# Approximation of Constrained Average Cost Markov Control Processes

Tobias Sutter, Peyman Mohajerin Esfahani, and John Lygeros

***Abstract*— This paper considers discrete-time constrained Markov control processes (MCPs) under the long-run expected average cost optimality criterion. For Borel state and action spaces a two-step method is presented to numerically approximate the optimal value of this constrained MCPs. The proposed method employs the infinite-dimensional linear programming (LP) representation of the constrained MCPs. In particular, we establish a bridge from the infinite-dimensional LP characterization to a finite LP consisting of a first asymptotic step and a second step that provides explicit bounds on the approximation error. Finally, the applicability and performance of the theoretical results are demonstrated on an LQG example.**

## I. Introduction

Discrete-time Markov control processes (MCPs) are a class of stochastic control problems that appear in many fields, for example engineering, economics, operations research, etc. Oftentimes it is impossible to obtain an explicit solution of such MCP problems, which motivates the task of finding tractable approximations leading to explicit solutions. Such approximation schemes are the core of a methodology known as *approximate dynamic programming* [1], which has been extensively studied in the literature from different perspectives [2], [3], [4]; see [5] for a comprehensive survey on this field.

Most MCPs (discrete or continuous time, finite or infinite space, constrained or unconstrained and finite or infinite horizon) can be recast as abstract "static" optimization problems over a closed convex set of measures and become infinite-dimensional convex programs, see for example [6], [7]. Hernández-Lerma and Lasserre investigate the linear programming (LP) approach [8], [9], [10] to discrete-time MCPs with Borel state and control spaces for infinite-horizon expected average and discounted costs. This reformulation allows the use of tools from the well-established field of mathematical programming to tackle MCPs. Furthermore, representing an MCP by means of an infinite-dimensional linear program is particularly appealing from the perspective of dealing with unconventional MCPs involving additional constrains or secondary costs, where traditional dynamic programming techniques are not applicable [11], [12], [13]. It is therefore desirable to derive an approximation scheme for such infinite-dimensional LPs that is computationally tractable while providing a performance bound on the approximation error.

Research supported by the the ETH grant (ETH-15 12-2) and the HYCON2 Network of Excellence (FP7-ICT-2009-5).

The authors are with the Automatic Control Laboratory, ETH Zürich, 8092 Zürich, Switzerland; Emails: {sutter, mohajerin, lygeros}@control.ee.ethz.ch

In the literature, to the best of our knowledge, there are two approximation schemes to tackle such infinite LPs. The first method [9], [14] is based on approximating the infinite-dimensional LPs by finite LPs and provides asymptotic convergence guarantees. The main difficulty in practically using this scheme is that the convergence proof is an existence proof and is not constructive. Furthermore, there are no explicit error bounds available. The second, quite recent method [12] is based on approximating a probability measure that underlies the random transitions of the dynamics of the system using a discretization procedure, known as quantization. While the method [12] can provide explicit error bounds, it is based on solving non-convex optimization problems, which in general are NP-hard.

The objective here is to build an approximation method for the linear programming formulation of constrained Markov control problems with special emphasis on its computational efficiency. Our approach consists of two steps: First, we show how to build a semi-infinite relaxation of the original infinite linear program. This step is specifically designed to lead to linear programs with a particular structure which is numerically desirable for the successive step. Then, the semi-infinite relaxation is approximated in a second step by finite linear programs. For this second step we propose two independent methods, one based on probabilistic approximation techniques for robust convex programs and the other on an adaptive cutting plane algorithm.

The layout of this paper is as follows: Section II introduces the notation and general framework of MCPs on Borel spaces. In Section III we present the problem statement, namely the infinite-dimensional linear program characterizing the constrained average cost MCPs. The two stage approximation scheme for those LPs is introduced in Sections IV-A and IV-B. To illustrate the proposed methodology, in Section V, the theoretical results are applied to an infinite-horizon average cost LQG problem and compared with the explicit optimal solution. We conclude in Section VI with a summary of our work and comment on possible subjects of further research.

## II. Preliminaries and Notation

We briefly recall standard definitions below and refer interested readers to [8], [11], [15], [16] for further details. A *constrained Markov control model* is the tuple

$$(X, A, \{A(x)|x \in X\}, Q, c, d, \ell)\,,$$

where $X$ (resp. $A$) is a Borel space, i.e., a Borel subset of a complete and separable metric space called the *state space*

(resp. *action* or *control space*). $\{A(x)|x \in X\}$ is a family of nonempty measurable subsets of $A$, where $A(x)$ denotes the set of *feasible actions* when the system is in state $x \in X$. The *transition law* is a stochastic kernel $Q$ on $X$ given the feasible state-action pairs $\mathbb{K} := \{(x,a)|x \in X, a \in A(x)\}$. A stochastic kernel acts on measurable functions $u$ from the left as

$$Qu(x,a) := \int_X u(y)Q(\,\mathrm{d}y|x,a), \quad \forall (x,a) \in \mathbb{K}$$

and on probability measures $\mu$ on $\mathbb{K}$ from the right as

$$\mu Q(B) := \int_{\mathbb{K}} Q(B|x,a)\mu(\,\mathrm{d}(x,a)), \quad \forall B \in \mathcal{B}(X)\,.$$

Finally $c, d : \mathbb{K} \to \mathbb{R}_{\geq 0}$ denote measurable functions, where $c$ is called the *one-stage cost function* and $\ell \in \mathbb{R}$ is a constant. The *admissible history spaces* are defined recursively as $H_0 := X$ and $H_t := H_{t-1} \times \mathbb{K}$ for $t \in \mathbb{N}$ and the canonical sample space is defined as $\Omega := (X \times A)^\infty$. These spaces are endowed with their respective product topologies and are therefore Borel spaces. A generic element $\omega \in \Omega$ is of the form $\omega = (x_0, a_0, x_1, a_1, \ldots)$, $x_i \in X$, $a_i \in A$; all random variables will be defined on the measurable space $(\Omega, \mathcal{B}(\Omega))$. The projections $x_t$ and $a_t$ from $\Omega$ to the sets $X$ and $A$ are called state and action variables, respectively. An *admissible policy* is a sequence $\pi = (\pi_t)_{t\in\mathbb{N}_0}$ of stochastic kernels $\pi_t$ on $A$ given $H_t$, satisfying the constraints $\pi_t(A(x_t)|h_t) = 1$, $x_t \in X$ and $h_t \in H_t$. The set of admissible policies will be denoted by $\Pi$. Given a probability measure $\nu \in \mathcal{P}(X)$ and $\pi \in \Pi$, there exists a unique probability measure $\mathbb{P}^\pi_\nu$ on $(\Omega, \mathcal{B}(\Omega))$ such that for all $B \in \mathcal{B}(X)$, $C \in \mathcal{B}(A)$ and $h_t \in H_t$, $t \in \mathbb{N}_0$

$$\begin{aligned}\mathbb{P}^\pi_\nu\big(x_0 \in B\big) &= \nu(B)\\ \mathbb{P}^\pi_\nu\big(a_t \in C|h_t\big) &= \pi_t(C|h_t)\\ \mathbb{P}^\pi_\nu\big(x_{t+1} \in B|h_t, a_t\big) &= Q(B|x_t, a_t).\end{aligned}$$

The expectation operator with respect to $\mathbb{P}^\pi_\nu$ is denoted by $\mathbb{E}^\pi_\nu$.

