - y\big\rangle < \zeta\right] = \inf_{\kappa\in\mathcal{K}} \mathbb{P}\left[y \ : \ \big\langle -\mathcal{A}_n\alpha + b, \kappa - y\big\rangle < \zeta\right] \\ &\ge \inf_{\kappa\in\mathcal{K}} \mathbb{P}\left[y \ : \ \|\mathcal{A}_n\alpha - b\|\|y - \kappa\|_* < \zeta\right] = \inf_{\kappa\in\mathcal{K}} \mathbb{P}\left[\mathsf{B}_{\gamma(\zeta)}(\kappa)\right] \ge g(\gamma(\zeta)), \end{aligned}$$

where $\gamma(\zeta) := \zeta\|\mathcal{A}_n\alpha - b\|^{-1}$. Thus, if $p(\alpha,\zeta) \le \varepsilon$, then $g(\gamma(\zeta)) \le \varepsilon$ and by construction of the inverse function g^{-1} we have $\zeta\|\mathcal{A}_n\alpha - b\|^{-1} \le g^{-1}(\varepsilon)$. In view of Definition 4.2, this observation readily suggests that the function $h(\alpha,\varepsilon) := \|\mathcal{A}_n\alpha - b\|g^{-1}(\varepsilon)$ is indeed a TB candidate, and the suggested upper bound follows readily from Lemma 4.1.

THEOREM 4.4 (randomized approximation error). *Consider the programs* $\mathcal{P}_n$ *and* $\mathcal{P}_{n,N}$ *with the associated optimum values* J_n *and* $J_{n,N}$*, respectively. Let Assumption* 3.1 *hold,* $\alpha_N^\star$ *be the optimizer of the program* $\mathcal{P}_{n,N}$*, and the function* h *be a TB as in Definition* 4.2*. Given* ε, β *in* $(0,1)$*, we define*

$$\mathsf{N}(n,\varepsilon,\beta) := \min\Big\{N \in \mathbb{N} \ : \ \sum_{i=0}^{n-1} \binom{N}{i} \varepsilon^i (1-\varepsilon)^{N-i} \le \beta\Big\}. \tag{4.3}$$

For all positive parameters ε, β, and $N \geq \mathsf{N}(n, \varepsilon, \beta)$ we have

$$\mathbb{P}^N\Big[0 \leq J_n - J_{n,N} \leq \theta_{\mathcal{D}} h\big(\alpha^\star_N, \varepsilon\big)\Big] \geq 1 - \beta, \tag{4.4a}$$

where the constant $\theta_{\mathcal{D}}$ is defined as in (3.3)*. In particular, suppose the function h is the TB candidate from Example* 4.3 *with corresponding g function, and*

$$N \geq \mathsf{N}\big(n, g(z_n\varepsilon), \beta\big), \qquad z_n := \Big(\theta_{\mathcal{D}}\big(\theta_{\mathcal{P}}\varrho_n\|\mathcal{A}\| + \|b\|\big)\Big)^{-1}, \tag{4.4b}$$

where ϱ_n is the ratio constant defined in Lemma 4.1*. We then have*

$$\mathbb{P}^N\Big[0 \leq J_n - J_{n,N} \leq \varepsilon\Big] \geq 1 - \beta\,. \tag{4.4c}$$

Theorem 4.4 extends the result [35, Theorem 3.6] in two respects:

- The bounds (4.4) are described in terms of a generic norm and the corresponding dual optimizer bound.
- Through the optimizer of $\mathcal{P}_{n,N}$, the bounds involve an a posteriori element (cf. (4.4a) to (4.4c)).

Before proceeding with the proof, we first remark on the complexity of the a priori bound of Theorem 4.4, its implications for an appropriate choice of $\theta_{\mathcal{P}}$, and its dependence on the dual pair norms.

Remark 4.5 (curse of dimensionality). The TB function h of Example 4.3 may grow exponentially in the dimension of the support set $\mathcal{K}$ (i.e., $h(\alpha, \varepsilon) \propto \varepsilon^{-\dim(\mathcal{K})}$). Since $\mathsf{N}(n, \cdot, \beta)$ admits a linear growth rate, the a priori bound (4.4c) effectively leads to an exponential number of samples in the precision level ε, an observation related to the curse of dimensionality [35, Remark 3.9]. To mitigate this inherent computational complexity, one may resort to a more elegant sampling approach so that the required number of samples N has a sublinear rate in the second argument; see [36].

Remark 4.6 (Optimal choice of $\theta_{\mathcal{P}}$). In view of the a priori error in Theorem 4.4, the parameter $\theta_{\mathcal{P}}$ may be chosen so as to minimize the required number of samples. To this end, it suffices to maximize z_n defined in (4.4b) over all $\theta_{\mathcal{P}} > \|b\|\gamma^{-1}$ (see Assumption 3.1(ii)), where $\theta_{\mathcal{D}}$ is defined in (3.3). One can show that the optimal choice in this respect is analytically available as

$$\theta^\star_{\mathcal{P}} := \frac{\|b\|}{\gamma} + \sqrt{\Big(\frac{\|b\|}{\gamma} + \frac{\|b\|}{\varrho_n\|\mathcal{A}\|}\Big)\Big(\frac{\|b\|}{\gamma} - \frac{J^{\mathrm{LB}}_n}{\|\mathbf{c}\|_{\mathfrak{R}_*}}\Big)}\,,$$

where J^{LB}_n is a lower bound on the optimal value of $\mathcal{P}_n$ used in (3.3).

Remark 4.7 (norm impact on finite approximation). In addition to what has already been highlighted in Remark 3.4, the choice of norms in the dual pairs of normed vector spaces also has an impact on the function $g^{-1}(\varepsilon)$. More specifically, the stronger the norm in the space $\mathbb{B}$, the larger the balls in the dual space $\mathbb{Y}$, and thus the smaller the function g^{-1}.

To prove Theorem 4.4 we need a few preparatory results.

LEMMA 4.8 (perturbation function). *Given $\delta \in \mathbb{B}$, consider the* δ-perturbed *program of $\mathcal{P}_n$ defined as*

$$(\mathcal{P}_n(\delta)) \qquad J_n(\delta) := \left\{\begin{array}{ll} \inf\limits_{\alpha\in\mathbb{R}^n} & \alpha \bullet \mathbf{c} \\ \text{s.t.} & \mathcal{A}_n\alpha \succeq_{\mathbb{K}} b - \delta, \\ & \|\alpha\|_{\mathfrak{R}} \leq \theta_{\mathcal{P}}. \end{array}\right.$$

Under Assumption 3.1*, we then have* $J_n - J_n(\delta) \leq \langle \delta, y_n^\star \rangle$*, where* $y_n^\star$ *is an optimizer of* $\mathcal{D}_n$*.*

Proof. For the proof we first introduce the dual program of $\mathcal{P}_n(\delta)$:

$$(\mathcal{D}_n(\delta)) \qquad \widetilde{J}_n(\delta) := \left\{ \begin{array}{ll} \sup\limits_{y} & \langle b - \delta, y \rangle - \theta_{\mathcal{P}} \|\mathcal{A}_n^* y - \mathbf{c}\|_{\mathfrak{R}^*} \\ \text{s.t.} & y \in \mathbb{K}^*. \end{array} \right.$$

We then have

$$\begin{aligned} J_n - J_n(\delta) = \widetilde{J}_n - J_n(\delta) &= \langle b, y_n^\star \rangle - \theta_{\mathcal{P}} \|\mathcal{A}_n^* y_n^\star - \mathbf{c}\|_{\mathfrak{R}^*} - J_n(\delta) \\ &= \langle \delta, y_n^\star \rangle + \langle b - \delta, y_n^\star \rangle - \theta_{\mathcal{P}} \|\mathcal{A}_n^* y_n^\star - \mathbf{c}\|_{\mathfrak{R}^*} - J_n(\delta) \\ &\leq \langle \delta, y_n^\star \rangle + \widetilde{J}_n(\delta) - J_n(\delta) \leq \langle \delta, y_n^\star \rangle, \end{aligned}$$

where the first line follows from the strong duality (gap-free) between $\mathcal{P}_n$ and $\mathcal{D}_n$ by Proposition 3.2. The third line is due to the fact that $y_n^\star$ is a feasible solution of $\mathcal{D}_n(\delta)$, and the last line follows from weak duality between $\mathcal{P}_n(\delta)$ and $\mathcal{D}_n(\delta)$. ∎

LEMMA 4.9 (perturbation error). *Let* $\alpha_N^\star$ *be an optimal solution of* $\mathcal{P}_{n,N}$ *and assume that* $\delta \in \mathbb{B}$ *satisfies the conic inequality* $\mathcal{A}_n \alpha_N^\star \succeq_{\mathbb{K}} b - \delta$*. Then, under Assumption* 3.1*, we have* $0 \leq J_n - J_{n,N} \leq \langle \delta, y_n^\star \rangle$*.*

Proof. The lower bound on $J_n - J_{n,N}$ is trivial since $\mathcal{P}_{n,N}$ is a relaxation of $\mathcal{P}_n$. For the upper bound the requirement on δ in the program $\mathcal{P}_n(\delta)$ implies that $\alpha_N^\star$ is a feasible solution of $\mathcal{P}_n(\delta)$. We then have $J_{n,N} \geq J_n(\delta)$, and thus $0 \leq J_n - J_{n,N} \leq J_n - J_n(\delta)$. Applying Lemma 4.8 completes the proof. ∎

The following fact follows readily from Definition 4.2; see [34, Lemma 4.10] for a formal proof in this regard.

FACT 4.10 (TB lower bound). *If* $\alpha \in \mathbb{R}^n$ *satisfies* $\mathbb{P}\left[y : \langle \mathcal{A}_n \alpha - b, y \rangle < 0\right] \leq \varepsilon$*, for any TB function in the sense of Definition* 4.2 *we have* $\sigma_{\mathcal{K}}(-\mathcal{A}_n \alpha + b) \leq h(\alpha, \varepsilon)$*.*

We follow our discussion with a result from randomized optimization in a convex setting.

