Asymptotic Optimality and Rates of Convergence of Quantized Stationary Policies in Stochastic Control

Naci Saldi, Tamás Linder, and Serdar Yüksel

***Abstract*—We consider the discrete approximation of stationary policies for a discrete-time Markov decision process with Polish state and action spaces under total, discounted, and average cost criteria. Deterministic stationary quantizer policies are introduced and shown to be able to approximate optimal deterministic stationary policies with arbitrary precision under mild technical conditions, thus demonstrating that one can search for ε-optimal policies within the class of quantized control policies. We also derive explicit bounds on the approximation error in terms of the quantization rate.**

***Index Terms*—Approximation, Markov decision processes, quantization, stationary policies, stochastic control.**

I. Introduction

In the theory of Markov decision processes (MDPs), control policies induced by measurable mappings from the state space to the action space are called stationary. For a large class of infinite horizon optimization problems, the set of stationary policies is the smallest structured set of control policies in which one can find a globally optimal policy. However, computing an optimal policy even in this class is in general computationally prohibitive for non-finite Polish (that is, complete and separable metric) state and action spaces. Furthermore, in applications to networked control, the transmission of such control actions to an actuator is not realistic when there is an information transmission constraint (imposed by the presence of a communication channel) between a plant, a controller, or an actuator.

Hence, it is of interest to study the approximation of optimal stationary policies. Several approaches have been developed in the literature to tackle this problem, most of which assume finite or countable state spaces, see [2]–[4], [17]. In this technical note, we study the following question: for infinite Borel state and action spaces, how much is lost in performance if optimal policy is represented with a finite number of bits? This formulation appears to be new in the networked control literature, where stability properties of quantized control actions have been studied extensively, but the optimization of quantized control actions has not been studied as much in the context of cost minimization.

This technical note contains two main contributions: (i) We establish conditions under which quantized control policies are asymptotically optimal; that is, as the accuracy of quantization increases, the optimal cost is achieved as the limit of the cost of quantized policies. (ii) We establish rates of convergence under further conditions; that is, we obtain bounds on the approximation loss due to quantization. These findings are somewhat analogous to results in optimal quantization theory [21].

Organization: In Section II we review the definition of discrete time Markov decision processes (MDP) in our setting. In Section III-A we tackle the approximation problem for the total and discounted cost cases using strategic measures. In Section III-B an analogous approximation result is obtained for the average cost case using ergodic invariant probability measures of the induced Markov chains. In Section IV we derive quantitative bounds on the approximation error in terms of the rate of the approximating quantizers for both discounted and average costs.

Manuscript received April 26, 2014; revised July 14, 2014; accepted July 15, 2014. Date of publication July 28, 2014; date of current version January 21, 2015. This technical note was presented in part at the 51st Annual Allerton Conference on Communication, Control and Computing, Monticello, Illinois, October 2013. This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada. Recommended by Associate Editor P. Shi.

The authors are with the Department of Mathematics and Statistics, Queen's University, Kingston, ON, Canada (e-mail: nsaldi@mast.queensu.ca; linder@mast.queensu.ca; yuksel@mast.queensu.ca).

Digital Object Identifier 10.1109/TAC.2014.2343831

II. Markov Decision Processes

For a metric space E, let $\mathcal{B}(\mathsf{E})$ denote its Borel σ-algebra. Unless otherwise specified, the term "measurable" will refer to Borel measurability. We denote by $\mathcal{P}(\mathsf{E})$ the set of all probability measures on E.

Consider a discrete time Markov decision process (MDP) with *state space* X and *action space* A, where X and A are complete, separable metric (Polish) spaces equipped with their Borel σ-algebras $\mathcal{B}(\mathsf{X})$ and $\mathcal{B}(\mathsf{A})$, respectively. For all $x \in \mathsf{X}$, we assume that the *set of admissible actions* is A. Let the *stochastic kernel* $p(\cdot|x,a)$ denote the *transition probability* of the next state given that previous state-action pair is (x,a) [6]. The probability measure μ over X denotes the initial distribution.

Define the history spaces $\mathsf{H}_n = (\mathsf{X} \times \mathsf{A})^n \times \mathsf{X},\ n = 0,1,2,\ldots$ endowed with their product Borel σ-algebras generated by $\mathcal{B}(\mathsf{X})$ and $\mathcal{B}(\mathsf{A})$. A *policy* is a sequence $\pi = \{\pi_n\}_{n\geq 0}$ of stochastic kernels on A given H_n. A policy π is said to be *deterministic* if the stochastic kernels π_n are realized by a sequence of measurable functions $\{f_n\}$ from H_n to A, i.e., $\pi_n(\cdot|h_n) = \delta_{f_n(h_n)}(\cdot)$ where $f_n : \mathsf{H}_n \to \mathsf{A}$ is measurable. A policy π is called *stationary* if the stochastic kernels π_n depend only on the current state; that is, $\pi_n = \pi_m\ (m,n \geq 0)$ and π_n is a stochastic kernel on A given X. A policy π that is both deterministic and stationary is called *deterministic stationary*. Hence, deterministic stationary policies are defined by a measurable function $f : \mathsf{X} \to \mathsf{A}$. We denote by S the set of deterministic stationary policies.

According to the Ionescu Tulcea theorem [6], an initial distribution μ on X and a policy π define a unique probability measure P_μ^π on $\mathsf{H}_\infty = (\mathsf{X} \times \mathsf{A})^\infty$, which is called a *strategic measure* [5]. The expectation with respect to P_μ^π is denoted by E_μ^π. If $\mu = \delta_x$ for some $x \in \mathsf{X}$, we write P_x^π and E_x^π instead of $P_{\delta_x}^\pi$ and $E_{\delta_x}^\pi$, respectively.