*Definition 1:* The stochastic process $(\Omega, \mathcal{B}(\Omega), \mathbb{P}^\pi_\nu, \{x_t\}_{t\in\mathbb{N}_0})$ is called a discrete-time *Markov control process*.

Let $\mathcal{P}(X)$ denote the space of all probability measures on $X$. In the following we focus on constrained long-run average cost problems, i.e.,

$$\rho_{\min} = \inf_{\nu,\pi}\{J_c(\pi,\nu) \;:\; (\pi,\nu) \in \Delta\}, \tag{1}$$

where $\Delta := \{(\pi,\nu) \in \Pi \times \mathcal{P}(X) \;:\; J_c(\pi,\nu) < \infty$ and $J_d(\pi,\nu) \leq \ell\}$ and

$$\begin{aligned}J_c(\pi,\nu) &:= \limsup_{n\to\infty} \frac{1}{n}\mathbb{E}^\pi_\nu\left(\sum_{t=0}^{n-1} c(x_t,a_t)\right)\\ J_d(\pi,\nu) &:= \limsup_{n\to\infty} \frac{1}{n}\mathbb{E}^\pi_\nu\left(\sum_{t=0}^{n-1} d(x_t,a_t)\right).\end{aligned}$$

Let $S$ be a Borel space with Borel $\sigma$-algebra $\mathcal{B}(S)$. We denote by $\mathbb{M}(S)$ the vector space of finite signed measures on $\mathcal{B}(S)$ and by $\mathbb{B}(S)$ the vector space of real-valued measurable functions on $S$. Let $\langle\cdot,\cdot\rangle$ be a bilinear form on $\mathbb{M}(S)\times\mathbb{B}(S)$ defined by $\langle\mu, f\rangle := \int_S f(x)\,\mathrm{d}\mu(x)$. Define $\mathcal{P}_c(S)$ as the set of probability measures $\mu$ on $S$ such that $\langle\mu, c\rangle < \infty$ for some $c \in \mathbb{B}(S)$. The space of continuous bounded functions (with respect to the sup-norm) is denoted by $\mathcal{C}_b(S)$ and $\mathcal{C}_0(S)$ denotes the continuous functions vanishing at infinity.

## III. Infinite LP Characterization

We start by stating all the assumptions imposed on the control model which hold throughout the paper.

*Assumption 2:*

(i) The set $\Delta$ is nonempty.

(ii) The cost function $c$ is lower semicontinuous and inf-compact, i.e., for each $r \in \mathbb{R}$ the set $\{(x,a) \in \mathbb{K} \mid c(x,a) \leq r\}$ is compact.

(iii) The cost function $c$ is strictly unbounded (coercive), i.e., there is a nondecreasing sequence of compact sets $K_n \uparrow \mathbb{K}$ such that

$$\liminf_{n\to\infty}\{c(x,a)|(x,a) \notin K_n\} = \infty.$$

(iv) The transition law $Q$ is weakly continuous, i.e., $Qu \in \mathcal{C}_b(\mathbb{K})$ for any $u \in \mathcal{C}_b(X)$

(v) $d(\cdot)$ is lower semicontinuous and bounded.

If $\mathbb{K}$ is a compact set, Assumption (iii) readily holds, see [9, Remark 11.4.2]. Note that $\mathcal{C}_0(X)$ is a separable Banach space [9, p. 207] and let $\mathcal{C}(X) := \{u_k\}_{k\in\mathbb{N}}$ be a countable dense subset of $\mathcal{C}_0(X)$. Consider the (infinite) linear program

$$\mathsf{P}: \begin{cases} \min\limits_{\mu} & \langle\mu, c\rangle \\ \text{s.t.} & \langle \mathsf{L}\mu, u\rangle = 0 \quad \forall u \in \mathcal{C}(X) \\ & \langle\mu, d\rangle \leq \ell \\ & \mu \in \mathcal{P}_c\big(\mathbb{K}\big)\,, \end{cases} \tag{2}$$

where $\mathsf{L} : \mathbb{M}(\mathbb{K}) \to \mathbb{M}(X)$ denotes a linear, weakly continuous operator defined as [9]

$$\mathsf{L}\mu(B) := \mu(B \times A) - \mu Q(B) \quad \forall B \in \mathcal{B}(X)\,.$$

We denote the optimal solution (that exists [9, Theorem 12.3.3]) and optimum value to (2), respectively, by $\mu^\star$ and $J^\star$. The linear programming formulation (2) is an alternative characterization of the problem (1) in the sense of the following theorem.

*Theorem 3:* Under Assumption 2, $\rho_{\min} = J^\star$.

*Proof:* The proof follows directly by combining [9, Lemma 12.5.2], [9, Theorem 12.3.3] and [11, Lemma 3.5]. ∎

The focus of our study is on providing an approximation scheme for the linear program (2). The proposed method consists of two steps: First $\mathsf{P}$ is relaxed by a semi-infinite linear program, which then in the second step is approximated by a finite linear program.

## IV. Finite LP Approximation

### A. Step (I): From infinite to semi-infinite LP

For each $k \in \mathbb{N}$ and $\lambda \in \mathbb{R}_{\geq 0}$ we consider the relaxed, semi-infinite linear program

$$\mathsf{P}^{(k)}(\lambda): \begin{cases} \min\limits_{\mu,\eta} & \langle \mu, c\rangle + \lambda\eta \\ \text{s.t.} & \big|\langle \mathsf{L}\mu, u_i\rangle\big| \leq \eta \quad \forall i \leq k \\ & \langle \mu, d\rangle \leq \ell + \eta \\ & \mu \in \mathcal{P}_c\big(\mathbb{K}\big)\,,\ \eta \in \mathbb{R}_{\geq 0}, \end{cases} \tag{3}$$

where $\mathcal{C}_k := \{u_1, \ldots, u_k\}$ denotes an increasing sequence such that $\bigcup_{k\in\mathbb{N}} \mathcal{C}_k = \mathcal{C}(X)$. We denote the optimal solution and optimum value, respectively, by $(\mu_k^\star(\lambda), \eta_k^\star(\lambda))$ and $J_k^\star(\lambda)$. The penalization term $\lambda\eta$ in the cost function of the relaxed program (3) allows us to provide a priori bounds for the dual variables associated with the constraints in (3). This property is of particular interest for the next step, which will be elaborated in Section IV-B.