THEOREM 4.11 (finite-sample probabilistic feasibility [10, Theorem 1]). *Assume that the program* $\mathcal{P}_{n,N}$ *admits a unique minimizer* $\alpha_N^\star$*.*[2] *If* $N \geq \mathsf{N}(n, \varepsilon, \beta)$ *as defined in* (4.3)*, then with confidence at least* $1 - \beta$ *(across multiscenarios* $\{y_j\}_{j \leq N} \subset \mathcal{K}$*) we have* $\mathbb{P}\big[y \ : \ \langle \mathcal{A}_n \alpha_N - b, y \rangle < 0\big] \leq \varepsilon$*.*

We are now in a position to prove Theorem 4.4.

Proof of Theorem 4.4. By definition of the support function we know that $\sigma_{\mathcal{K}}(\delta) = \sigma_{\mathsf{conv}(\mathcal{K})}(\delta)$, where $\mathsf{conv}(\mathcal{K})$ is the convex hull of $\mathcal{K}$. Recall that by definition of the set $\mathcal{K}$ in (4.2), we also have $y/\|y\|_* \in \mathsf{conv}(\mathcal{K})$ for any $y \in \mathbb{K}^*$. Thus, for any $\delta \in \mathbb{B}$ and $y \in \mathbb{K}^*$ we have $\langle \delta, y \rangle \leq \|y\|_* \sigma_{\mathcal{K}}(\delta)$. This leads to

$$0 \leq J_n - J_{n,N} \leq \big\langle -\mathcal{A}_n \alpha_N^\star + b, y_n^\star \big\rangle \leq \|y_n^\star\|_* \sigma_{\mathcal{K}}(-\mathcal{A}_n \alpha_N^\star + b),$$

where the second inequality is due to Lemma 4.9 as $\delta = -\mathcal{A}_n \alpha_N^\star + b$ clearly satisfies the requirements. By Fact 4.10 and Theorem 4.11, we know that with probability at least $1 - \beta$ we have $\sigma_{\mathcal{K}}(-\mathcal{A}_n \alpha_N + b) \leq h(\alpha_N, \varepsilon)$, which in conjunction with the dual

[2] The uniqueness assumption may be relaxed at the expense of solving an auxiliary convex program; see [35, section 3.3].

optimizer bound in Proposition 3.2 results in (4.4a). Now using the TB candidate in Example 4.3 immediately leads to the first assertion of (4.4c). Recall that the solution $\mathcal{P}_{n,N}$ obeys the norm bound $\|\alpha_N^\star\|_{\mathfrak{R}} \leq \theta_{\mathcal{P}}$. Thus, by employing the triangle inequality together with Lemma 4.1 we arrive at the second assertion (4.4c). ∎

Theorem 4.4 quantifies the approximation error between programs $\mathcal{P}_n$ and $\mathcal{P}_{n,N}$ probabilistically in terms of the TB functions as introduced in Definition 4.2. The natural question is under what conditions the proposed bound can be made arbitrarily small. This question is intimately related to the behavior of TB functions. For the TB candidate proposed in Example 4.3, the question translates to when the measure of a ball $\mathsf{B}_\gamma(\kappa) \subset \mathcal{K}$ has a lower bound $g(\gamma)$ uniformly away from 0 with respect to the location of its center: the answer to this question also depends on the properties of the norm on $(\mathbb{B}, \mathbb{Y}, \|\cdot\|)$. A positive answer to this question requires that the set $\mathcal{K}$ can be covered by finitely many balls, indicating that $\mathcal{K}$ is indeed compact with respect to the (dual) norm topology. In the next subsection we study this requirement in more detail in the MDP setting.

4.2. Randomized results in the MDP setting. We return to the MDP setting and discuss the implication of Theorem 4.4 as the bridge from the semi-infinite program $\mathcal{P}_n$ to the finite counterpart $\mathcal{P}_{n,N}$. Recall the dual pairs of vector spaces setting in (3.7) with the assigned norms (3.8). To construct the finite program $\mathcal{P}_{n,N}$, we need to sample from the set of extreme points of $\mathcal{P}(K)$, i.e., the set of point measures

$$\mathcal{K} := \mathcal{E}\big(\mathcal{P}(K)\big) = \big\{\boldsymbol{\delta}_{(s,a)} : (s,a) \in K\big\},$$

where $\boldsymbol{\delta}_{(s,a)}$ denotes a point probability distribution at $(s,a) \in K$. In this view, in order to sample elements from $\mathcal{K}$ it suffices to sample from the state-action feasible pairs $(s,a) \in K$.

COROLLARY 4.12 (MDP finite randomized approximation error). *Let $\{(s_j,a_j)\}_{j\leq N}$ be N i.i.d. samples generated from the uniform distribution on K. Consider the program*

$$-J_{n,N}^{\mathrm{AC}} = \begin{cases} \inf\limits_{(\rho,\alpha)\in\mathbb{R}^{n+1}} & -\rho \\ \text{s.t.} & \rho + \sum\limits_{i=1}^{n} \alpha_i\big(u_i(s_j) - Qu_i(s_j,a_j)\big) \leq \psi(s_j,a_j)\ \forall j \leq N, \\ & \|\alpha\|_{\mathfrak{R}} \leq \theta_{\mathcal{P}}. \end{cases} \tag{4.5}$$

where the basis functions $\{u_i\}_{i\leq n}$ introduced in (3.9) are normalized (i.e., $\|u_i\|_{\mathrm{L}} = 1$). Let L_Q be the Lipschitz constant from Assumption 2.1(ii), and define the constant $z_n := \big(\theta_{\mathcal{P}}\varrho_n(\max\{L_Q,1\}+1) + \|\psi\|_{\mathrm{L}}\big)^{-1}$, where ϱ_n is the ratio constant introduced in (4.1). Then, $\forall\, \varepsilon,\beta$ in $(0,1)$ and $N \geq \mathsf{N}\big(n+1, (z_n\varepsilon)^{\dim(K)}, \beta\big)$ defined in (4.3), we have $\mathbb{P}^N[0 \leq J_{n,N}^{\mathrm{AC}} - J_n^{\mathrm{AC}} \leq \varepsilon] \geq 1-\beta$.

Proof. Let $(\rho_N^\star, \alpha_N^\star)$ be the optimal solution for (4.5). Observe that in the MDP setting, Assumption 2.1(ii) implies

$$\begin{aligned} \|\mathcal{A}_n\alpha_N^\star - b\| &= \Big\| -\rho_N^\star + \sum_{i=1}^{n} \alpha_{N(i)}^\star (Q-I)u_i + \psi \Big\|_{\mathrm{L}} \\ &\leq (\max\{L_Q,1\}+1)\Big\| \sum_{i=1}^{n} \alpha_{N(i)}^\star u_i \Big\|_{\mathrm{L}} + \| -\rho_N^\star + \psi\|_{\mathrm{L}} \\ &\leq (\max\{L_Q,1\}+1)\theta_{\mathcal{P}}\varrho_n \left(\max_{i\leq n} \|u_i\|_{\mathrm{L}}\right) + \|\psi\|_{\mathrm{L}}, \end{aligned} \tag{4.6}$$

where the equality $\| - \rho_N^\star + \psi\|_{\mathrm{L}} = \|\psi\|_{\mathrm{L}}$ leading to (4.6) follows from the fact that ψ and $\rho^\star$ are nonnegative (note that $\alpha = 0, \rho = 0$ is a trivial feasible solution for (4.5)). In the second step, we propose a TB candidate in the sense of Definition 4.2. Note that for any $k, k' \in K$, by the definition of the Wasserstein norm we have $\|\boldsymbol{\delta}_{\{k\}} - \boldsymbol{\delta}_{\{k'\}}\|_{\mathrm{W}} = \min\{1, \|k - k'\|_\infty\}$. Thus, generating samples uniformly from K leads to

$$\mathbb{P}\big[\mathsf{B}_\gamma(\kappa)\big] \geq \mathbb{P}\big[\mathsf{B}_\gamma(k)\big] \geq \gamma^{\dim(K)} \qquad \forall \kappa \in \mathcal{K}, \quad \forall k \in K, \tag{4.7}$$

where, with slight abuse of notation, the first ball $\mathsf{B}_\gamma(\kappa)$ is a subset of the infinite dimensional space $\mathbb{Y}$ with respect to the dual norm $\|\cdot\|_{\mathrm{W}}$, while the second ball $\mathsf{B}_\gamma(k)$ is a subset of the finite dimensional space K whose respective norm is $\|\cdot\|_\infty$. The relation (4.7) readily suggests a function $g : \mathbb{R}_+ \to [0, 1]$ for Example 4.3, which together with (4.6) and the fact that the basis functions are normalized yields

$$h(\alpha, \varepsilon) := \|\mathcal{A}_n\alpha - b\| g^{-1}(\varepsilon) \leq \big(\theta_{\mathcal{P}} \varrho_n(\max\{L_Q, 1\} + 1) + \|\psi\|_{\mathrm{L}}\big)\varepsilon^{1/\dim K}.$$

Recall from Lemma 3.7 that the dual multiplier bound is $\theta_{\mathcal{D}} = 1$, and feasible solution α is bounded by $\theta_{\mathcal{P}}$. Finally, note that the decision variable of the program (4.5) is the $n + 1$ dimensional pair (ρ, α). Given all the information above, the claim then readily follows from the second result of Theorem 4.4 in (4.4c). ∎

To select $\theta_{\mathcal{P}}$, one may minimize the complexity of the a priori bound in Corollary 4.12, which is reflected through the required number of samples. At the same time, the impact of the bound $\theta_{\mathcal{P}}$ on the approximation step from infinite to semi-infinite in Corollary 3.9 should also be taken into account. The first factor is monotonically decreasing with respect to $\theta_{\mathcal{P}}$, i.e., the smaller the parameter $\theta_{\mathcal{P}}$, the lower the number of the required samples. The second factor is presented through the projection residual (cf. Remark 3.10). Therefore, an acceptable choice of $\theta_{\mathcal{P}}$ is an upper bound for the projection error of the optimal solution onto the ball $\mathbb{U}_n$ uniformly in $n \in \mathbb{N}$, i.e.,

$$\theta_{\mathcal{P}} \geq \sup\left\{ \|\alpha^\star\|_{\mathfrak{R}} \ : \ \Pi_{\mathbb{U}_n}(x^\star) = \sum_{i=1}^n \alpha_i^\star u_i, \quad n \in \mathbb{N} \right\}. \tag{4.8a}$$

The above bound may be available in particular cases, e.g., when $\|\cdot\|_{\mathfrak{R}} = \|\cdot\|_{\ell_2}$ it yields the bound

$$\|\alpha^\star\|_{\ell_2} = \sqrt{\int_S u^{\star 2}(s)\mathrm{d}s} \leq \|u^\star\|_{\mathrm{L}} \leq \max\{L_Q, 1\}\|\psi\|_\infty, \tag{4.8b}$$

where L_Q is the Lipschitz constant in Assumption 2.1(ii). We note that the first inequality in (4.8b) follows since S is a unit hypercube, and the second inequality follows from [19, Lemma 2.3]; see also [19, section 5] for further detailed analysis.