Let c and c_n, $n = 0,1,2,\ldots$, be measurable functions from $\mathsf{X} \times \mathsf{A}$ to $[0,\infty)$. The cost functions w considered in this technical note are *expected total cost* i.e., $w_t(\pi,\mu) := E_\mu^\pi[\sum_{n=0}^\infty c_n(x_n,a_n)]$, *expected discounted cost* i.e., $w_\beta(\pi,\mu) := E_\mu^\pi[\sum_{n=0}^\infty \beta^n c(x_n,a_n)]$ for some $\beta \in (0,1)$, and *expected average cost*, i.e., $w_A(\pi,\mu) := \limsup_{N\to\infty}(1/N)E_\mu^\pi[\sum_{n=0}^{N-1} c(x_n,a_n)]$. Note that the expected discounted cost is a special case of the expected total cost.

We write $w(\pi,\mu)$ to denote the cost function (either i), ii), or iii)) of the policy π for the initial distribution μ. If $\mu = \delta_x$, we write $w(\pi,x)$ instead of $w(\pi,\delta_x)$. A policy π^* is called optimal if

$w(\pi^*,\mu)=\inf_\pi w(\pi,\mu)$ for all $\mu\in\mathcal{P}(\mathsf{X})$. It is well known that the set of deterministic stationary policies contains optimal policies for a large class of infinite horizon discounted cost problems (see, e.g., [6], [14]) and average cost optimal control problems (see, e.g., [1], [14]).

Throughout the technical note, the initial distribution μ is assumed to be an arbitrary fixed distribution unless otherwise specified.

A. Notation and Conventions

The set of all bounded measurable real functions and bounded continuous real functions on a metric space E are denoted by $B(\mathsf{E})$ and $C_b(\mathsf{E})$, respectively. For any $\nu\in\mathcal{P}(\mathsf{E})$ and measurable real function g on E, define $\nu(g):=\int g d\nu$. Let $\mathsf{E}_n=\prod_{i=1}^n \mathsf{E}_i$ $(2\le n\le\infty)$ be a finite or a infinite product space. By an abuse of notation, any function g on $\prod_{j=i_1}^{i_m}\mathsf{E}_j$, where $\{i_1,\dots,i_m\}\subseteq\{1,\dots,n\}$ $(m\le n)$, is also treated as a function on E_n by identifying it with its natural extension to E_n. For any π and initial distribution μ, let $\lambda_n^{\pi,\mu}$, $\lambda_{(n)}^{\pi,\mu}$, and $\gamma_n^{\pi,\mu}$, respectively, denote the law of x_n, $(x_0,\dots,x_n)$ and (x_n,a_n) for all $n\ge 0$. Hence, for instance, we may write $\lambda_{(n+1)}^{\pi,\mu}(h)=\lambda_{(n)}^{\pi,\mu}(\lambda_{(1)}^{\pi,x_n}(h))$ where $h\in B(\mathsf{X}^{n+2})$. Let $\mathbb{F}$ denote the set of all measurable functions from X to A. For any $g\in B(\mathsf{H}_n)$ $(n\ge 1)$ and $f\in\mathbb{F}$, define $g_f(x_0,\dots,x_n):=g(x_0,f(x_0),\dots,f(x_{n-1}),x_n)$. Hence, when $c\in B(\mathsf{X}\times\mathsf{A})$, $c_f(x_n)=c(x_n,f(x_n))$ since $c\in B(\mathsf{H}_{n+1})$ by our conventions.

B. Problem Formulation

In this section we give a formal definition of the problems considered in this technical note. To this end, we first give the definition of a quantizer.

Definition 2.1: A measurable function $q:\mathsf{X}\to\mathsf{A}$ is called a *quantizer* from X to A if the range of q, i.e., $q(\mathsf{X})=\{q(x)\in\mathsf{A}:x\in\mathsf{X}\}$, is finite.

The elements of $q(\mathsf{X})$ (i.e., the possible values of q) are called the *levels* of q. The rate R of a quantizer q is defined as the logarithm of the number of its levels: $R=\log_2|q(\mathsf{X})|$. Note that R (approximately) represents the number of bits needed to losslessly encode the output levels of q using binary codewords of equal length. Let $\mathcal{Q}$ denote the set of all quantizers from X to A. In this technical note we introduce a new type of policy called a *deterministic stationary quantizer policy*. Such a policy is a constant sequence $\pi=\{\pi_n\}$ of stochastic kernels on A given X such that $\pi_n(\cdot|x)=\delta_{q(x)}(\cdot)$ for all n for some $q\in\mathcal{Q}$. For any finite set $\Lambda\subset\mathsf{A}$, let $\mathcal{Q}(\Lambda)$ denote the set of all quantizers having range Λ and let $S\mathcal{Q}(\Lambda)$ denote the set of all deterministic stationary quantizer policies induced by $\mathcal{Q}(\Lambda)$.

The principal goal in this technical note is to determine conditions such that there exists a sequence of finite subsets $\{\Lambda_k\}_{k\ge1}$ of A for which the following statements hold:

(**P1**) For any $\pi\in S$ there exists an approximating sequence $\{\pi^k\}$ satisfying $\lim_{k\to\infty}w(\pi^k,\mu)=w(\pi,\mu)$, where $\pi^k\in S\mathcal{Q}(\Lambda_k)$ $(k\ge1)$.

(**P2**) For any $\pi\in S$ the approximating sequence $\{\pi^k\}$ in (**P1**) is such that $|w(\pi,\mu)-w(\pi^k,\mu)|$ can be explicitly upper bounded by a term depending on the cardinality of Λ_k.

Thus (**P1**) implies the existence of a sequence of stationary quantizer policies converging to an optimal stationary policy, while (**P2**) implies that the approximation error can be explicitly controlled.