The following result, Theorem 4, establishes an asymptotic link from the infinite linear program $\mathsf{P}$ to the semi-infinite relaxation $\mathsf{P}^{(k)}(\lambda)$.

*Theorem 4:* Under Assumption 2, we have

1) $\mathsf{P}^{(k)}(\lambda)$ is solvable for every $k \in \mathbb{N}$ and $\lambda \in \mathbb{R}_{\geq 0}$, i.e., the minimum in (3) exists.
2) Let $(\mu_k^\star(\lambda), \eta_k^\star(\lambda))$ be an optimizer of $\mathsf{P}^{(k)}(\lambda)$ and denote $J_k^\star(\lambda) := \langle \mu_k^\star(\lambda), c\rangle + \lambda\eta_k^\star(\lambda)$. Then, the sequence $\{J_k^\star(\lambda)\}_{k\in\mathbb{N},\ \lambda\in\mathbb{R}_{\geq 0}}$ is monotonically increasing in $(k, \lambda)$. Furthermore,

$$\lim_{k,\lambda\to\infty} J_k^\star(\lambda) = J^\star.$$

*Proof:* The proof effectively follows the same lines as in [14, Theorem 12.5.3] with an extension to allow for a penalization term as well as the constraint MCP setting. We refer to Appendix I for further details. ∎

Preparatory to the second step toward our approximation scheme, we dualize the problem $\mathsf{P}^{(k)}(\lambda)$. As shown in Appendix II with a detailed derivation, the dual of the linear program (3) is given by

$$\mathsf{D}^{(k)}(\lambda): \begin{cases} \max\limits_{\rho,\gamma,\alpha,\beta} & \rho - \gamma\ell \\ \text{s.t.} & \rho + \sum_{i=1}^k (\alpha_i - \beta_i)\mathsf{L}^\star u_i(x,a) \\ & \qquad \leq \gamma d(x,a) + c(x,a) \quad \forall (x,a) \in \mathbb{K} \\ & \gamma + \sum_{i=1}^k (\alpha_i + \beta_i) \leq \lambda \\ & \rho \in \mathbb{R},\ \gamma \in \mathbb{R}_{\geq 0},\ \alpha, \beta \in \mathbb{R}^k_{\geq 0}, \end{cases}$$

where $\mathsf{L}^\star : \mathbb{B}(X) \to \mathbb{B}(\mathbb{K})$ is the adjoint operator of $\mathsf{L}$ given by $(\mathsf{L}^\star u)(x,a) := u(x) - Qu(x,a)$. The optimization problem $\mathsf{D}^{(k)}(\lambda)$ is a standard robust linear program [17]. As mentioned before, by looking at the optimization problem $\mathsf{D}^{(k)}(\lambda)$, the constraint $\gamma + \sum_{i=1}^k (\alpha_i + \beta_i) \leq \lambda$ provides an a priori bound for the optimzation variables $\gamma, \alpha$ and $\beta$. Furthermore, the optimal value of $\mathsf{D}^{(k)}(\lambda)$ is upper bounded by $J^\star$, as shown in Lemma 5 below. This implies that all the optimization variables of the robust linear program $\mathsf{D}^{(k)}(\lambda)$ are bounded, which is a desirable property for both numerical solvers as well as our second approximation step presented in the subsequent section.

*Lemma 5:* Under Assumption 2 and for any $\lambda \in \mathbb{R}_{\geq 0}$ there is no duality gap between $\mathsf{P}^{(k)}(\lambda)$ and $\mathsf{D}^{(k)}(\lambda)$.

*Proof:* As $\mathsf{P}^{(k)}(\lambda)$ has a min-max problem structure, it can be seen that there is pair $(\mu_0, \eta_0) \in \mathcal{P}_c\big(\mathbb{K}\big) \times \mathbb{R}_{\geq 0}$ such that $\max\limits_{i=1,\ldots,k} \big|\langle \mathsf{L}\mu_0, u_i\rangle\big| < \eta_0$. According to Theorem 4 $\mathsf{P}^{(k)}(\lambda)$ has a finite optimal value for any $\lambda \in \mathbb{R}_{\geq 0}$. Hence, according to [18, Theorem 3.13] there is no duality gap. ∎

### B. Step (II): From semi-infinite to finite LP

The second approximation step establishes a link from the relaxation of the preceding step to a finite linear program. To this end, we propose two approaches. One relies on (random) sampling techniques whose performance is quantified based on the relation between the robust convex program and its corresponding scenario convex program, as recently derived in [19]. The second is a (deterministic) adaptive method, based on a cutting plane iteration scheme. Both methods lead to explicit bounds on the approximation error.

*1) Scenario Based Approximation Scheme:* In this subsection we provide a tractable approximation to the semi-infinite linear programs of the form $\mathsf{D}^{(k)}(\lambda)$, that are in general known to be computationally intractable — NP-hard [17, p. 16]. We propose an approximation by using the scenario approach which is based on sampling techniques. To this end, we endow the set $\mathbb{K}$ with its Borel $\sigma$-algebra $\mathcal{B}(\mathbb{K})$ and consider a probability measure $\mathbb{P}$ on $(\mathbb{K}, \mathcal{B}(\mathbb{K}))$. Suppose $\{(x_i, a_i)\}_{i=1}^N$ are $N$ independent and identically distributed (i.i.d.) samples extracted according to the probability measure $\mathbb{P}$. We introduce the following random (scenario) linear program

$$\mathsf{D}_N^{(k)}(\lambda): \begin{cases} \max\limits_{\rho,\gamma,\alpha,\beta} & \rho - \gamma\ell \\ \text{s.t.} & \rho + \sum_{j=1}^k (\alpha_j - \beta_j)\mathsf{L}^\star u_j(x_i,a_i) \\ & \qquad \leq \gamma d(x_i,a_i) + c(x_i,a_i) \quad \forall i \leq N \\ & \gamma + \sum_{j=1}^M (\alpha_j + \beta_j) \leq \lambda \\ & \rho \in \mathbb{R},\ \gamma \in \mathbb{R}_{\geq 0},\ \alpha, \beta \in \mathbb{R}^k_{\geq 0}, \end{cases} \tag{4}$$

where the optimal solution and optimum value are denoted, respectively, by $(\rho_{k,N}^\star(\lambda), \gamma_{k,N}^\star(\lambda), \alpha_{k,N}^\star(\lambda), \beta_{k,N}^\star(\lambda))$ and $J_{k,N}^\star(\lambda)$. We introduce the following technical assumption.