5. Semi-infinite to finite program: Structural convex optimization. This section approaches the approximation of the semi-infinite program $\mathcal{P}_n$ from an alternative perspective relying on an iterative first-order descent method. As opposed to the scenario approach presented in section 4, which is probabilistic and starts from the program $\mathcal{P}_n$, the method of this section is deterministic and starts with the dual counterpart $\mathcal{D}_n$, in particular a *regularized* version whose solutions can be computed efficiently. It turns out that the regularized solution allows one to reconstruct a nearly feasible solution for both programs $\mathcal{P}_n$ and $\mathcal{D}_n$, offering a meaningful performance bound for the approximation step from the semi-infinite program to a finite program.

5.1. Structural convex optimization. The basis of our approach is the fast gradient method that significantly improves the theoretical and, in many cases, also the practical convergence speed of the gradient method. The main idea is based on a well-known technique of smoothing nonsmooth functions [38]. To simplify the notation, for a given $\theta_{\mathcal{P}}$ we define the sets

$$\mathscr{A} := \big\{\alpha \in \mathbb{R}^n : \|\alpha\|_{\mathfrak{R}} \leq \theta_{\mathcal{P}}\big\}, \qquad \mathscr{Y} := \Big\{y \in \mathbb{K}^* : \|y\|_* \leq \theta_{\mathcal{D}}\Big\},$$

where $\theta_{\mathcal{D}}$ is the constant defined in (3.3). Recall that in the wake of Proposition 3.2 we know that the decision variables of the dual program $\mathcal{D}_n$ may be restricted to the set $\mathscr{Y}$ without loss of generality. We modify the program $\mathcal{D}_n$ with a regularization term scaled with the nonnegative parameter η and define the *regularized* program

$$(\mathcal{D}_{n,\eta}) \qquad \widetilde{J}_{n,\eta} := \sup_{y\in\mathscr{Y}} \Big\{\big\langle b, y\big\rangle - \theta_{\mathcal{P}}\|\mathcal{A}_n^* y - \mathbf{c}\|_{\mathfrak{R}^*} - \eta d(y)\Big\},$$

where the regularization function $d : \mathscr{Y} \to \mathbb{R}_+$, also known as the *prox-function*, is strongly convex. The choice of the prox-function depends on the specific problem structure and may have significant impact on the approximation errors. Given the regularization term η and the parameter $\alpha \in \mathbb{R}^n$, we introduce the auxiliary quantity

$$y_\eta^\star(\alpha) := \arg\max_{y\in\mathscr{Y}} \Big\{\big\langle b - \mathcal{A}_n\alpha, y\big\rangle - \eta d(y)\Big\}. \tag{5.1}$$

It is computationally crucial for the solution method proposed in this part that the prox-function allows us to have access to the auxiliary variable $y_\eta^\star(\alpha)$ for each $\alpha \in \mathbb{R}^n$. This requirement is formalized as follows.

Assumption 5.1 (Lipschitz gradient). Consider the adjoint operator $\mathcal{A}_n^*$ in (3.2) and the optimizer $y_\eta^\star(\alpha)$ of the auxiliary quantity (5.1). We assume that for each $\alpha \in \mathscr{A}$ the vector $\mathcal{A}_n^* y_\eta^\star(\alpha) \in \mathbb{R}^n$ can be approximated to an arbitrary precision, and the mapping $\alpha \mapsto \mathcal{A}_n^* y_\eta^\star(\alpha)$ is Lipschitz continuous with a constant $\frac{L}{\eta}$, i.e.,

$$\|\mathcal{A}_n^* y_\eta^\star(\alpha) - \mathcal{A}_n^* y_\eta^\star(\alpha')\|_{\mathfrak{R}^*} \leq \frac{L}{\eta}\|\alpha - \alpha'\|_{\mathfrak{R}} \qquad \forall \alpha, \alpha' \in \mathscr{A}.$$

Let $\vartheta > 0$ be the strong convexity parameter of the mapping $\alpha \mapsto \frac{1}{2}\|\alpha\|_{\mathfrak{R}}^2$ with respect to the $\mathfrak{R}$-norm. We then define the operator $\mathbb{T} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ as

$$\mathbb{T}(q, \alpha) := \arg\min_{\beta\in\mathscr{A}} \Big\{q \cdot \beta + \frac{1}{2\vartheta}\|\beta - \alpha\|_{\mathfrak{R}}^2\Big\}. \tag{5.2}$$

More generally, a different norm can be used in the second term in (5.2) when ϑ is a different strong convexity parameter. However, we forgo this additional generality to keep the exposition simple. The operator $\mathbb{T}$ is defined implicitly through a finite convex optimization program whose computational complexity may depend on the $\mathfrak{R}$-norm through the constraint set $\mathscr{A}$. For typical norms in $\mathbb{R}^n$ (e.g., $\|\cdot\|_{\ell_p}$) the pointwise evaluation of the operator $\mathbb{T}$ is computationally tractable. Furthermore, if $\|\cdot\|_{\mathfrak{R}} = \|\cdot\|_{\ell_2}$, then the definition of (5.2) has an explicit analytical description for any pair (q, α) as follows.

LEMMA 5.2 (explicit description of $\mathbb{T}$). *Suppose in the definition of the operator* (5.2) *the $\mathfrak{R}$-norm is the classical ℓ_2-norm. Then, the operator $\mathbb{T}$ admits the analytical description $\mathbb{T}(q, \alpha) = \xi\,(\alpha - q)$, where $\xi := \min\big\{1, \theta_{\mathcal{P}}\|q - \alpha\|_{\ell_2}^{-1}\big\}$.*

Algorithm 1. Optimal scheme for smooth convex optimization.

Choose some $w^{(0)} \in \mathscr{A}$
For $k \geq 0$ do
1. Define $r^{(k)} := \frac{\eta}{L}\big(\mathbf{c} - \mathcal{A}_n^* y_\eta^\star(w^{(k)})\big)$;
2. Compute $z^{(k)} := \mathbb{T}\big(\sum_{j=0}^{k} \frac{j+1}{2} r^{(j)}, 0\big), \quad \alpha^{(k)} := \mathbb{T}\big(\frac{1}{\vartheta} r^{(k)}, w^{(k)}\big)$;
3. Set $w^{(k+1)} = \frac{2}{k+3} z^{(k)} + \frac{k+1}{k+3} \alpha^{(k)}$.

Proof. See [34, Lemma 5.2]. ∎

Algorithm 1 exploits the information revealed under Assumption 5.1 as well as the operator $\mathbb{T}$ to approximate the solution of the program $\mathcal{D}_n$. The following proposition provides explicit error bounds for the solution provided by Algorithm 1 after k iterations. The result is a slight extension of the classical smoothing technique in finite dimensional convex optimization [38, Theorem 3] where the prox-function is not necessarily uniformly bounded, a potential difficulty in infinite dimensional spaces. We address this difficulty by considering a growth rate for the prox-function d evaluated at the optimal solution $y_\eta^\star$. We later show how this extension will help in the MDP setting.

THEOREM 5.3 (smoothing approximation error). *Suppose Assumption* 5.1 *holds with constant L and ϑ is the strong convexity parameter in the definition of the operator $\mathbb{T}$ in* (5.2). *Given the regularization term $\eta > 0$ and k iterations of Algorithm* 1, *we define*

$$\widehat{\alpha}_\eta := \alpha^{(k)}, \qquad \widehat{y}_\eta := \sum_{j=0}^{k} \frac{2(j+1)}{(k+1)(k+2)} y_\eta^\star(w^{(j)}).$$

Under Assumption 3.1, *the optimal value of the program $\mathcal{P}_n$ is bounded by $J_{n,\eta}^{\mathrm{LB}} \leq J_n \leq J_{n,\eta}^{\mathrm{UB}}$, where*

$$J_{n,\eta}^{\mathrm{LB}} := \big\langle b, \widehat{y}_\eta \big\rangle - \theta_{\mathcal{P}} \|\mathcal{A}_n^* \widehat{y}_\eta - \mathbf{c}\|_{\mathfrak{R}^*}, \qquad J_{n,\eta}^{\mathrm{UB}} := \widehat{\alpha}_\eta \bullet \mathbf{c} + \sup_{y \in \mathscr{Y}} \big\langle b - \mathcal{A}_n \widehat{\alpha}_\eta, y \big\rangle. \tag{5.3}$$