III. Approximation of Deterministic Stationary Policies

A sequence $\{\mu_n\}$ of measures on a measurable space $(\mathsf{E},\mathcal{E})$ is said to converge setwise [7] to a measure μ if $\mu_n(B)\to\mu(B)$ for all $B\in\mathcal{E}$, or equivalently, $\mu_n(g)\to\mu(g)$ for all $g\in B(\mathsf{E})$. In this section, we will impose the following assumptions:

(a) The stochastic kernel $p(\cdot|x,a)$ is setwise continuous in $a\in\mathsf{A}$, i.e., if $a_n\to a$, then $p(\cdot|x,a_n)\to p(\cdot|x,a)$ setwise for all $x\in\mathsf{X}$.

(b) A is compact.

Remark 3.1: Note that if X is countable, then $B(\mathsf{X})=C_b(\mathsf{X})$ which implies the equivalence of setwise convergence and weak convergence. Hence, results developed in this technical note are applicable to the MDPs having weakly continuous, in the action variable, transition probabilities when the state space is countable.

Remark 3.2: Note that any MDP can be modeled by a discrete time dynamical system of the form $x_{n+1}=F(x_n,a_n,v_n)$, where the v_n's are independent and identically distributed (i.i.d.) random variables with values in some space V and common distribution ν. In many applications, the function F has a well behaved structure and is in the form $F(x,a,v)=H(x,a)G(v)$ or $F(x,a,v)=H(x,a)+G(v)$, e.g., the *fisheries management model* [6, p. 5], the *cash balance model* [15], and the *Pacific halibut fisheries management model* [18]. In these systems, assumption (a) holds for common noise processes. For instance, if ν admits a continuous density, which is often the case in practice, then assumption (a) usually holds. We refer the reader to [15, Section 4] for a discussion on the relevance of the setwise continuity assumption on inventory control problems. In addition, the widely studied and practically important case of the additive noise system in our Example 3.1 in the next section also satisfies assumption (a).

We now define the ws^∞ topology on $\mathcal{P}(\mathsf{H}_\infty)$ which was first introduced by Schäl in [8]. Let $\mathcal{C}(\mathsf{H}_0)=B(\mathsf{X})$ and let $\mathcal{C}(\mathsf{H}_n)$ $(n\ge1)$ be the set of real valued functions g on H_n such that $g\in B(\mathsf{H}_n)$ and $g(x_0,\cdot,x_1,\cdot,\dots,x_{n-1},\cdot,x_n)\in C_b(\mathsf{A}^n)$ for all $(x_0,\dots,x_n)\in\mathsf{X}^{n+1}$. The ws^∞ topology on $\mathcal{P}(\mathsf{H}_\infty)$ is defined as the smallest topology which renders all mappings $P\mapsto P(g)$, $g\in\bigcup_{n=0}^\infty\mathcal{C}(\mathsf{H}_n)$, continuous.

Let d_{A} denote the metric on A. Since the action space A is compact and thus totally bounded, one can find a sequence of finite sets $(\{a_i\}_{i=1}^{m_k})_{k\ge1}$ such that for all k, $\min_{i\in\{1,\dots,m_k\}}d_{\mathsf{A}}(a,a_i)<1/k$ for all $a\in\mathsf{A}$. In other words, $\{a_i\}_{i=1}^{m_k}$ is a $1/k$-net in A. Let $\Lambda_k:=\{a_1,\dots,a_{m_k}\}$ and for any $f\in\mathbb{F}$ define the sequence $\{q_k\}$ by letting

$$q_k(x):=\arg\min_{a\in\Lambda_k}d_{\mathsf{A}}(f(x),a) \tag{1}$$

where ties are broken so that q_k are measurable. Note that, $q_k\in\mathcal{Q}(\Lambda_k)$ for all k and q_k converges *uniformly* to f as $k\to\infty$. Let $\pi\in S$ and $\pi^k\in S\mathcal{Q}(\Lambda_k)$ be induced by f and q_k, respectively. We call each π_k a *quantized approximation* of π. In the rest of this technical note, we assume that the sequence $\{\Lambda_k\}$, as defined above, is fixed.

A. Expected Total and Discounted Costs

Recall that w_t and w_β denote the expected total and discounted costs, respectively. We impose the following assumptions in addition to assumptions (a) and (b):

(c) c and c_n $(n\ge1)$ are non-negative, bounded functions satisfying $c(x,\cdot),c_n(x,\cdot)\in C_b(\mathsf{A})$ for all $x\in\mathsf{X}$.

(d) $\sup_{\tilde\pi\in S}\sum_{n=N+1}^\infty\gamma_n^{\tilde\pi,\mu}(c_n)\to0$ as $N\to\infty$.

Remark 3.3: We note that all the results in this technical note remain valid if it is only assumed that c and c_n $(n\ge0)$ are bounded and satisfies $c(x,\cdot),c_n(x,\cdot)\in C_b(\mathsf{A})$ for all $x\in\mathsf{X}$.

Since the one-stage cost functions c_n are non-negative, assumption (d) is equivalent to Condition (C) in [8, pg. 349]. Clearly, the expected

discounted cost satisfies assumption (d) under assumption (c). We now state our main theorem in this subsection.

Theorem 3.1: Suppose assumptions (a), (b), (c) hold. Let $\pi \in S$ and $\{\pi^k\}$ be the quantized approximations of π. Then, $w_\beta(\pi^k, \mu) \to w_\beta(\pi, \mu)$ as $k \to \infty$. The same statement is true for w_t if we further impose assumption (d).

The proof of Theorem 3.1 requires the following proposition which is proved in Appendix V-A.

Proposition 3.1: Suppose assumptions (a) and (b) hold. Then for any $\pi \in S$, the strategic measures $\{P_\mu^{\pi^k}\}$ induced by the quantized approximations $\{\pi^k\}$ of π converge to the strategic measure P_μ^π of π in the ws^∞ topology. Hence, $\gamma_n^{\pi^k,\mu}(c_n) \to \gamma_n^{\pi,\mu}(c_n)$ as $k \to \infty$ under assumption (c).