*Assumption 6:* The problem $\mathsf{D}_N^{(k)}(\lambda)$ admits unique and measurable optimizers.

See [19, p. 6] how one may rigorously address this issue without any assumption. The optimization program $\mathsf{D}_N^{(k)}(\lambda)$ in (4) is a standard linear program, and hence tractable for large number of constraints and decision variables. A natural question is whether there exist theoretical links from $\mathsf{D}_N^{(k)}(\lambda)$

to $\mathsf{D}^{(k)}(\lambda)$ in terms of objective performance. The answer requires the following definition.

*Definition 7 ([19]):* The *tail probability of the worst-case violation* is the function $p : \mathbb{R}_{\geq 0} \times \mathbb{R}^M \times \mathbb{R}^M \times \mathbb{R}_{\geq 0} \to [0,1]$ defined as

$$p(\gamma,\alpha,\beta,\delta) := \mathbb{P}\Big[\sup_{(\tilde{x},\tilde{a})\in\mathbb{K}} \Big\{ \sum_{j=1}^{k}(\alpha_j-\beta_j)\mathsf{L}^\star u_j(\tilde{x},\tilde{a}) - \gamma d(\tilde{x},\tilde{a}) - c(\tilde{x},\tilde{a})\Big\} - \delta < \sum_{j=1}^{k}(\alpha_j-\beta_j)\mathsf{L}^\star u_j(x,a) - \gamma d(x,a) - c(x,a)\Big].$$

We call $h : [0,1] \to \mathbb{R}_{\geq 0}$ a *uniform level-set bound* (ULB) of $p$ if for all $\varepsilon \in [0,1]$

$$h(\varepsilon) \geq \sup\Big\{\delta \in \mathbb{R}_{\geq 0} \;\Big|\; \min_{\substack{\alpha,\beta,\gamma \\ \|(\alpha,\beta,\gamma)\|_\infty \leq \lambda}} p(\gamma,\alpha,\beta,\delta) \leq \varepsilon\Big\}.$$

A ULB can be used to derive a probabilistic bound on the quality of $\mathsf{D}_N^{(k)}(\lambda)$ as an approximation to $\mathsf{D}^{(k)}(\lambda)$ as shown in the following theorem.

*Theorem 8:* Consider the programs $\mathsf{D}^{(k)}(\lambda)$ and $\mathsf{D}_N^{(k)}(\lambda)$ with the associated optimum values $J_k^\star(\lambda)$ and $J_{k,N}^\star(\lambda)$, respectively. Assume that Assumption 6 holds and let $h$ be a ULB as introduced in Definition 7. Given $\varepsilon, \beta$ in $[0,1]$, for all $N \geq N(\varepsilon,\beta)$, where

$$N(\varepsilon,\beta) := \min\Big\{N \in \mathbb{N} \;\Big|\; \sum_{i=0}^{n-1}\binom{N}{i}\varepsilon^i(1-\varepsilon)^{N-i} \leq \beta\Big\},$$

we have

$$\mathbb{P}^N\Big[J_{k,N}^\star(\lambda) - J_k^\star(\lambda) \in \big[0, h(\varepsilon)\big]\Big] \geq 1-\beta. \tag{5}$$

*Proof:* The proof follows by the proof of Theorem 3.5 in [19] and by observing that problem $\mathsf{D}^{(k)}(\lambda)$ is a min-max problem (see [19, Remark 3.8]). ∎

*Remark 9 (ULB candidate):* If $\mathbb{K}$ is a compact set, a uniform level-set bound can be proposed under some mild assumption on $\mathbb{P}$, where $h(\varepsilon)$ converges to zero as $\varepsilon \to 0$, [19, Proposition 3.8].

*2) Adaptive Approximation Scheme:* The idea of the second method basically relies on a finite linear program with "important" sample constraints, as opposed to the random ones in the first method. For this purpose we propose an (adaptive) iterative scheme in which at each iteration an active constraint is added. For $\mathbb{K}_m := \{(x_1,a_1),\ldots,(x_m,a_m)\}$, consider the finite linear program

$$\mathsf{D}_{\mathbb{K}_m}^{(k)}(\lambda): \begin{cases} \max\limits_{\rho,\gamma,\alpha,\beta} & \rho - \gamma\ell \\ \text{s.t.} & \rho + \sum_{j=1}^{k}(\alpha_j-\beta_j)\mathsf{L}^\star u_j(x_i,a_i) \\ & \quad \leq \gamma d(x_i,a_i) + c(x_i,a_i) \quad \forall i \leq m \\ & \gamma + \sum_{j=1}^{M}(\alpha_j+\beta_j) \leq \lambda \\ & \rho \in \mathbb{R},\ \gamma \in \mathbb{R}_{\geq 0},\ \alpha,\beta \in \mathbb{R}_{\geq 0}^k \end{cases}$$

and denote its optimal value by $J_{k,m}^\star(\lambda)$. Define

$$\delta(\rho,\gamma,\alpha,\beta) := \sup_{(x,a)\in\mathbb{K}} \Big\{\rho + \sum_{j=1}^{k}(\alpha_j-\beta_j)\mathsf{L}^\star u_j(x,a) - \gamma d(x,a) - c(x,a)\Big\}.$$

The constraints of the approximating linear program $\mathsf{D}_{\mathbb{K}_m}^{(k)}(\lambda)$, given by $\mathbb{K}_m$, are constructed iteratively via a basic cutting plane algorithm that is described in Algorithm 1.

**Algorithm 1:** Cutting Plane Method

| | |
|---|---|
| **Step 1:** | Set $m = s > 0$, $\mathbb{K}_m := \{(x_1,a_1),\ldots,(x_m,a_m)\}$, $\varepsilon \in \mathbb{R}_{\geq 0}$ arbitrary small |
| **Step 2:** | Solve $\mathsf{D}_{\mathbb{K}_m}^{(k)}(\lambda)$, denote by $(\rho^m,\gamma^m,\alpha^m,\beta^m)$ its optimizer |
| **Step 3:** | Calculate $\delta(\rho^m,\gamma^m,\alpha^m,\beta^m)$ and denote its maximizer by $(x^{m+1},a^{m+1})$ |
| **Step 4:** | If $\delta(\rho^m,\gamma^m,\alpha^m,\beta^m) < \varepsilon$, stop and output $\rho^m - \gamma^m\ell$ as the solution |
| **Step 5:** | Set $\mathbb{K}_{m+1} := \mathbb{K}_m \cup \{(x^{m+1},a^{m+1})\}$, update $m := m+1$, then go to Step 2 |