Moreover, suppose there exist positive constants c, C such that

$$C \max\big\{ \log\big(c\eta^{-1}\big), 1 \big\} \geq d\big(y_\eta^\star(\alpha)\big) \qquad \forall \eta > 0 \quad \forall \alpha \in \mathscr{A},$$

and, given an a priori precision $\varepsilon > 0$, the regularization parameter η and the number of iterations k satisfy

$$\eta \leq \frac{\varepsilon}{2C \max\{2\log(2cC\varepsilon^{-1}), 1\}}, \qquad k \geq 2\theta_{\mathcal{P}} \varrho_n \frac{\sqrt{CL \max\{2\log(2cC\varepsilon^{-1}), 1\}}}{\sqrt{\vartheta}\ \varepsilon}, \tag{5.4}$$

where ϱ_n is the constant defined in (4.1). *Then, after k iterations of Algorithm* 1 *we have $J_{n,\eta}^{\mathrm{UB}} - J_{n,\eta}^{\mathrm{LB}} \leq \varepsilon$.*

Proof. Observe that the bounds $J_{n,\eta}^{\mathrm{LB}}$ and $J_{n,\eta}^{\mathrm{UB}}$ in (5.3) are the values of the programs $\mathcal{D}_n$ and $\mathcal{P}_n$ evaluated at $\widehat{y}_\eta$ and $\widehat{\alpha}_\eta$, respectively. As such, the first assertion follows immediately. Toward the second part, thanks to the compactness of the set $\mathscr{A}$, the strong duality argument of Sion's minimax theorem [45] allows us to describe the program $\mathcal{D}_{n,\eta}$ through

$$\begin{aligned}\widetilde{J}_{n,\eta} &:= \sup_{y\in\mathscr{Y}} \big\langle b, y\big\rangle - \Big[\sup_{\alpha\in\mathscr{A}} \big\langle \mathcal{A}_n\alpha, y\big\rangle - \alpha\cdot\mathbf{c} + \eta d(y)\Big] \\ &= \inf_{\alpha\in\mathscr{A}} \alpha\cdot\mathbf{c} + \sup_{y\in\mathscr{Y}} \big[\big\langle b - \mathcal{A}_n\alpha, y\big\rangle - \eta d(y)\big] \\ &= \inf_{\alpha\in\mathscr{A}} \alpha\cdot\mathbf{c} + \big\langle b - \mathcal{A}_n\alpha, y_\eta^\star(\alpha)\big\rangle - \eta d\big(y_\eta^\star(\alpha)\big), \end{aligned} \tag{5.5}$$

where the last equality follows from the definition in (5.1). Note that the problem (5.5) belongs to the class of smooth and strongly convex optimization problems and can be solved using a fast gradient method developed by [38]. For this purpose, we define the function

$$\phi_\eta(\alpha) := \alpha\cdot\mathbf{c} + \big\langle b - \mathcal{A}_n\alpha, y_\eta^\star(\alpha)\big\rangle - \eta d\big(y_\eta^\star(\alpha)\big). \tag{5.6}$$

Invoking techniques similar to [38, Theorem 1], it can be shown that the mapping $\alpha \mapsto \phi_\eta(\alpha)$ is smooth with the gradient $\nabla\phi_\eta(\alpha) = \mathbf{c} - \mathcal{A}_n^* y_\eta^\star(\alpha)$. The gradient $\nabla\phi_\eta(\alpha)$ is Lipschitz continuous by Assumption 5.1 with constant $\frac{L}{\eta}$. Thus, following similar arguments as in the proof of [38, Theorem 3] we have

$$\begin{aligned} 0 \le J_{n,\eta}^{\mathrm{UB}} - J_{n,\eta}^{\mathrm{LB}} &\le \frac{L\|\alpha^\star\|_{\mathfrak{R}}^2}{\vartheta(k+1)(k+2)\eta} + \eta d\big(y_\eta^\star(\alpha^*)\big) \\ &\le \frac{L(\theta_{\mathcal{P}}\varrho_n)^2}{\vartheta k^2\eta} + C\eta\max\big\{\log(c\eta^{-1}), 1\big\}. \end{aligned} \tag{5.7}$$

Now, it is enough to bound each of the terms in the right-hand side of the above inequality by $\frac{1}{2}\varepsilon$. It should be noted that this may not lead to an optimal choice of the parameter η, but it is good enough to achieve a reasonable precision order with respect to ε. To ensure $\eta\log(\eta^{-1}) \le \varepsilon$ for an $\varepsilon \in (0,1)$, it is not difficult to see that it suffices to set $\eta \le \frac{\varepsilon}{2\log(\varepsilon^{-1})}$. In this observation if we replace η and ε with $\frac{1}{c}\eta$ and $\frac{1}{2cC}\varepsilon$, respectively, we deduce that the second term on the right-hand side in (5.7) is bounded by $\frac{1}{2}\varepsilon$. Thus, the desired assertion follows by equating the first term on the right-hand side in (5.7) to $\frac{1}{2}\varepsilon$ while the parameter η is set as just suggested. ∎

Remark 5.4 (computational complexity). Adding the prox-function to the problem $\mathcal{D}_n$ ensures that the regularized counterpart $\mathcal{D}_{n,\eta}$ admits an efficiency estimate (in terms of iteration numbers) of the order $\mathcal{O}(\sqrt{\frac{L}{\eta}\varepsilon^{-1}})$. To construct a smooth ε-approximation for the original problem $\mathcal{D}_n$, the Lipschitz constant $\frac{L}{\eta}$ can be chosen of the order $\mathcal{O}(\varepsilon^{-1}\log(\varepsilon^{-1}))$. Thus, the presented gradient scheme has an efficiency estimate of the order $\mathcal{O}\big(\varepsilon^{-1}\sqrt{\log(\varepsilon^{-1})}\big)$; see [38] for a more detailed discussion along similar objective.

Remark 5.5 (inexact gradient). The error bounds in Theorem 5.3 are introduced based on the availability of the exact first-order information, i.e., it is assumed that at each iteration the vector $r^{(k)}$ that due to the bilinear form potentially involves a multidimensional integration can be computed exactly. In general, the evaluation of those vectors may only be available approximately. This gives rise to the question of how the fast gradient method performs in the case of *inexact* first-order information. We refer the interested reader to [15] for further details.

The a priori bound proposed by Theorem 5.3 involves the positive constants c, C, which are used to introduce an upper bound for the proxy-term. These constants potentially depend on $\theta_{\mathcal{D}}$, the size of the dual feasible set, hence also on $\theta_{\mathcal{P}}$. Therefore,

unlike the randomized approach in section 4, it is not immediately clear how $\theta_{\mathcal{P}}$ can be chosen to minimize the complexity of the proposed method, which in this case is the required number of iterations k suggested in (5.4) (cf. Remark 4.6). In the next section, we shall discuss how to address this issue in the MDP setting for particular constants c, C.

5.2. Structural convex optimization results in the MDP setting. To link the approximation method presented in section 5.1 to the AC program in (3.9), let us recall the dual pairs (3.7) equipped with the norms (3.8). To simplify the analysis, we refine the assertion in Lemma 3.7 and argue that the dual optimizers are indeed probability measures, i.e.,

$$\mathscr{Y} := \Big\{ y \in \mathcal{M}_+(K) \ : \|y\|_{\mathrm{W}} = \theta_{\mathcal{D}} = 1 \Big\}. \tag{5.8}$$

To see this, one can consider the norm $\|(\rho,\alpha)\| := \|\alpha\|_{\mathfrak{R}}$ and follow similar arguments in the proof of Proposition 3.2. Strictly speaking, this is not a true norm on $\mathbb{R}^{n+1}$ but it does not affect the technical argument, in particular strong duality between $\mathcal{P}_n$ and $\mathcal{D}_n$. The details are omitted here in the interest of space. We consider the prox-function as a relative entropy defined by

$$d(y) := \begin{cases} \big\langle \log\big(\frac{\mathrm{d}y}{\mathrm{d}\lambda}\big), y \big\rangle, & y \ll \lambda, \\ \infty & \text{otherwise}, \end{cases} \tag{5.9}$$

where λ is the uniform measure supported on the set K and $\frac{\mathrm{d}y}{\mathrm{d}\lambda} \in \mathcal{F}_+(K)$ is the Radon–Nikodym derivative between two measures y and λ. One can inspect that the prox-function (5.9) is indeed a nonnegative function. The optimizer of the regularized program $\mathcal{D}_{n,\eta}$ for the AC program (3.9) is

$$y^\star_\eta(\rho,\alpha) := \arg\max_{y\in\mathscr{Y}} \left\{ \left\langle -\psi + \rho - \sum_{i=1}^n \alpha_i (Q-I) u_i y \right\rangle - \eta \big\langle \log\big(\tfrac{\mathrm{d}y}{\mathrm{d}\lambda}\big), y \big\rangle \right\}. \tag{5.10}$$

To see (5.10), check (5.1) together with the definitions of the operator $\mathcal{A}_n$ in (3.2) and the AC problem parameters in (2.3). The main reason for such a choice of the regularization term is the fact that the optimizer of the regularized program (5.10) admits an analytical expression.

Lemma 5.6 (entropy maximization [12]). *Given a (measurable) function $g : K \to \mathbb{R}$ and the set $\mathscr{Y} \subset \mathcal{M}_+(K)$ as defined in* (5.8) *we have*

$$y^\star(\mathrm{d}k) := \arg\max_{y\in\mathscr{Y}} \Big\{ \big\langle g, y \big\rangle - \eta d(y) \Big\} = \frac{\exp\big(\eta^{-1} g(k)\big)\lambda(\mathrm{d}k)}{\big\langle \exp\big(\eta^{-1} g(k)\big), \lambda \big\rangle}.$$

Thanks to Lemma 5.6, the analytical description of the dual optimizer in (5.10) is readily available by setting

$$g(k) := [b - \mathcal{A}_n \alpha](k) = -\psi(k) + \rho - \sum_{i=1}^n \alpha_i (Q-I) u_i(k). \tag{5.11}$$

The last requirement to implement Algorithm 1 is to verify Assumption 5.1, i.e., we need to compute the Lipschitz constant of the mapping $(\rho,\alpha) \mapsto \mathcal{A}_n^* y^\star_\eta(\rho,\alpha)$ in which the respective norm is $\|(\rho,\alpha)\| := \|\alpha\|_{\mathfrak{R}}$. By definition of the adjoint operator $\mathcal{A}_n^*$ in (3.2), it is not difficult to observe that

$$\mathcal{A}_n^* y_\eta^\star(\rho,\alpha) = \begin{bmatrix} \big\langle -\mathbb{1}, y_\eta^\star(\rho,\alpha)\big\rangle \\ \big\langle (Q-I)u_1, y_\eta^\star(\rho,\alpha)\big\rangle \\ \vdots \\ \big\langle (Q-I)u_n, y_\eta^\star(\rho,\alpha)\big\rangle \end{bmatrix} = \begin{bmatrix} -1 \\ \big\langle (Q-I)u_1, y_\eta^\star(\rho,\alpha)\big\rangle \\ \vdots \\ \big\langle (Q-I)u_n, y_\eta^\star(\rho,\alpha)\big\rangle \end{bmatrix}. \tag{5.12}$$

The next lemma addresses the requirement of Assumption 5.1 for the mapping (5.12).