Proof of Theorem 3.1: Since w_β is a special case of w_t and satisfies (d) under assumption (c), it is enough to prove the theorem for w_t. By Proposition 3.1, $\gamma_n^{\pi^k,\mu}(c_n) \to \gamma_n^{\pi,\mu}(c_n)$ as $k \to \infty$ for all n. Then, we have

$$\begin{aligned}
&\limsup_{k\to\infty} \left|w_t(\pi^k,\mu) - w_t(\pi,\mu)\right| \\
&\le \limsup_{k\to\infty} \sum_{n=0}^{\infty} \left|\gamma_n^{\pi^k,\mu}(c_n) - \gamma_n^{\pi,\mu}(c_n)\right| \\
&\le \lim_{k\to\infty} \sum_{n=0}^{N} \left|\gamma_n^{\pi^k,\mu}(c_n) - \gamma_n^{\pi,\mu}(c_n)\right| + 2\sup_{\tilde\pi\in S} \sum_{n=N+1}^{\infty} \gamma_n^{\tilde\pi,\mu}(c_n). \quad (2)
\end{aligned}$$

Since the first and second terms in the last expression converge to zero as $N \to \infty$ by Proposition 3.1 and assumption (d), respectively, the proof is complete. ■

Remark 3.4: Notice that this proof implicitly shows that w_t and w_β are sequentially continuous with respect to the strategic measures in the ws^∞ topology.

The following is a generic example frequently considered in the theory of Markov decision processes (see [12]).

Example 3.1: Let us consider an additive-noise system given by

$$x_{n+1} = F(x_n, a_n) + v_n,\ n = 0, 1, 2, \ldots$$

where $\mathsf{X} = \mathbb{R}^n$ and the v_n's are independent and identically distributed (i.i.d.) random vectors whose common distribution has a continuous, bounded, and strictly positive probability density function. A non-degenerate Gaussian distribution satisfies this condition. We assume that the action space A is a compact subset of $\mathbb{R}^d$ for some $d \ge 1$, the one stage cost functions c and c_n $(n \ge 1)$ satisfy assumption (c), and $F(x, \cdot)$ is continuous for all $x \in \mathsf{X}$. It is straightforward to show that assumption (a) holds under these conditions. Hence, under assumption (d) on the cost functions c_n, Theorem 3.1 holds for this system.

B. Expected Average Cost

We are still assuming (a), (b), and (c). In contrast to the expected total and discounted cost criteria, the expected average cost is in general not sequentially continuous with respect to strategic measures for the ws^∞ topology under practical assumptions. Instead, we develop an approach based on the convergence of the sequence of invariant probability measures under quantized stationary policies.

Recall that w_A denotes the expected average cost. Observe that any deterministic stationary policy π, induced by f, defines a stochastic kernel on X given X via

$$Q_\pi(\cdot|x) := \lambda_1^{\pi,x}(\cdot) = p\left(\cdot|x, f(x)\right). \quad (3)$$

Let us write $Q_\pi g(x) := \lambda_1^{\pi,x}(g)$. If Q_π admits an ergodic invariant probability measure ν_π, then by Theorem 2.3.4 and Proposition 2.4.2 in [7], there exists an invariant set with full ν_π measure such that for all x in that set we have

$$\begin{aligned}
w_A(\pi, x) &= \limsup_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \gamma_n^{\pi,\mu}(c) \\
&= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \lambda_n^{\pi,x}(c_f) = \nu_\pi(c_f). \quad (4)
\end{aligned}$$

Let $\mathsf{M}_\pi \in \mathcal{B}(\mathsf{X})$ be the set of all $x \in \mathsf{X}$ such that convergence in (4) holds. Hence, $\nu_\pi(\mathsf{M}_\pi) = 1$ if ν_π exists. The following assumptions will be imposed in the main theorem of this section.

(e) For any $\pi \in S$, Q_π has a unique invariant probability measure ν_π.
(f1) The set $\Gamma_S := \{\nu \in \mathcal{P}(\mathsf{X}) : \nu Q_\pi = \nu \text{ for some } \pi \in S\}$ is relatively sequentially compact in the setwise topology.
(f2) There exists $x \in \mathsf{X}$ such that for all $B \in \mathcal{B}(\mathsf{X})$, $\lambda_n^{\pi,x}(B) \to \nu_\pi(B)$ uniformly in $\pi \in S$.
(g) $\mathsf{M} := \bigcap_{\pi\in S} \mathsf{M}_\pi \neq \emptyset$.

Theorem 3.2: Let the initial distribution μ be concentrated on some $x \in \mathsf{M}$. Let $\pi \in S$ and $\{\pi^k\}$ be the quantized approximations of π. Then, $w_A(\pi^k, \mu) \to w_A(\pi, \mu)$ under the assumptions (e), (f1) or (f2), and (g).

Proof: See Appendix V-B. ■

In the rest of this section we will derive conditions under which assumptions (e), (f1), (f2), and (g) hold. To begin with, assumptions (e), (f2), and (g) are satisfied under any of the conditions Ri, $i \in \{0, 1, 1(a), 1(b), 2, \ldots, 6\}$ in [10]. Moreover, $\mathsf{M} = \mathsf{X}$ in (g) if at least one of the above conditions holds. The next step is to find sufficient conditions for assumptions (e), (f1), and (g) to hold.