*Lemma 10:* Let $(\rho^m,\gamma^m,\alpha^m,\beta^m)$ be an optimal solution for problem $\mathsf{D}_{\mathbb{K}_m}^{(k)}(\lambda)$. Then

$$J_{k,m}^\star(\lambda) - J_k^\star(\lambda) \leq \delta(\rho^m,\gamma^m,\alpha^m,\beta^m).$$

*Proof:* The proof is based on the fact that under the strong duality condition the so-called perturbation function of convex optimization problems, is Lipschitz continuous (see [20, p. 250] for the proof and [21, Section 28] for more details in this direction). By the particular min-max structure of problem $\mathsf{D}_{\mathbb{K}_m}^{(k)}(\lambda)$ one can see that the Lipschitz constant is given by 1 [19, Remark 3.5], which concludes the proof. ∎

Thanks to the boundedness of the decision variables of $\mathsf{D}^{(k)}(\lambda)$, which is inherited from the penalization term in the first approximation step, one can deduce the convergence behaviour of the adaptive scheme as follows.

*Remark 11:* Under the assumption of compactness of state and action space, and continuity of the cost-functions $c$ and $d$ one can prove that $\delta(\rho^m,\gamma^m,\alpha^m,\beta^m) \to 0$ as $m \to \infty$, where $(\rho^m,\gamma^m,\alpha^m,\beta^m)$ is obtained by Algorithm 1. See [22, Lemma 2.2] and [23] for further details.

We conclude the presentation of the second step of the approximation scheme with a short remark on its computational tractability. The randomized approximation scheme is attractive since it only requires to solve one LP, which can be done efficiently in practice for a very large number of constraints and decision variables. Its downside, however, is the number of samples required to achieve an $\varepsilon$-precise solution which may grow exponentially in the dimension of $\mathbb{K}$ [19, Remark 3.9]. The adaptive approach, as simulation results in the subsequent section reveal, often requires considerably less constraints. This, however, comes at the cost of solving a non-convex optimization problem in each

iteration step, namely in Step 3 of Algorithm 1, which might be computationally expensive. As another important difference, in the adaptive scheme, the explicit error bound is an a posteriori bound, in the sense that one cannot predict the number of iteration steps (in Algorithm 1) needed to achieve a certain precision. In contrast, the scenario-based approximation method provides an a priori error bound.

## V. Numerical Example — LQG Problem

The standard LQG problem consists of a linear system and a quadratic one-stage cost which emerges in many applications ranging from control engineering to mathematical finance. To illustrate our results we consider the simplest version of this problem, where state and input are scalar. Consider the linear system

$$x_{t+1} = \theta x_t + \rho a_t + \xi_t, \quad t = 0, 1, \dots, \tag{6}$$

with one-stage cost $c(x, a) = qx^2 + ra^2$, where $q \geq 0$ and $r > 0$ are given constants. The disturbances $\xi_t$ are i.i.d. random Gaussian variables, independent of the initial state $x_0$, with zero mean and finite variance $\sigma^2$. We assume that $X = A = \mathbb{R}$, that $\theta, \rho \in \mathbb{R}$ are given constants, that the function $d$ and the constant $\ell$ are both 0 and that $A(x) = A$ for all $x \in X$. The transition kernel $Q$ has a density function $q(y|x, a)$, i.e., $Q(B|x, a) = \int_B q(y|x, a)\,\mathrm{d}y$ for all $B \in \mathcal{B}(X)$, that is given by

$$q(y|x, a) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y - \theta x - \rho a)^2}{2\sigma^2}\right).$$

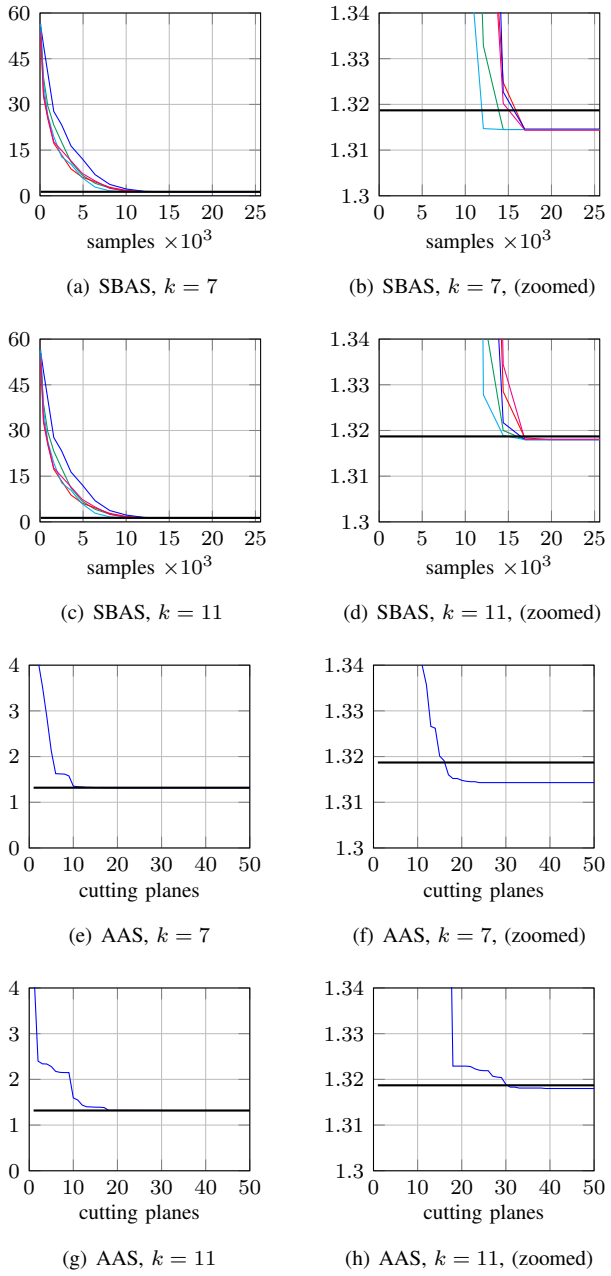
It is well known that the solution of the above LQG problem can be obtained via the algebraic Riccati equation [24, p. 372]. This value is depicted by the solid line in all the plots in Fig. 2. To validate our theoretical results, due to numerical purposes, we restrict our computations in the first approximation step, in Section IV-A, to a compact set $[-L, L]$, where $L$ is chosen large enough. On the space $\mathcal{C}_b([-L, L])$, we consider the Fourier basis

$$\mathcal{C}_{2k+1} := \{u_0, u_1, \dots, u_{2k}\},$$

where $u_0(x) := 1$, $u_{2k-1}(x) := \cos\left(\frac{k\pi x}{L}\right)$, and $u_{2k}(x) := \sin\left(\frac{k\pi x}{L}\right)$.