LEMMA 5.7 (Lipschitz constant in MDP). *Consider the entropy maximizers in Lemma* 5.6 *with* g *as defined in* (5.11) *and the adjoint operator in* (5.12). *An upper bound for the Lipschitz constant in Assumption* 5.1 *is* $L \leq 4\varrho_n^2$, *where the constant* ϱ_n *is the equivalence ratio between the norms* $\|\cdot\|_{\mathfrak{R}}$ *and* $\|\cdot\|_{\ell_1}$ *introduced in* (4.1).

Proof. See [34, Lemma 5.7]. ∎

The performance of Algorithm 1 can now be characterized through the following corollary.

COROLLARY 5.8 (MDP smoothing approximation error). *Consider the operator* (3.2) *with the parameters described in* (2.3) *for the semi-infinite AC program* (3.9). *Given this setting and the Lipschitz constant in Lemma* 5.7, *we run Algorithm* 1 *for* k *iterations using the entropy function* (5.9) *with analytical solution* (5.10) *as the prox-function. We define the constants*

$$C_1 := 2\,\mathrm{e}\,\big(\varrho_n\theta_{\mathcal{P}}(\max\{L_Q,1\}+1) + \|\psi\|_{\mathrm{L}}\big), \qquad C_2 := 4\theta_{\mathcal{P}}\varrho_n^2\sqrt{\frac{2\dim(K)}{\vartheta}}.$$

For every $\varepsilon \leq C_1$ *we set the smoothing factor* η *and the number of iterations* k *by*

$$\eta \leq \frac{\varepsilon}{4\dim(K)\log(C_1\varepsilon^{-1})}, \qquad k \geq C_2\frac{\sqrt{\log(C_1\varepsilon^{-1})}}{\varepsilon}.$$

Then, the outcome of Algorithm 1 *as defined in* (5.3) *is an* ε *approximation of the optimal value* J_n^{AC} *in the sense of Theorem* 5.3.

Corollary 5.8 requires one to compute the constants c, C to quantify the a priori bounds. The following two technical lemmas provide supplementary materials to address this issue.

LEMMA 5.9. *Let* $K \subseteq [0,1]^m$ *and* $g : K \to \mathbb{R}$ *be a Lipschitz continuous function with constant* $L_g > 0$ *(with respect to the* ℓ_∞*-norm) and the maximum value* $g_{\max} := \max_{k\in K} g(k)$. *Then, for every* $\eta > 0$ *we have*

$$\int_K \exp\Big(\eta^{-1}\big(g(k)-g_{\max}\big)\Big)\mathrm{d}k \geq \min\Big\{\Big(\frac{m\eta}{L_g}\Big)^m, 1\Big\}\exp\big(-\min\big\{m, L_g\eta^{-1}\big\}\big).$$

Proof. Let us define the set $Z(\delta) := \{k \in K : g_{\max} - g(k) < \delta\}$. Thanks to the Lipschitz continuity of the function g, we have $g_{\max} - g(k) \leq L_g\|k^\star - k\|_{\ell_\infty}$, where $g(k^\star) = g_{\max}$. Thus, using this inequality one can bound the size of the set $Z(\delta)$ in the sense of

$$\int_{Z(\delta)} \mathrm{d}k \geq \min\{(\delta L_g^{-1})^m, 1\} \qquad \forall \delta \geq 0.$$

By virtue of the above result, one can observe that for every $\delta > 0$

$$\begin{aligned}\int_K \exp\Big(\eta^{-1}\big(g(k)-g_{\max}\big)\Big)\mathrm{d}k &\geq \int_{Z(\delta)} \exp\Big(\eta^{-1}\big(g(k)-g_{\max}\big)\Big)\mathrm{d}k \\ &\geq \exp(-\eta^{-1}\delta)\int_{Z(\delta)}\mathrm{d}k \geq \exp(-\eta^{-1}\delta)\min\{(\delta L_g^{-1})^m, 1\}.\end{aligned}$$

Maximizing the right-hand side of the above inequality over δ suggests to set $\delta = \min\{m\eta, L_g\}$, which yields the desired assertion. ∎

In light of Lemma 5.9, we can bound the entropy prox-function (5.9) evaluated at the optimizer (5.10).

LEMMA 5.10 (entropy prox-bound). *Consider the prox-function* (5.9) *and let* $y^\star_\eta(\rho,\alpha)$ *be the optimizer of* (5.10). *Then, for every* $\eta > 0$, ρ, *and* $\|\alpha\|_{\mathfrak{R}} \le \theta_{\mathcal{P}}$, *we have* $d\big(y^\star_\eta(\rho,\alpha)\big) \le C \max\big\{\log(c\eta^{-1}), 1\big\}$, *where*

$$C := \dim(K), \qquad c := \frac{\mathrm{e}}{\dim(K)}\big(\theta_{\mathcal{P}}\varrho_n(\max\{L_Q,1\}+1) + \|\psi\|_{\mathrm{L}}\big),$$

and ϱ_n *is the equivalence ratio between the norms* $\|\cdot\|_{\ell_1}$ *and* $\|\cdot\|_{\mathfrak{R}}$ *as defined in* (4.1).

Proof. The result is a direct application of Lemma 5.9. Consider the function g as defined in (5.11) with Lipschitz constant $L_g \ge 0$; note that the function g, as well as its Lipschitz constant L_g, depends also on the pair (ρ,α). Observe that

$$\begin{aligned}
d\big(y^\star_\eta(\rho,\alpha)\big) &= \big\langle \log\big(\exp(\eta^{-1}g)\big), y^\star_\eta(\rho,\alpha)\big\rangle - \log\big(\big\langle \exp(\eta^{-1}g), \lambda\big\rangle\big)\\
&= \big\langle \eta^{-1}g, y^\star_\eta(\rho,\alpha)\big\rangle - \log\big(\big\langle \exp(\eta^{-1}g), \lambda\big\rangle\big)\\
&= \big\langle \eta^{-1}g, y^\star_\eta(\rho,\alpha)\big\rangle - \eta^{-1}g_{\max} - \log\big(\big\langle \exp(\eta^{-1}(g-g_{\max}), \lambda\big\rangle\big)\\
&\le -\log\big(\big\langle \exp(\eta^{-1}(g-g_{\max})), \lambda\big\rangle\big)\\
&\le -\log\left(\min\left\{\left(\frac{\dim(K)\eta}{L_g}\right)^{\dim(K)}, 1\right\}\exp\big(-\min\big\{\dim(K), L_g\eta^{-1}\big\}\big)\right) \qquad (5.13)\\
&\le \dim(K)\max\left\{\log\left(\left(\frac{\mathrm{e}\,L_g}{\dim(K)}\right)\eta^{-1}\right), 1\right\},
\end{aligned}$$

where the inequality (5.13) follows from Lemma 5.9. Note also that the Lipschitz constant L_g for the function g defined in (5.11) is upper bounded, uniformly in (ρ,α) where $\|\alpha\|_{\mathfrak{R}} \le \theta_{\mathcal{P}}$, by

$$\begin{aligned}
L_g \le \|g-\rho\|_{\mathrm{L}} &\le \Big\|\sum_{i=1}^n \alpha_i(Q-I)u_i + \psi\Big\|_{\mathrm{L}} \le (\max\{L_Q,1\}+1)\Big\|\sum_{i=1}^n \alpha_i u_i\Big\|_{\mathrm{L}} + \|\psi\|_{\mathrm{L}}\\
&\le \theta_{\mathcal{P}}\varrho_n(\max\{L_Q,1\}+1) + \|\psi\|_{\mathrm{L}}.
\end{aligned}$$

We refer to the proof of Corollary 4.12, and in particular the paragraph following (4.6), for further discussions regarding L_g. The desired assertion follows from the last two inequalities and the definition of the constant $\theta_{\mathcal{D}}$ in (5.8). ∎

The proof of Corollary 5.8 follows by replacing the constants in Lemma 5.10 in Theorem 5.3. By contrast to the randomized approach in Corollary 4.12 where the computational complexity scales exponentially in dimensional of state-action space, the complexity of the smoothing technique grows effectively linearly (more precisely $\mathcal{O}\big(\varepsilon^{-1}\sqrt{\log(\varepsilon^{-1})}\big)$; cf. Remark 5.4). The computational difficulty is, however, transferred to step 1 of Algorithm 1 for computation of $\mathcal{A}_n^* y^\star_\eta$ as defined in (5.12). The following remark elaborates this.

Remark 5.11 (efficient computation of (5.12)). When the transition kernel Q and the basis functions u_i are such that the relation (5.12) involves integration of exponentials of polynomials over simple sets (e.g., a box or a simplex), one may utilize efficient methods that require solving a hierarchy of semidefinite programming problems to generate upper and lower bounds which asymptotically converge to the true

value of integral; see [31, section 12.2] and [9]. It is also worth noting that a straightforward computation of (5.12) for a small parameter η may be numerically difficult due to the exponential functions. This issue can, however, be circumvented by a numerically stable technique presented in [38, p. 148].

Regarding the choice of $\theta_{\mathcal{P}}$, in a similar spirit to section 4, one can target minimizing the complexity of the a priori bound, in other words, the number of iterations k in (5.4). In the setting of Corollary 5.8, one can observe that the smaller the parameter $\theta_{\mathcal{P}}$, the lower the number of the required iterations, leading to the choice described as in (4.8).