Observe that the stochastic kernel p on X given $\mathsf{X} \times \mathsf{A}$ can be written as a measurable mapping from $\mathsf{X} \times \mathsf{A}$ to $\mathcal{P}(\mathsf{X})$ if $\mathcal{P}(\mathsf{X})$ is equipped with its Borel σ-algebra generated by the weak topology [7], i.e., $p(\cdot|x, a) : \mathsf{X} \times \mathsf{A} \to \mathcal{P}(\mathsf{X})$. We impose the following assumption:

(e1) $p(\cdot|x, a) \le \zeta(\cdot)$ for all $x \in \mathsf{X}$, $a \in \mathsf{A}$ for some finite measure ζ on X.

Proposition 3.2: Suppose (e1) holds. Then, for any $\pi \in S$ induced by f, Q_π has an invariant probability measure ν_π. Furthermore, Γ_S is sequentially relatively compact in the setwise topology. Hence, (e1) implies assumption (f1). In addition, if these invariant measures are unique, then assumptions (e) and (g) also hold with $\mathsf{M} = \mathsf{X}$ in (g).

Proof: For any $\pi \in S$, define $Q_{\pi,x}^{(N)}(\cdot) := (1/N) \sum_{n=0}^{N-1} \lambda_n^{\pi,x}(\cdot)$ for some $x \in \mathsf{X}$. Clearly, $Q_{\pi,x}^{(N)} \le \zeta$ for all N. Hence, by [7, Corollary 1.4.5] there exists a subsequence $\{Q_{\pi,x}^{(N_k)}\}$ which converges to some probability measure ν_π setwise. Following the same steps in [11, Theorem 4.17] one can show that $\nu_\pi(g) = \nu_\pi(Q_\pi g)$, for all $g \in B(\mathsf{X})$. Hence, ν_π is an invariant probability measure for Q_π.

Furthermore, assumption (e1) implies $\nu_\pi \le \zeta$ for all $\nu_\pi \in \Gamma_s$. Thus, Γ_s is relatively sequentially compact in the setwise topology by again [7, Corollary 1.4.5].

Finally, for any π, if the invariant measure ν_π is unique, then every setwise convergent subsequence of the relatively sequentially compact sequence $\{Q_{\pi,x}^{(N)}\}$ must converge to ν_π. Hence, $Q_{\pi,x}^{(N)} \to \nu_\pi$ setwise which implies that $w_A(\pi, x) = \limsup_{N\to\infty} Q_{\pi,x}^{(N)}(c_f) = \lim_{N\to\infty} Q_{\pi,x}^{(N)}(c_f) = \nu_\pi(c_f)$ for all $x \in \mathsf{X}$ since $c_f \in B(\mathsf{X})$. Thus, $\mathsf{M} = \mathsf{X}$ in (g). ■

Example 3.2: Let us consider an additive-noise system in Example 3.1 with the same assumptions. Furthermore, we assume F is bounded. Observe that for any $\pi \in S$, if Q_π has an invariant probability measure, then it has to be unique [7, Lemma 2.2.3] since there cannot exist disjoint invariant sets due to the positivity of probability density function. Since this system satisfies (e1) and $R1(a)$ in [10]

due to the boundedness of F, assumptions (e), (f1), (f2), and (g) hold with $\mathsf{M} = \mathsf{X}$. This means that Theorem 3.2 holds for an additive noise system under the above conditions.

IV. Rates of Convergence

Let $\|\cdot\|_{TV}$ [7] denote the total variation distance between measures. We will impose a new set of assumptions in this section:

(h) A is an infinite compact subset of $\mathbb{R}^d$ for some $d \geq 1$.

(j) c is bounded and $|c(x,\tilde{a}) - c(x,a)| \leq K_1 d_{\mathsf{A}}(\tilde{a},a)$ for all x, and some $K_1 \geq 0$.

(k) $\|p(\cdot\,|x,\tilde{a}) - p(\cdot\,|x,a)\|_{TV} \leq K_2 d_{\mathsf{A}}(\tilde{a},a)$ for all x, and some $K_2 \geq 0$.

(l) There exists positive constants C and $\kappa \in (0,1)$ such that for all $\pi \in S$, there is a (necessarily unique) probability measure $\nu_\pi \in \mathcal{P}(\mathsf{X})$ satisfying $\|\lambda_n^{\pi,x} - \nu_\pi\|_{TV} \leq C\kappa^n$ for all $x \in \mathsf{X}$ and $n \geq 1$.

Assumption (l) implies that for any policy $\pi \in S$, the stochastic kernel Q_π, defined in (3), has a unique invariant probability measure ν_π and satisfies *geometric ergodicity* [7]. Note that (l) holds under any of the conditions Ri, $i \in \{0,1,1(a),1(b),2,\ldots,5\}$ in [10]. Moreover, one can explicitly compute the constants C and κ for certain systems. For instance, consider an additive-noise system in Example 3.1 with Gaussian noise. Let $\mathsf{X} = \mathbb{R}$. Assume F has a bounded range so that $F(\mathbb{R}) \subset [-L,L]$ for some $L > 0$. Then, assumption (l) holds with $C = 2$ and $\kappa = 1 - \varepsilon L$, where $\varepsilon = (1/\sigma\sqrt{2\pi})\exp^{-(2L)^2/2\sigma^2}$. For further conditions that imply (l) we refer the reader to [7], [10].

The following example gives the sufficient conditions for the additive noise system under which (j), (k), and (l) hold.

Example 4.3: Consider the additive-noise system in Example 3.1. In addition to the assumptions there, suppose $F(x,\cdot)$ is Lipschitz uniformly in $x \in \mathsf{X}$ and the common density g of the v_n is Lipschitz on all compact subsets of X. Note that a Gaussian density has these properties. Let $c(x,a) := \|x-a\|^2$. Under these conditions, assumptions (j) and (k) hold for the additive noise system. If we further assume that F is bounded, then assumption (l) holds as well.