For the second approximation step in Section IV-B we verify the two presented methods. Regarding the scenario based approximation method, in Section IV-B.1, suppose that $\{(x_i, a_i)\}_{i=1}^N$ are $N$ independent and identically distributed (i.i.d.) samples according to the uniform distribution on $[-L, L]^2$. Then, we solve the respective scenario linear program $\mathsf{D}_N^{(k)}(\lambda)$, given in (4).

The numerical simulations of the resulting finite LP are depicted in Fig. 2(a) and 2(b) (respectively Fig. 2(c) and 2(d)) when the number of the basis functions in the first step is $k = 7$ (respectively $k = 11$). The corresponding results of the adaptive method described by Algorithm 1 are shown in Fig. 1(e) to 1(h). Let us remark that the finite number $k$, which corresponds to the first step relaxation,




(a) SBAS, $k = 7$ (b) SBAS, $k = 7$, (zoomed)

(c) SBAS, $k = 11$ (d) SBAS, $k = 11$, (zoomed)

(e) AAS, $k = 7$ (f) AAS, $k = 7$, (zoomed)

(g) AAS, $k = 11$ (h) AAS, $k = 11$, (zoomed)

Fig. 1. Numerical results for parameters $\theta = 0.8$, $\rho = 0.5$, $q = 1$, $r = 0.2$, $\sigma = 1$, with different number of basis functions $k$ and comparing the scenario based approximation scheme (SBAS), shown for 5 experiments, with the adaptive approximation scheme (AAS). The solid black line represents the exact solution obtained via the Riccati equation ($J^\star = 1.3187$).

influences the "steady state" level of the finite LPs coming out from the second step. As expected and also confirmed by simulation results in Fig. 2, the scenario approach requires significantly more samples in comparison with the adaptive approach proposed as the second method in this step. It, however, should be highlighted that this improvement comes at the cost of solving a static non-convex program at Step 3 in Algorithm 1. This is of course not a problem in this low dimensional example, but can become critical as the

dimension of the problem increases.

## VI. Conclusions

In this paper we presented a two-step approximation scheme for the linear programming formulation of discrete-time MCPs with Borel state and action spaces under the long-run average cost optimality criterion. The first approximation step bridges the infinite-dimensional LP characterization of MCPs to a semi-infinite relaxed LP. The second step transforms the latter problem to a finite LP through two different approaches: a randomized method based on the so-called scenario approach, and an adaptive method employing a cutting plane approach. Both methods lead to explicit bounds on the approximation error in the second step of the proposed scheme.

For future work, in light of the first part of our proposed two-stage approximation scheme, we aim to study the derivation of an explicit error bound. This, together with the second step presented here, would then lead to an explicit bound on the approximation error for the infinite-dimensional LP (2) and therefore for the average-cost MCP. Another open question is, given such an approximating scheme, how to find $\varepsilon$-approximating policies, i.e., policies whose corresponding cost is $\varepsilon$ away from the optimal value.

## Appendix I
## Proof of Theorem 4

Before proving the Theorem 4 we state a preliminary lemma.

*Lemma 12 ([14]):* Let the sequence $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathcal{P}_c(\mathbb{K})$ converge weakly to $\mu$ and let $c$ be a nonnegative lower semicontinuous function on $\mathbb{K}$. Then,

$$\liminf_{n\to\infty} \langle \mu_n, c\rangle \geq \langle \mu, c\rangle .$$

*Proof:* [Proof of Theorem 4] Since for any positive $\lambda$ and for any $k \in \mathbb{N}$, $\mathsf{P}^{(k)}(\lambda)$ is a relaxation of $\mathsf{P}$

$$0 \leq \inf \mathsf{P}^{(k)}(\lambda) \leq \min \mathsf{P} \quad \forall \lambda \in \mathbb{R}_{\geq 0}, \tag{7}$$

where solvability of $\mathsf{P}$ (Assumption 2(i)) and the fact that $c(\cdot,\cdot)$ is nonnegative were used. In a first step we assume that there exists a finite real positive number $M$ such that $\eta \in [0, M]$. This is without loss of generality; suppose there does not exist such a number, i.e., consider $\eta = \infty$. This, however, contradicts the fact that $\inf \mathsf{P}^{(k)}(\lambda)$ is finite for all $\lambda$ according to (7). Fix $\lambda$ and consider a minimizing sequence $\{(\mu_n, \eta_n)\}_{n\in\mathbb{N}}$ for $\mathsf{P}^{(k)}(\lambda)$, that is, each $\mu_n$ and $\eta_n$ satisfy $\mu_n \in \mathcal{P}_c(\mathbb{K})$, $\eta_n \in \mathbb{R}_{\geq 0}$, $|\langle \mathsf{L}\mu_n, u_i\rangle| \leq \eta_n$ for all $i = 1,\dots,k$, $\langle \mu_n, d\rangle \leq \ell$ and

$$\langle \mu_n, c\rangle + \lambda\eta_n \downarrow \inf \mathsf{P}^{(k)}(\lambda) .$$

Since $\lambda$, $\eta_n \in \mathbb{R}_{\geq 0}$, according to (7) there exists $M, N$ such that $\langle \mu_n, c\rangle \leq M$ for all $n \geq N$. Therefore invoking Assumptions 2(ii), (iii) and Theorem 12.2.15 in [9, p. 216] the family $\{\mu_n\}_{n\geq N}$ is tight. According to Prohorov's Theorem there exists a subsequence $\{\mu_m\}$ of $\{\mu_n\}$ and a probability measure $\mu$ on $\mathbb{K}$ such that $\langle \mu_m, v\rangle \to \langle \mu, v\rangle$ for all $v \in \mathcal{C}_b(\mathbb{K})$. Since $c(\cdot,\cdot)$ is nonnegative and lower semicontinuous by Assumption 2(ii), Lemma 12 states $\liminf_{m\to\infty} \langle \mu_m, c\rangle \geq \langle \mu, c\rangle$. Consider the subsequence $\{\eta_m\}_{m\in\mathbb{N}} \subset [0, M]$, by compactness there is a subsequence $\{\eta_j\}$ of $\{\eta_m\}$ such that $\eta_j \to \eta$ as $j \to \infty$. Obviously $\langle \mu_j, v\rangle \to \langle \mu, v\rangle$ for all $v \in \mathcal{C}_b(\mathbb{K})$ which gives

$$\liminf_{j\to\infty} \langle \mu_j, c\rangle + \lambda\eta_j \geq \langle \mu, c\rangle + \lambda\eta \tag{8}$$