6. Full infinite to finite programs. The intention in this short section is to combine the two-step process from infinite to semi-infinite programs in section 3 and from semi-infinite to finite programs in sections 4 and 5 and hence establish a link from the original infinite program to finite counterparts. We only present the final result for the general infinite programs without discussing its implication in the MDP setting, as it is essentially a similar assertion.

THEOREM 6.1 (infinite to finite approximation error). *Consider the infinite program $\mathcal{P}$ with a solution $\{x^\star, J\}$, the finite (random) convex program $\mathcal{P}_{n,N}$ with the (random) solution $\{\alpha_N^\star, J_{n,N}\}$, and the output of Algorithm 1 with values $\{J_{n,\eta}^{\mathrm{LB}}, J_{n,\eta}^{\mathrm{UB}}\}$. Suppose Assumption 3.1 holds and assume further that there exists constant d, D so that the projection residual of the optimizer $x^\star$ onto the finite dimensional ball defined in Theorem 3.3 is bounded by $\|r_n\| \leq Dn^{-1/d}\ \forall n \in \mathbb{N}$. Then, for any number of scenario samples N and prox-term coefficient η, with probability $1-\beta$ we have*

$$\max\{J_{n,N}, J_{n,\eta}^{\mathrm{LB}}\} - D\big(\|c\|_* + \theta_{\mathcal{D}}\|\mathcal{A}\|\big)n^{-1/d} \leq J \leq \min\{J_{n,\eta}^{\mathrm{UB}}\ ,\ J_{n,N} + \theta_{\mathcal{D}} h(\alpha_N^\star, \varepsilon)\},$$

where $\theta_{\mathcal{D}}$ is as defined in (3.3) *and the function h is a TB in the sense of Definition* 4.2. *Moreover, given an a priori precision level ε, if $n \geq \big(D(\|c\|_* + \theta_{\mathcal{D}}\|\mathcal{A}\|)\varepsilon^{-1}\big)^d$, and the number samples N are chosen as in* (4.4b) or *the parameter η together with the number of iterations of Algorithm 1 is chosen as in* (5.4), *then with probability $1-\beta$ we have*

$$\min\Big\{|J - J_{n,N}|, |J - J_{n,\eta}^{\mathrm{LB}}|\Big\} \leq \varepsilon\,.$$

The proof follows readily from the link between the infinite program $\mathcal{P}$ to the semi-infinite counterpart $\mathcal{P}_n$ in Theorem 3.3, in conjunction with the link between $\mathcal{P}_n$ to the finite programs $\mathcal{P}_{n,N}$ and $\mathcal{D}_{n,\eta}$ in Theorems 4.4 and 5.3, respectively.

The assertion of Theorem 6.1 can be readily translated into the MDP problem by replacing the dual optimizer bound $\theta_{\mathcal{D}}$ with 1 thanks to Lemma 3.7 and the term $(\|c\|_* + \theta_{\mathcal{D}}\|\mathcal{A}\|)$ with $(1 + \max\{L_Q, 1\})$ thanks to Corollary 3.9. In this case, the requirement concerning the projection residual bound $\|r_n\| \leq Dn^{-1/d}$ is fulfilled due to the Lipschitz continuity of the value function when $d = \dim(S)$ and the finite dimensional approximation is generated by, among others, polynomials [22] or the Fourier basis [41] (cf. Remark 3.10).

7. Numerical examples. We present two numerical examples to illustrate the solution methods and corresponding performance bounds. Throughout this section we consider the norm $\|\cdot\|_{\mathfrak{R}} = \|\cdot\|_{\ell_2}$, leading to $\varrho_n = \sqrt{n}$ in (4.1), and we choose the Fourier basis functions.

7.1. Example 1: Truncated LQG. Consider the linear system

$$s_{t+1} = \vartheta s_t + \rho a_t + \xi_t, \quad t \in \mathbb{N},$$

with quadratic stage cost $\psi(s,a) = qs^2 + ra^2$, where $q \geq 0$ and $r > 0$ are given constants. We assume that $S = A = [-L, L]$ and the parameters $\vartheta, \rho \in \mathbb{R}$ are known. The disturbances $\{\xi_t\}_{t\in\mathbb{N}}$ are i.i.d. random variables generated by a truncated normal distribution with known parameters μ and σ, independent of the initial state s_0. Thus, the process ξ_t has a distribution density

$$f(s,\mu,\sigma,L) = \begin{cases} \frac{\frac{1}{\sigma}\phi\left(\frac{s-\mu}{\sigma}\right)}{\Phi\left(\frac{L-\mu}{\sigma}\right)-\Phi\left(\frac{-L-\mu}{\sigma}\right)}, & s \in [-L, L], \\ 0 & \text{otherwise}, \end{cases}$$

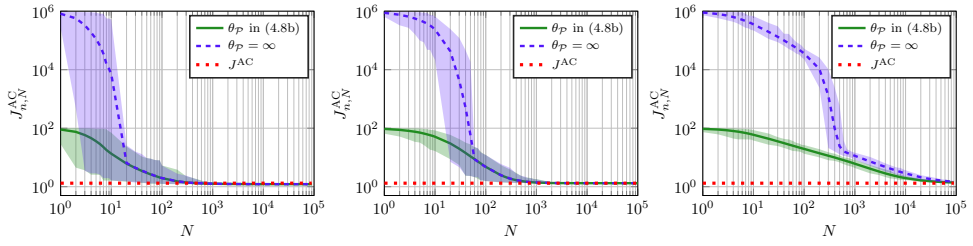
where ϕ is the probability density function of the standard normal distribution, and Φ is its cumulative distribution function. The transition kernel Q has a density function $q(y|s,a)$, i.e., $Q(B|s,a) = \int_B q(y|s,a)\mathrm{d}y \ \forall\, B \in \mathcal{B}(S)$, that is given by $q(y|s,a) = f(y - \vartheta s - \rho a, \mu, \sigma, L)$. In the special case that $L = +\infty$ the above problem represents the classical LQG problem, whose solution can be obtained via the algebraic Riccati equation [6, p. 372]. By a simple change of coordinates it can be seen that the presented system fulfills Assumption 2.1. The following lemma provides the technical parameters required for the proposed error bounds.

LEMMA 7.1 (truncated LQG properties). *The error bounds provided by Corollaries* 4.12 *and* 5.8 *hold with the norms* $\|\psi\|_\infty = L^2(q+r)$, $\|\psi\|_{\mathrm{L}} = 4L^2\sqrt{q^2+r^2}$, *and the Lipschitz constant of the kernel is* $L_Q = \frac{2L\max\{\vartheta,\rho\}}{\sigma^2\sqrt{2\pi}\left(\Phi\left(\frac{L-\mu}{\sigma}\right)-\Phi\left(\frac{-L-\mu}{\sigma}\right)\right)}$.

Proof. In regard to Assumption 2.1(i), we consider the change of coordinates $\bar{s}_t := \frac{s_t}{2L} + \frac{1}{2}$ and $\bar{a}_t := \frac{a_t}{2L} + \frac{1}{2}$. In the new coordinates, the constants of Lemma 7.1 follow from a standard computation. ∎

For the simulation results we choose the numerical values $\vartheta = 0.8$, $\rho = 0.5$, $\sigma = 1$, $\mu = 0$, $q = 1$, $r = 0.5$, and $L = 10$. In the first approximation step discussed in section 3.3, we consider the Fourier basis $u_{2k-1}(s) = \frac{L}{k\pi}\cos\left(\frac{k\pi s}{L}\right)$ and $u_{2k}(s) = \frac{L}{k\pi}\sin\left(\frac{k\pi s}{L}\right)$.

Randomized approach. We implement the methodology presented in section 4.2, resulting in a finite random convex program as in (4.5), where the uniform distribution on $K = S \times A = [-L, L]^2$ is used to draw the random samples. Figures 2(a), 2(b), and 2(c) visualize three cases with different number of basis functions



(a) $n = 2$ basis functions. (b) $n = 10$ basis functions. (c) $n = 100$ basis functions.

FIG. 2. *The quantity* $J^{\mathrm{AC}}_{n,N}$ *is computed using* (4.5). *The optimal value* J^{AC} *(red dotted line) is approximated by* $n = 10^3$ *and* $N = 10^6$.

$n \in \{2, 10, 100\}$, respectively. To show the impact of the additional norm constraint, in each case two approximation settings are examined: the constrained (regularized) one proposed in this article (i.e., $\theta_{\mathcal{P}} < \infty$) and the unconstrained one (i.e., $\theta_{\mathcal{P}} = \infty$). In the former we choose the bound suggested by (4.8b). In the latter, the resulting optimization programs of (4.5) may happen to be unbounded, particularly when the number of samples N is low; numerically, we capture the behavior of the unbounded $\theta_{\mathcal{P}}$ through a large bound such as $\theta = 10^6$. In each subfigure, the colored tubes represent the results of 400 independent experiments (shaded areas) as well as the mean value across different experiments (solid and dashed lines) of the objective performance $J^{\mathrm{AC}}_{n,N}$ as a function of the sample size N.

All the results in Figure 2 are obtained based on 400 independent simulation experiments. It is perhaps not surprising that the optimal value depicted by the red dotted line is very close to the classical LQG example whose exact solution is analytically available. It can be seen that the randomized approximations asymptotically converge, as suggested by Theorem 6.1.

The simulation results suggest three interesting features concerning n, the number of basis functions: The higher the number of basis functions,

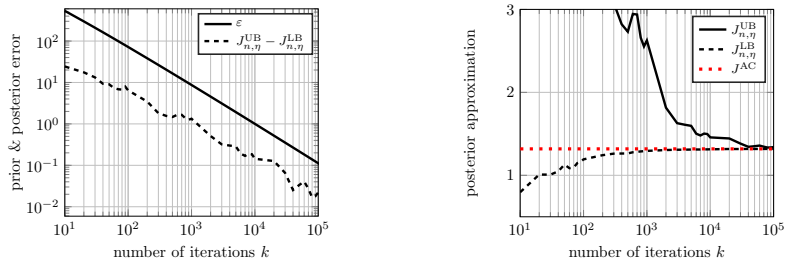
(i) the smaller the approximation error (i.e., asymptotic distance for $N \to \infty$ to the red dotted line),
(ii) the lower the variance of approximation with respect to the sampling distribution for each N, and
(iii) the slower the convergence behavior with respect to the sample size N.