The following result is a consequence of the fact that if A is a compact subset of $\mathbb{R}^d$ then there exist a constant $\alpha > 0$ and finite subsets $\Lambda_k \subset \mathsf{A}$ with cardinality $|\Lambda_k| = k$ such that $\max_{x\in\mathsf{A}} \min_{y\in\Lambda_k} d_{\mathsf{A}}(x,y) \leq \alpha(1/k)^{1/d}$ for all k, where d_{A} is the Euclidean distance on A inherited from $\mathbb{R}^d$.

Lemma 4.1: Let $\mathsf{A} \subset \mathbb{R}^d$ be compact. Then for any measurable function $f : \mathsf{X} \to \mathsf{A}$ we can construct a sequence of quantizers $\{q_k\}$ from X to A which satisfy $\sup_{x\in\mathsf{X}} d_{\mathsf{A}}(q_k(x), f(x)) \leq \alpha(1/k)^{1/d}$ for some constant α.

The following proposition is the key result in this section. It is proved in Appendix V-C.

Proposition 4.3: Let $\pi \in S$ and $\{\pi^k\}$ be the quantized approximations of π. For any initial distribution μ we have

$$\left\|\lambda_n^{\pi,\mu} - \lambda_n^{\pi^k,\mu}\right\|_{TV} \leq \alpha K_2 (2n-1)\left(\frac{1}{k}\right)^{\frac{1}{d}} \tag{5}$$

for all $n \geq 1$ under assumptions (h), (j), and (k).

A. Expected Discounted Cost

The proof of the following theorem essentially follows from Proposition 4.3. The proof is given in Appendix V-D.

Theorem 4.1: Let $\pi \in S$ and $\{\pi^k\}$ be the quantized approximations of π. For any initial distribution μ, we have

$$\left|w_\beta(\pi,\mu) - w_\beta(\pi^k,\mu)\right| \leq K\left(\frac{1}{k}\right)^{\frac{1}{d}} \tag{6}$$

where $K = \alpha/(1-\beta)(K_1 - \beta K_2 M + (2\beta M K_2/(1-\beta)))$ with $M := \sup_{(x,a)\in\mathsf{X}\times\mathsf{A}} |c(x,a)|$ under assumptions (h), (j), and (k).

B. Expected Average Cost

Note that for any $\pi \in S$, induced by f, assumption (l) implies that ν_π is a unique invariant probability measure for Q_π and that $w_A(\pi,x) = \nu_\pi(c_f)$ for all x when c is as in the assumption (c). The following theorem basically follows from Proposition 4.3 and assumption (l). It is proved in Appendix V-E.

Theorem 4.2: Let $\pi \in S$ and $\{\pi^k\}$ be the quantized approximations of π. Under assumptions (h), (j), (k), and (l), for any $x \in \mathsf{X}$ we have

$$\left|w_A(\pi,x) - w_A(\pi^k,x)\right| \leq 2MC\kappa^n + K_n\left(\frac{1}{k}\right)^{\frac{1}{d}} \tag{7}$$

for all $n \geq 0$, where $K_n = ((2n-1)K_2\alpha M + K_1\alpha)$ and $M := \sup_{(x,a)\in\mathsf{X}\times\mathsf{A}} |c(x,a)|$.

Observe that depending on the values of C and κ, we can first make the first term in the upper bound small enough by choosing sufficiently large n, and then for this n we can choose k large enough such that the second term in the upper bound is small.

Order Optimality: The following example demonstrates that the order of approximation errors in Theorems 4.2 and 4.1 cannot be better than $O((1/k)^{1/d})$. More precisely, we exhibit a simple standard example where we can lower bound the approximation errors for the optimal stationary policy by $L(1/k)^{1/d}$, for some positive constant L.

In what follows $h(\cdot)$ and $h(\cdot|\cdot)$ denote differential and conditional differential entropies, respectively [19].

Example 4.4: Consider the linear system

$$x_{n+1} = Ax_n + Ba_n + v_n,\ n = 0,1,2,\ldots$$

where $\mathsf{X} = \mathsf{A} = \mathbb{R}^d$ and the v_n's are i.i.d. random vectors whose common distribution has density g. For simplicity suppose that the initial distribution μ has the same density g. It is assumed that the differential entropy $h(g) := -\int_{\mathsf{X}} g(x)\log g(x)dx$ is finite. Let the one stage cost function be $c(x,a) := \|x-a\|$. Clearly, the optimal stationary policy π^* is induced by the identity $f(x) = x$, having the optimal cost $w_i(\pi,\mu) = 0$, where $i \in \{\beta, A\}$. Let $\{\pi^k\}$ be the quantized approximations of π^*. Fix any k and define $D_n := E_\mu^{\pi^k}[c(x_n,a_n)]$ for all n. Then, by the Shannon lower bound (SLB) [20, p. 12] we have for $n \geq 1$

$$\begin{aligned}
\log k \geq R(D_n) &\geq h(x_n) + \theta(D_n) \\
&= h(Ax_{n-1} + Ba_{n-1} + v_{n-1}) + \theta(D_n) \\
&\geq h(Ax_{n-1} + Ba_{n-1} + v_{n-1}|x_{n-1},a_{n-1}) + \theta(D_n) \\
&= h(v_{n-1}) + \theta(D_n)
\end{aligned} \tag{8}$$

where $\theta(D_n) = -d + \log((1/(dV_d\Gamma(d))))(d/D_n)^d)$, $R(D_n)$ is the rate-distortion function of x_n, V_d is the volume of the unit sphere $S_d = \{x : \|x\| \leq 1\}$, and Γ is the gamma function. Here, (8) follows from the independence of v_{n-1} and the pair (x_{n-1},a_{n-1}). Note that $h(v_{n-1}) = h(g)$ for all n. Hence, we obtain $D_n \geq L(1/k)^{1/d}$, where $L := (d/2)(2^{h(g)}/(dV_d\Gamma(d)))^{1/d}$. This gives $|w_\beta(\pi^*,\mu) - w_\beta(\pi^k,\mu)| \geq L/(1-\beta)(1/k)^{1/d}$ and $|w_A(\pi^*,\mu) - w_A(\pi^k,\mu)| \geq L(1/k)^{1/d}$.