Hence, it remains to show that the pair $(\mu, \eta)$ is feasible for $\mathsf{P}^{(k)}(\lambda)$. First note that $\langle \mu, c\rangle < \infty$ which implies that $\mu \in \mathcal{P}_c(\mathbb{K})$. Second $\eta \geq 0$ is without loss of generality. It remains to show that $|\langle \mathsf{L}\mu, u_i\rangle| \leq \eta$ for all $i = 1,\dots,k$. As a preliminary step observe that $\langle \mathsf{L}\mu_m, u\rangle \to \langle \mathsf{L}\mu, u\rangle$ for all $u \in \mathcal{C}_b(X)$. This is a consequence of $\mu_m$ converging weakly to $\mu$ as $\langle \mathsf{L}\mu_m, u\rangle = \langle \mu_m, \mathsf{L}^\star u\rangle \to \langle \mu, \mathsf{L}^\star u\rangle = \langle \mathsf{L}\mu, u\rangle$, where we used that $\mathsf{L}^\star$ maps $\mathcal{C}_b(X)$ into $\mathcal{C}_b(\mathbb{K})$. Therefore, $|\langle \mathsf{L}\mu_j, u\rangle| - \eta_j \to |\langle \mathsf{L}\mu, u\rangle| - \eta$. Since $|\langle \mathsf{L}\mu_j, u_i\rangle| - \eta_j \leq 0$ for all $j \in \mathbb{N}$ and for all $i = 1,\dots,k$ we get $|\langle \mathsf{L}\mu, u_i\rangle| \leq \eta$. Finally by Assumption 2(v) and Lemma 12 we get $\langle \mu, d\rangle \leq \ell$. This settles the assertion (1).

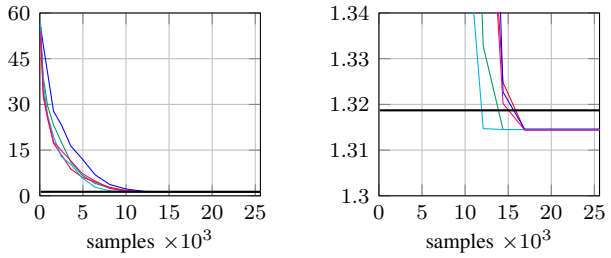
In order to show (2), for any $k = 1, 2, \dots$, let $(\mu_k, \eta_k)$ be an optimal solution for $\mathsf{P}^{(k)}(\lambda)$, where $\{\lambda_k\}_{k\in\mathbb{N}}$ is an arbitrary increasing sequence. Clearly $\langle \mu_k, c\rangle + \lambda_k\eta_k$ is nondecreasing. Therefore, combined with (7), there is a number $\rho$ such that $\langle \mu_k, c\rangle + \lambda_k\eta_k \uparrow \rho$, where $\rho \leq \min \mathsf{P}$. Following the same argumentation as in the proof of assertion (1) there is a subsequence $\{\mu_j\}$ of $\{\mu_k\}$ and a subsequence $\{\eta_j\}$ of $\{\eta_k\}$ and a probability measure $\mu$ on $\mathbb{K}$ such that $\mu_j \rightharpoonup \mu$ and an $\eta \geq 0$ such that $\eta_j \to \eta$. Also

$$\liminf_{j\to\infty} \langle \mu_j, c\rangle + \lambda_j\eta_j \geq \langle \mu, c\rangle + \lim_{j\to\infty} \lambda_j\eta_j. \tag{9}$$

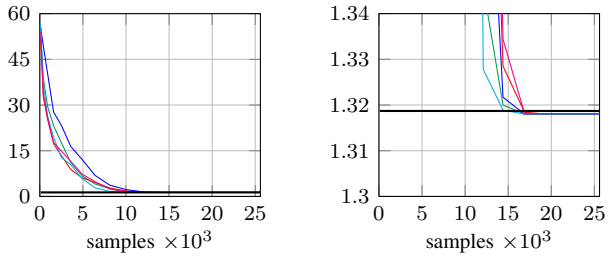
We claim that $\eta = 0$. Suppose not, i.e., $\eta > 0$, take $\lambda_k \uparrow \infty$ and recall that $\langle \mu_k, c\rangle \geq 0$. This, however, contradicts boundedness of $\mathsf{P}^{(k)}(\lambda)$ (7). Having (9) it remains to show that the pair $(\mu, \eta)$ is feasible for $\mathsf{P}$. First note that $\langle \mu, c\rangle < \infty$ and therefore $\mu \in \mathcal{P}_c(\mathbb{K})$, which follows directly from (7). To show that $\langle \mathsf{L}\mu_j, u\rangle = 0$ $\forall u \in \mathcal{C}(X)$, note that $\mathcal{C}(X) = \bigcup_{i=1}^{\infty}\{u_i\}$. Therefore, if $u \in \mathcal{C}(X)$, then there exists $N$ such that $u \in \bigcup_{i=1}^{k}\{u_i\}$ $\forall k \geq N$ and as such $|\langle \mathsf{L}\mu_j, u\rangle| \leq \eta_j$ $\forall j \geq N$ since $(\mu_j, \eta_j)$ is feasible for $\mathsf{P}^{(j)}(\lambda_j)$. As in the proof of (1) we have $|\langle \mathsf{L}\mu_j, u\rangle| \to |\langle \mathsf{L}\mu, u\rangle|$ and hence $\langle \mathsf{L}\mu, u\rangle = 0$ as well as $\langle \mu, d\rangle \leq \ell$. Therefore $\mu$ is feasible for $\mathsf{P}$. This together with (9) leads to

$$\begin{aligned}\min \mathsf{P} &\geq \liminf_{j\to\infty} \langle \mu_j, c\rangle + \lambda_j\eta_j \geq \langle \mu, c\rangle + \lim_{j\to\infty} \lambda_j\eta_j \\ &\geq \min \mathsf{P} + \lim_{j\to\infty} \lambda_j\eta_j,\end{aligned}$$

where $\lambda_j, \eta_j \in \mathbb{R}_{\geq 0}$ for all $j \in \mathbb{N}$. Hence, $\langle \mu_k, c\rangle + \lambda_k\eta_k \uparrow \min \mathsf{P}$, which settles the Theorem. ∎




(a) $(k = 7)$ (b) $(k = 7)$, (zoomed)




(c) $(k = 11)$ (d) $(k = 11)$, (zoomed)

Fig. 2. Numerical results for parameters $\theta = 0.8$, $\rho = 0.5$, $q = 1$, $r = 0.2$, $\sigma = 1$, with different number of basis functions $k$. The solid black line represents the exact solution obtained via the Riccati equation ($J^\star = 1.3187$).