Features (i) and (ii) are positive impacts of increasing the number of basis functions. While (i) is predicted by Corollary 3.9, since the error due to the projection term becomes smaller, it is not entirely clear how to formally explain (ii). On the contrary, feature (iii) is indeed a negative impact, as a high number of basis functions requires a large number of samples N to produce reasonable approximation errors. This phenomena can be justified through the lens of Corollary 4.12, where the approximation errors grows proportionally to n.

Structural convex optimization. Algorithm 1 is implemented with the parameters described in Corollary 5.8 leading to deterministic upper and lower bounds ($J^{\mathrm{UB}}_{n,\eta}$ and $J^{\mathrm{LB}}_{n,\eta}$, respectively) for the cost function J^{AC}_n; see also Theorem 5.3. These bounds are computationally appealing as they provide a posteriori bounds on the approximation error that often is significantly smaller than the a priori bounds given by Theorem 5.3. This behavior can be seen in the simulation results summarized in Figure 3, where the number of basis functions is $n = 10$. Similar to Figure 2, the red dotted line is the optimal value of the original infinite program $\mathcal{P}$, which we approximated by using 10^3 basis functions and 10^6 iterations of Algorithm 1; it coincides with the one from the randomized method.

7.2. Example 2: A fisheries management problem. A natural approximation approach toward dynamic programming problems goes through a discretization scheme (e.g., discretization of the state and/or action spaces). The main objective of this example is to compare the proposed LP-based approximation of this article with more standard discretization schemes. To this end, we borrow an example from [25, section 1.3] and compare our results with the recent discretization method proposed by [43]. Consider the population growth model, known as the Ricker model,

$$s_{t+1} = \vartheta_1 a_t \exp(-\vartheta_2 a_t + \xi_t), \quad t \in \mathbb{N},$$



(a) Prior ε and posterior $J^{\mathrm{UB}}_{n,\eta} - J^{\mathrm{LB}}_{n,\eta}$.

(b) Posterior error bounds.

FIG. 3. *The results and error bounds are obtained by Algorithm* 1 *with* $n = 10$. *The optimal value* J^{AC} *(red dotted line) is computed as indicated in Figure* 2.

where $\vartheta_1, \vartheta_2 \in \mathbb{R}_+$, s_t is the population size in season t, and a_t is the population to be left for spawning for the next season, i.e., the difference $s_t - a_t$ is the amount of fish captured in season t. The running reward function to be maximized is $\psi(a, s) = \varphi(s - a)$, where φ is the so-called shifted isoelastic utility function $\varphi(z) := 3(z + 0.5)^{1/3} - (0.5)^{1/3}$ [16], [13, section 4.1]. The state space is $S = [\underline{\kappa}, \overline{\kappa}]$ for some $\underline{\kappa}, \overline{\kappa} \in \mathbb{R}_+$. Since the population left for spawning cannot be greater than the total population, for each $s \in S$, the set of admissible actions is $A(s) = [\underline{\kappa}, s]$. To fulfill Assumption 2.1(i), following the transformation suggested by [43], we equivalently reformulate the above problem using the dynamics

$$s_{t+1} = \vartheta_1 \min(a_t, s_t) \exp(-\vartheta_2 \min(a_t, s_t) + \xi_t), \quad t \in \mathbb{N},$$

where the admissible actions set is now the state-independent set $A = [\underline{\kappa}, \overline{\kappa}]$, and the running reward function is $\psi(a, s) = \varphi(s - a)\mathbf{1}_{\{s \geq a\}}$. The noise process $(\xi_t)_{t \in \mathbb{N}}$ is a sequence of i.i.d. random variables which have a uniform density function g supported on the interval $[0, \lambda]$. Thus, the corresponding kernel is

$$Q(B|s, a) = \int_B g\Big(\log \xi - \log\big(\vartheta_1 \min(a, s)\big) + \vartheta_2 \min(a, s)\Big)\frac{1}{\xi}\mathrm{d}\xi \quad \forall B \in \mathcal{B}(\mathbb{R})\,.$$

Note that to make the model consistent, we must have $\vartheta_1 a \exp(-\vartheta_2 a + \xi) \in [\underline{\kappa}, \overline{\kappa}]\ \forall\, (a, \xi) \in [\underline{\kappa}, \overline{\kappa}] \times [0, \lambda]$. By defining an appropriate change of coordinate similar to Lemma 7.1, Assumption 2.1 is fulfilled; we refer the reader to [43, section 7.2] for further information and detailed analysis.

The chosen numerical values are $\lambda = 0.5$, $\vartheta_1 = 1.1$, $\vartheta_2 = 0.1$, $\overline{\kappa} = 7$, and $\underline{\kappa} = 0.005$.

Randomized approach. We implement the methodology presented in section 4.2, resulting in a finite random convex program (4.5), where the uniform distribution on $K = S \times A = [\underline{\kappa}, \overline{\kappa}]^2$ is used to draw the random samples. Figure 4 illustrates three cases with the number of basis functions $n \in \{2, 10, 100\}$ and the bound (4.8b). The colored tubes represent the results between $[10\%, 90\%]$ quantiles (shaded areas) as well as the means (solid lines) across 400 independent experiments of the objective performance $J^{\mathrm{AC}}_{n,N}$ as a function of the sample size N. It is interesting to note that in this example the optimal solution is captured even with two basis functions and only $N = 20$ random samples. This becomes even more attractive when we compare

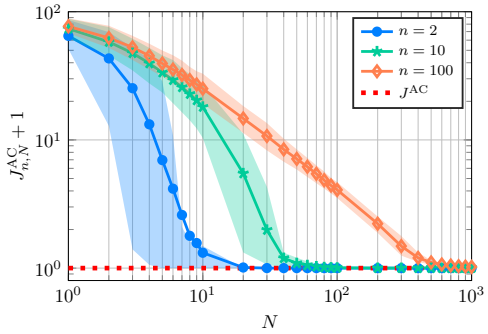


FIG. 4. *The quantity* $J^{\mathrm{AC}}_{n,N}$ *is computed using* (4.5). *The optimal value* J^{AC} *is approximated by* $n = 10^3$ *and* $N = 10^6$, *which amounts to* 0 *as also reported in* [43].

the results with a direct discretization scheme depicted in [43, Figure 2]. The numerical simulations concerning the structural convex optimization (Algorithm 1 and the bounds in Corollary 5.8) are reported in [34, Figure 5].

Appendix A. Infinite-horizon discounted-cost problems. In the MDP setting, introduced in section 2.1, let us consider *long-run* τ*-discounted cost* (DC) problems with the discount factor $\tau \in (0, 1)$ and initial distribution $\nu \in \mathcal{P}(X)$ described as $J^{\mathrm{DC}}(\nu) := \inf_{\pi\in\Pi} \limsup_{n\to\infty} \mathbb{E}^{\pi}_{\nu}[\sum_{t=0}^{n-1} \tau^t \psi(x_t, a_t)]$.

As in the AC setting, in section 2, we assume that the control model satisfies Assumption 2.1. We refer to [25, Chapter 4] and [27, Chapter 8] for a detailed exposition and required technical assumptions in more general settings. As for the AC problems, it is well known that the DC problem can be alternatively characterized by means of infinite LPs $(\mathcal{P})$ and $(\mathcal{D})$ introduced in section 3.1, where

$$\text{(A.1)} \quad \begin{cases} (\mathbb{X}, \mathbb{C}) = (\mathcal{C}(S), \mathcal{M}(S)), & c(B) = -\nu(B), \quad B \in \mathfrak{B}(S), \\ (\mathbb{B}, \mathbb{Y}) = (\mathcal{C}(K), \mathcal{M}(K)), & b(s,a) = -\psi(s,a), \\ \mathbb{K} = \mathcal{C}_+(K), & \mathcal{A} : \mathbb{X} \to \mathbb{B},\ \mathcal{A}x(s,a) = -x(s) + \tau Q x(s,a), \\ \mathbb{K}^* = \mathcal{M}_+(K), & \mathcal{A}^* : \mathbb{Y} \to \mathbb{C}, \mathcal{A}^* y(B) = y(B \times A) - \tau y Q(B). \end{cases}$$

THEOREM A.1 (LP characterization [25, Theorem 6.3.8]). *Under Assumption* 2.1, *the optimal value* J^{DC} *of the DC problem can be characterized by the LP problem* $(\mathcal{P})$ *in the setting* (A.1), *in the sense that* $J = -J^{\mathrm{DC}}$.

It is known that under similar conditions as in Assumption 2.1 on the control model, the value function $u^\star$ in the τ-DC optimality equation is Lipschitz continuous; see [23, section 2.6] or [17, Theorem 3.1]. We use the norms similar to the AC setting (3.7). The next step toward studying the approximation error (3.5) for the DC setting readily follows by Theorem 3.3 combined with the following lemma.