V. Discussion

Motivated by the fact that deterministic stationary policies may not be optimal in constrained MDPs even for the discounted cost (see, e.g., [16]), these approximation results are extended to randomized policies in [22]. One direction for future work is to establish similar results for approximations where the set of admissible quantizers has a certain structure, such as the set of quantizers having convex codecells [13],

which may give rise to practical design methods. As a final remark, since setwise continuity assumption might be too restrictive in certain important cases, it is of interest to study a version of this problem where the setwise continuity assumption is replaced with the weak continuity in the state-action variables.

APPENDIX

A. Proof of Proposition 3.1

Suppose $g \in \mathcal{C}(\mathsf{H}_n)$ for some n. Then we have $P_\mu^{\pi^k}(g) = \lambda_{(n)}^{\pi^k,\mu}(g_{q_k})$ and $P_\mu^{\pi}(g) = \lambda_{(n)}^{\pi,\mu}(g_f)$. Since g is continuous in the "a" terms by definition and q_k converges to f, we have $g_{q_k} \to g_f$. Hence, by [9, Theorem 2.4] it is enough to prove that $\lambda_{(n)}^{\pi^k,\mu} \to \lambda_{(n)}^{\pi,\mu}$ setwise as $k \to \infty$.

We will prove this by induction. Clearly, $\lambda_{(1)}^{\pi^k,\mu} \to \lambda_{(1)}^{\pi,\mu}$ setwise by assumption (a). Assume the claim is true for some $n \geq 1$. For any $h \in B(\mathsf{X}^{n+2})$ we can write $\lambda_{(n+1)}^{\pi^k,\mu}(h) = \lambda_{(n)}^{\pi^k,\mu}(\lambda_{(1)}^{\pi^k,x_n}(h))$ and $\lambda_{(n+1)}^{\pi,\mu}(h) = \lambda_{(n)}^{\pi,\mu}(\lambda_{(1)}^{\pi,x_n}(h))$. Since $\lambda_{(1)}^{\pi^k,x_n}(h) \to \lambda_{(1)}^{\pi,x_n}(h)$ for all $(x_0,\ldots,x_n) \in \mathsf{X}^{n+1}$ by assumption (a) and $\lambda_{(n)}^{\pi^k,\mu} \to \lambda_{(n)}^{\pi,\mu}$ setwise, we have $\lambda_{(n+1)}^{\pi^k,\mu}(h) \to \lambda_{(n+1)}^{\pi,\mu}(h)$ by [9, Theorem 2.4] which completes the proof.

B. Proof of Theorem 3.2

Let Q_π and Q_{π^k} be the stochastic kernels, respectively, for π and $\{\pi^k\}$ defined in (3). By assumption (e), Q_π and Q_{π^k} $(k \geq 1)$ have unique, and so ergodic, invariant probability measures ν_π and ν_{π^k}, respectively. Since μ is concentrated on some $x \in \mathsf{M}$, we have $w_A(\pi^k,\mu) = \nu_{\pi^k}(c_{q_k})$ and $w_A(\pi,\mu) = \nu_\pi(c_f)$. Observe that $c_{q_k}(x) \to c_f(x)$ for all x by assumption (c). Hence, if we prove $\nu_{\pi^k} \to \nu_\pi$ setwise, then by [9, Theorem 2.4] we have $J(\pi^k,\mu) \to J(\pi,\mu)$. We prove this first under (f1) and then under (f2).

1) Proof Under Assumption (f1): We show that every setwise convergent subsequence $\{\nu_{\pi^{k_l}}\}$ of $\{\nu_{\pi^k}\}$ must converge to ν_π. Then, since Γ_s is relatively sequentially compact in the setwise topology, there is at least one setwise convergent subsequence $\{\nu_{\pi^{k_l}}\}$ of $\{\nu_{\pi^k}\}$, which implies the result.

Let $\nu_{\pi^{k_l}} \to \nu$ setwise for some $\nu \in \mathcal{P}(\mathsf{X})$. We will show that $\nu = \nu_\pi$ or equivalently ν is an invariant probability measure of Q_π. For simplicity, we write $\{\nu_{\pi^l}\}$ instead of $\{\nu_{\pi^{k_l}}\}$. Let $g \in B(\mathsf{X})$. Then by assumption (e) we have $\nu_{\pi^l}(g) = \nu_{\pi^l}(Q_{\pi^l}g)$. Since $Q_{\pi^l}g(x) \to Q_\pi g(x)$ for all x by assumption (a) and $\nu_{\pi^l} \to \nu$ setwise, we have $\nu_{\pi^l}(Q_{\pi^l}g) \to \nu_\pi(Q_\pi g)$ by [9, Theorem 2.4]. On the other hand since $\nu_{\pi^l} \to \nu$ setwise we have $\nu_{\pi^l}(g) \to \nu(g)$. Thus $\nu(g) = \nu(Q_\pi g)$. Since g is arbitrary, ν is an invariant probability measure for Q_π.