## APPENDIX II
## DUAL OF THE LP (3)

This section shows how to derive the dual program of the LP (3). As a standard result in the theory of linear programming in infinite-dimensional spaces [18, p. 38], [25, p. 162] the dual of the LP (3) is given by

$$\left\{\begin{array}{ll} \sup\limits_{\rho,\gamma,\alpha,\beta} & \rho - \gamma\ell \\ \text{s.t.} & c + \gamma d - \rho - \sum_{i=1}^k (\alpha_i - \beta_i)\mathsf{L}^\star u_i \in \mathbb{M}(\mathbb{K})_+^\star \\ & \gamma + \sum_{i=1}^k (\alpha_i + \beta_i) \leq \lambda \\ & \rho \in \mathbb{R},\ \gamma \in \mathbb{R}_{\geq 0},\ \alpha, \beta \in \mathbb{R}_{\geq 0}^k, \end{array}\right.$$

where $\mathbb{M}(\mathbb{K})_+^\star$ denotes the dual cone of $\mathbb{M}(\mathbb{K})_+ := \{\mu \in \mathbb{M}(\mathbb{K}) \mid \mu \geq 0\}$, which coincides with the natural positive cone $\mathbb{B}(\mathbb{K})_+ := \{u \in \mathbb{B}(\mathbb{K}) | u \geq 0\}$, see [9, p. 212]. $\mathsf{L}^\star : \mathbb{B}(X) \to \mathbb{B}(\mathbb{K})$ is the adjoint operator of $\mathsf{L}$ given by $(\mathsf{L}^\star u)(x,a) := u(x) - Qu(x,a)$. Since in Euclidean spaces the intersection between a compact and closed set is compact and since the objective function is continuous the supremum is attained and therefore the linear program $\mathsf{D}^{(k)}(\lambda)$ is solvable. Hence the dual LP of (3) has the form

$$\left\{\begin{array}{ll} \max\limits_{\rho,\gamma,\alpha,\beta} & \rho - \gamma\ell \\ \text{s.t.} & \rho + \sum_{i=1}^k (\alpha_i - \beta_i)\mathsf{L}^\star u_i(x,a) \\ & \qquad \leq \gamma d(x,a) + c(x,a) \quad \forall (x,a) \in \mathbb{K} \\ & \gamma + \sum_{i=1}^k (\alpha_i + \beta_i) \leq \lambda \\ & \rho \in \mathbb{R},\ \gamma \in \mathbb{R}_{\geq 0},\ \alpha, \beta \in \mathbb{R}_{\geq 0}^k. \end{array}\right.$$

## REFERENCES


[1] W. B. Powell, *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. Wiley-Interscience, 2007.

[2] D. P. de Farias and B. Van Roy, "The linear programming approach to approximate dynamic programming," *Operations Research*, vol. 51, no. 6, pp. 850–865, 2003.

[3] D. P. De Farias and B. Van Roy, "On constraint sampling in the linear programming approach to approximate dynamic programming," *Mathematics of operations research*, vol. 29, no. 3, pp. 462–478, 2004.

[4] V. R. Konda and J. N. Tsitsiklis, "On actor-critic algorithms," *SIAM J CONTROL OPTIM*, vol. 42, no. 4, pp. 1143–1166, 2003.

[5] D. Bertsekas, "Dynamic programming and suboptimal control: A survey from adp to mpc," *European Journal of Control*, vol. 11, no. 45, pp. 310 – 334, 2005.

[6] V. Borkar, "A convex analytic approach to markov decision processes," *PROBAB THEORY REL*, vol. 78, no. 4, pp. 583–602, 1988.

[7] ——, *Handbook of Markov decision processes: methods and applications*, ser. International Series in Operations Research & Management Science, 40. Kluwer Academic Pub., 2002, ch. Convex Analytic Methods in Markov Decision Processes.

[8] O. Hernández-Lerma and J. Lasserre, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, ser. Applications of Mathematics Series. Springer, 1996.

[9] ——, *Further topics on discrete-time Markov control processes*, ser. Applications of Mathematics Series. Springer, 1999.

[10] ——, *Handbook of Markov decision processes: methods and applications*, ser. International Series in Operations Research & Management Science, 40. Kluwer Academic Pub., 2002, ch. The linear programming approach.

[11] O. Hernández-Lerma, J. González-Hernández, and R. López-Martínez, "Constrained average cost markov control processes in borel spaces," *SIAM J CONTROL OPTIM*, vol. 42, no. 2, pp. 442–468, 2003.

[12] F. Dufour and T. Prieto-Rumeau, "Finite linear programming approximations of constrained discounted markov decision processes," *SIAM J CONTROL OPTIM*, vol. 51, no. 2, pp. 1298–1324, 2013.

[13] E. Shafieepoorfard, M. Raginsky, and S. Meyn, "Rational inattention in controlled markov processes," in *American Control Conference (ACC), 2013*, June 2013, pp. 6790–6797.

[14] O. Hernández-Lerma and J. Lasserre, "Approximation schemes for infinite linear programs," *SIAM Journal on Optimization*, vol. 8, no. 4, pp. 973–988, 1998.

[15] A. Arapostathis, V. Borkar, E. Fernndez-Gaucherand, M. Ghosh, and S. Marcus, "Discrete-time controlled markov processes with average cost criterion: A survey," *SIAM Journal on Control and Optimization*, vol. 31, no. 2, pp. 282–344, 1993.

[16] D. Chatterjee, E. Cinquemani, and J. Lygeros, "Maximizing the probability of attaining a target prior to extinction," *Nonlinear Analysis: Hybrid Systems*, vol. 5, no. 2, pp. 367 – 381, 2011.

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

[18] E. Anderson and P. Nash, *Linear programming in infinite-dimensional spaces: theory and applications*, ser. Wiley-Interscience series in discrete mathematics and optimization. Wiley, 1987.

[19] 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 Transactions on Automatic Control, to appear*, 2014. [Online]. Available: http://dx.doi.org/10.1109/TAC.2014.2330702

[20] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[21] R. T. Rockafellar, *Convex analysis*. PRINCETON University Press, 1997.

[22] S. Ito, S.-Y. Wu, T.-J. Shiu, and K. L. Teo, "A numerical approach to infinite-dimensional linear programming in l1 spaces," *Journal of Industrial and Management Opt.*, vol. 6, no. 1, pp. 15–28, 2010.

[23] H. Lai and S. Wu, "Extremal points and optimal solutions for general capacity problems," *Mathematical Programming*, vol. 54, no. 1-3, pp. 87–113, 1992.

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

[25] A. Barvinok, *A Course in Convexity*, ser. Graduate studies in mathematics. American Mathematical Society, 2002.

[26] E. Feinberg and A. Shwartz, *Handbook of Markov decision processes: methods and applications*, ser. International Series in Operations Research & Management Science, 40. Kluwer Academic Pub., 2002.