LEMMA A.2 (DC semi-infinite regularity). *For the DC problem, characterized by the dual-pair vector spaces in* (A.1), *under Assumption* 2.1 *we have the operator norm* $\|\mathcal{A}\| \le 1 + \max\{L_Q, 1\}\tau$, *the inf-sup constant of Assumption* 3.1(ii) $\gamma = 1 - \tau$, *and the dual optimizer norm*

$$\text{(A.2)} \qquad \|y^\star\|_{\mathrm{W}} \le \theta_{\mathcal{D}} = \frac{\theta_{\mathcal{P}} + (1-\tau)^{-1}\|\psi\|_\infty}{(1-\tau)\theta_{\mathcal{P}} - \|\psi\|_{\mathrm{L}}}.$$

Proof. With the norms considered and following a proof similar to Lemma 3.6, the operator norm $\|\mathcal{A}\|$ can be upper bounded as $\|\mathcal{A}\| \leq 1+\tau$. The *inf-sup* condition, Assumption 3.1(ii), holds with $\gamma = 1-\tau$, since

$$\inf_{y\in\mathbb{K}^*} \sup_{x\in\mathbb{X}_n} \frac{\langle \mathcal{A}x, y\rangle}{\|x\|\|y\|_{\mathrm{W}}} \geq \inf_{y\in\mathbb{K}^*} \frac{(1-\tau)\langle \mathbb{1}, y\rangle}{\|y\|_{\mathrm{W}}} = 1-\tau.$$

Moreover $\|\nu\|_{\mathrm{W}} = 1$ since it is a probability measure. Thus, given the lower bound for the optimal value $J_n^{\mathrm{DC}} \geq -(1-\tau)^{-1}\|\psi\|_\infty$, the assertion of Proposition 3.2 (i.e, the dual optimizers bound in (3.3)) leads to the desired assertion (A.2). ∎

Note that when the norm constraint is neglected, the dual program enforces that any solution $y_n^\star$ in the program $\mathcal{D}_n$ satisfies $\langle x, \mathcal{A}^* y_n^\star - c\rangle = 0 \ \forall\, x \in \mathbb{X}_n$ (cf. the program $\mathcal{D}$). Assume that a constant function belongs to $\mathbb{X}_n$. Then, the constraint evaluated at the constant function reduces to $(1-\tau)\langle \mathbb{1}, y_n^\star\rangle = (1-\tau)\|y_n^\star\|_{\mathrm{W}} = 1$. It is worth noting that this observation can consistently be captured by Lemma A.2 when $\theta_{\mathcal{P}}$ tends to ∞, in which the bound (A.2) reduces to $\|y_n^\star\|_{\mathrm{W}} \leq (1-\tau)^{-1}$.

REFERENCES

[1] E. J. Anderson and P. Nash, *Linear Programming in Infinite-Dimensional Spaces*, John Wiley & Sons, Chichester, UK, 1987.

[2] A. Arapostathis, V. Borkar, E. Fernandez-Gaucherand, M. Ghosh, and S. Marcus, *Discrete-time controlled Markov processes with average cost criterion: A survey*, SIAM J. Control Optim., 31 (1993), pp. 282–344.

[3] A. Basu and V. S. Borkar, *Stochastic control with imperfect models*, SIAM J. Control Optim., 47 (2008), pp. 1274–1300.

[4] A. Ben-Tal, L. Ghaoui, and A. Nemirovski, *Robust Optimization*, Princeton University Press, Princeton, NJ, 2009.

[5] D. Bertsekas, *Convergence of discretization procedures in dynamic programming*, IEEE Trans. Automat. Control, 20 (1975), pp. 415–419.

[6] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, Vol. II, 4th ed., Athena Scientific, Belmont, MA, 2012.

[7] D. P. Bertsekas and S. E. Shreve, *Stochastic Optimal Control*, Academic Press, New York, 1978.

[8] D. P. Bertsekas and J. N. Tsitsiklis, *Neuro-Dynamic Programming*, Athena Scientific, Belmont, MA, 1996.

[9] D. Bertsimas, X. V. Doan, and J. Lasserre, *Approximating integrals of multivariate exponentials: A moment approach*, Oper. Res. Lett., 36 (2008), pp. 205–210.

[10] M. C. Campi and S. Garatti, *The exact feasibility of randomized solutions of uncertain convex programs*, SIAM J. Optim., 19 (2008), pp. 1211–1230.

[11] O. L. V. Costa and F. Dufour, *A linear programming formulation for constrained discounted continuous control for piecewise deterministic Markov processes*, J. Math. Anal. Appl., 424 (2015), pp. 892–914.

[12] I. Csiszár, *I-divergence geometry of probability distributions and minimization problems*, Ann. Probab., 3 (1975), pp. 146–158.

[13] D. P. de Farias and B. Van Roy, *The linear programming approach to approximate dynamic programming*, Oper. Res., 51 (2003), pp. 850–865.

[14] D. P. De Farias and B. Van Roy, *On constraint sampling in the linear programming approach to approximate dynamic programming*, Math. Oper. Res., 29 (2004), pp. 462–478.

[15] O. Devolder, F. Glineur, and Y. Nesterov, *First-order methods of smooth convex optimization with inexact oracle*, Math. Program., 146 (2014), pp. 37–75.

[16] F. Dufour and T. Prieto-Rumeau, *Approximation of Markov decision processes with general state space*, J. Math. Anal. Appl., 388 (2012), pp. 1254–1267.

[17] F. Dufour and T. Prieto-Rumeau, *Finite linear programming approximations of constrained discounted Markov decision processes*, SIAM J. Control Optim., 51 (2013), pp. 1298–1324.

[18] F. Dufour and T. Prieto-Rumeau, *Stochastic approximations of constrained discounted Markov decision processes*, J. Math. Anal. Appl., 413 (2014), pp. 856–879.

[19] F. Dufour and T. Prieto-Rumeau, *Approximation of average cost Markov decision processes using empirical distributions and concentration inequalities*, Stochastics, 87 (2015), pp. 273–307.

[20] M. J. Eisner and P. Olsen, *Duality for stochastic programming interpreted as lp in L_p-space*, SIAM J. Appl. Math., 28 (1975), pp. 779–792.
[21] A. Ern and J.-L. Guermond, *Theory and Practice of Finite Elements*, Springer, New York, 2004.
[22] R. T. Farouki, *The Bernstein polynomial basis: A centennial retrospective*, Comput. Aided Geom. Design, 29 (2012), pp. 379–419.
[23] O. Hernández-Lerma, *Adaptive Markov Control Processes*, Appl. Math. Sci. 79, Springer, New York, 1989.
[24] O. Hernández-Lerma, J. González-Hernández, and R. López-Martínez, *Constrained average cost Markov control processes in Borel spaces*, SIAM J. Control Optim., 42 (2003), pp. 442–468.
[25] O. Hernández-Lerma and J. Lasserre, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, Springer, New York, 1996.
[26] O. Hernández-Lerma and J. Lasserre, *Approximation schemes for infinite linear programs*, SIAM J. Optim., 8 (1998), pp. 973–988.
[27] O. Hernández-Lerma and J. Lasserre, *Further Topics on Discrete-Time Markov Control Processes*, Springer, New York, 1999.
[28] V. R. Konda and J. N. Tsitsiklis, *On actor-critic algorithms*, SIAM J. Control Optim., 42 (2003), pp. 1143–1166.
[29] H. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, Stoch. Model. Appl. Probab. 35, Springer, New York, 2003.
[30] H. Lai and S. Wu, *Extremal points and optimal solutions for general capacity problems*, Math. Program., 54 (1992), pp. 87–113.
[31] J. B. Lasserre, *Moments, Positive Polynomials and Their Applications*, Imperial College Press, London, 2009.
[32] P. Mohajerin Esfahani, D. Chatterjee, and J. Lygeros, *Motion planning for continuous time stochastic processes: A dynamic programming approach*, IEEE Trans. Automat. Control, 61 (2016), pp. 2155–2170.
[33] P. Mohajerin Esfahani, D. Chatterjee, and J. Lygeros, *The stochastic reach-avoid problem and set characterization for diffusions*, Automatica, 70 (2016), pp. 43–56.
[34] P. Mohajerin Esfahani, T. Sutter, D. Kuhn, and J. Lygeros, *From Infinite to Finite Programs: Explicit Error Bounds with Applications to Approximate Dynamic Programming*, arXiv:1701.06379, 2017.
[35] P. Mohajerin Esfahani, T. Sutter, and J. Lygeros, *Performance bounds for the scenario approach and an extension to a class of non-convex programs*, IEEE Trans. Automat. Control, 60 (2015), pp. 46–58.
[36] A. Nemirovski and A. Shapiro, *Scenario approximations of chance constraints*, in Probabilistic and Randomized Methods for Design Under Uncertainty, G. Calafiore and F. Dabbene, eds., Springer, New York, 2006, pp. 3–47.
[37] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, Springer, New York, 2004.
[38] Y. Nesterov, *Smooth minimization of non-smooth functions*, Math. Program., 103 (2005), pp. 127–152.
[39] P. Olsen, *Discretizations of multistage stochastic programming problems*, in Stochastic Systems: Modeling, Identification and Optimization, II, Springer, New York, 1976, pp. 111–124.
[40] P. Olsen, *Multistage stochastic programming with recourse as mathematical programming in an L_p space*, SIAM J. Control Optim., 14 (1976), pp. 528–537.
[41] S. Olver, *On the convergence rate of a modified Fourier series*, Math. Comp., 78 (2009), pp. 1629–1645.
[42] A. Ruszczyński, *Risk-averse dynamic programming for Markov decision processes*, Math. Program., 125 (2010), pp. 235–261.
[43] N. Saldi, S. Yüksel, and T. Linder, *Asymptotic Optimality of Finite Approximations to Markov Decision Processes with General State and Action Spaces*, arXiv:1503.02244, 2015.
[44] E. Shafieepoorfard, M. Raginsky, and S. P. Meyn, *Rationally inattentive control of Markov processes*, SIAM J. Control Optim., 54 (2016), pp. 987–1016.
[45] M. Sion, *On general minimax theorems*, Pacific J. Math., 8 (1958), pp. 171–176.
[46] T. Sutter, P. Mohajerin Esfahani, and J. Lygeros, *Approximation of constrained average cost Markov control processes*, in Proceedings of the 53rd IEEE Conference on Decision and Control, 2014, pp. 6597–6602.
[47] J. N. Tsitsiklis and B. V. Roy, *An analysis of temporal-difference learning with function approximation*, IEEE Trans. Automat. Control, 42 (1997), pp. 674–690.
[48] C. Villani, *Topics in Optimal Transportation*, AMS, Providence, RI, 2003.