2) Proof Under Assumption (f2): Observe that for all $x \in \mathsf{X}$ and n, $\lambda_n^{\pi^k,x} \to \lambda_n^{\pi,x}$ setwise as $k \to \infty$ since $P_x^{\pi^k} \to P_x^\pi$ in the ws^∞ topology (see Proposition 3.1). Let $B \in \mathcal{B}(\mathsf{X})$ be given and fix some $\varepsilon > 0$. By assumption (f2) we can choose N large enough such that $|\lambda_N^{\tilde{\pi},x}(B) - \nu_{\tilde{\pi}}(B)| < \varepsilon/3$ for all $\tilde{\pi} \in \{\pi, \pi^1, \pi^2, \cdots\}$. For this N, choose K large enough such that $|\lambda_N^{\pi^k,x}(B) - \lambda_N^{\pi,x}(B)| < \varepsilon/3$ for all $k \geq K$. Thus, for all $k \geq K$ we have $|\nu_{\pi^k}(B) - \nu_\pi(B)| \leq |\nu_{\pi^k}(B) - \lambda_N^{\pi^k,x}(B)| + |\lambda_N^{\pi^k,x}(B) - \lambda_N^{\pi,x}(B)| + |\lambda_N^{\pi,x}(B) - \nu_\pi(B)| < \varepsilon$. Since ε is arbitrary, we obtain $\nu_{\pi^k}(B) \to \nu_\pi(B)$, which completes the proof.

C. Proof of Proposition 4.3

We will prove this result by induction. Let μ be an arbitrary initial distribution and fix k. For $n = 1$ the claim holds by the following argument:

$$
\begin{aligned}
&\left\|\lambda_1^{\pi,\mu} - \lambda_1^{\pi^k,\mu}\right\|_{TV} \\
&= 2 \sup_{B \in \mathcal{B}(\mathsf{X})} \left|\mu\left(\lambda_1^{\pi,x}(B)\right) - \mu\left(\lambda_1^{\pi^k,x}(B)\right)\right| \\
&\leq \mu\left(\left\|\lambda_1^{\pi,x} - \lambda_1^{\pi^k,x}\right\|_{TV}\right) \\
&\leq \mu\left(K_2 d_{\mathsf{A}}\left(f(x), q_k(x)\right)\right) \text{ (by assumption(k))} \\
&\leq \sup_{x \in \mathsf{X}} K_2 d_{\mathsf{A}}\left(f(x), q_k(x)\right) \leq \left(\frac{1}{k}\right)^{\frac{1}{d}} K_2 \alpha \text{ (by Lemma 4.1).}
\end{aligned}
$$

Observe that the bound $\alpha K_2(2n-1)(1/k)^{1/d}$ is independent of the choice of initial distribution μ for $n = 1$. Assume the claim is true for $n \geq 1$. Then we have

$$
\begin{aligned}
&\left\|\lambda_{n+1}^{\pi,\mu} - \lambda_{n+1}^{\pi^k,\mu}\right\|_{TV} \\
&= 2 \sup_{B \in \mathcal{B}(\mathsf{X})} \left|\lambda_1^{\pi,\mu}\left(\lambda_n^{\pi,x_1}(B)\right) - \lambda_1^{\pi^k,\mu}\left(\lambda_n^{\pi^k,x_1}(B)\right)\right| \\
&= 2 \sup_{B \in \mathcal{B}(\mathsf{X})} \Big|\lambda_1^{\pi,\mu}\left(\lambda_n^{\pi,x_1}(B)\right) - \lambda_1^{\pi,\mu}\left(\lambda_n^{\pi^k,x_1}(B)\right) \\
&\qquad\qquad + \lambda_1^{\pi,\mu}\left(\lambda_n^{\pi^k,x_1}(B)\right) - \lambda_1^{\pi^k,\mu}\left(\lambda_n^{\pi^k,x_1}(B)\right)\Big| \\
&\leq \lambda_1^{\pi,\mu}\left(\left\|\lambda_n^{\pi,x_1} - \lambda_n^{\pi^k,x_1}\right\|_{TV}\right) + 2\left\|\lambda_1^{\pi,\mu} - \lambda_1^{\pi^k,\mu}\right\|_{TV} \qquad (9) \\
&\leq \left(\frac{1}{k}\right)^{\frac{1}{d}} (2n-1) K_2 \alpha + 2\left(\frac{1}{k}\right)^{\frac{1}{d}} K_2 \alpha \\
&= \alpha K_2 \left(2(n+1) - 1\right)\left(\frac{1}{k}\right)^{\frac{1}{d}} \alpha. \qquad (10)
\end{aligned}
$$

Here (9) follows since $|\mu(h) - \eta(h)| \leq \|\mu - \eta\|_{TV} \sup_{x \in \mathsf{X}} |h(x)|$ and (10) follows since the bound $\lambda K_2(2n-1)(1/k)^{1/d}$ is independent of the initial distribution.

D. Proof of Theorem 4.1

For any fixed k we have

$$
\begin{aligned}
&\left|w_\beta(\pi) - w_\beta(\pi^k)\right| \\
&= \left|\sum_{n=0}^{\infty} \beta^n \lambda_n^{\pi,\mu}(c_f) - \sum_{n=0}^{\infty} \beta^n \lambda_n^{\pi^k,\mu}(c_{q_k})\right| \\
&\leq \sum_{n=0}^{\infty} \beta^n \Big(\left|\lambda_n^{\pi,\mu}(c_f) - \lambda_n^{\pi,\mu}(c_{q_k})\right| + \left|\lambda_n^{\pi,\mu}(c_{q_k}) - \lambda_n^{\pi^k,\mu}(c_{q_k})\right|\Big) \\
&\leq \sum_{n=0}^{\infty} \beta^n \left(\sup_{x_n \in \mathsf{X}} |c_f - c_{q_k}| + \left\|\lambda_n^{\pi,\mu} - \lambda_n^{\pi,\mu}\right\|_{TV} M\right) \\
&\leq \sum_{n=0}^{\infty} \beta^n \left(\sup_{x_n \in \mathsf{X}} d_{\mathsf{A}}(f(x_n), q_k(x_n))\, K_1 + \|\lambda_n^{\pi,\mu} - \lambda_n^{\pi,\mu}\|_{TV} M\right) \\
&\leq \sum_{n=0}^{\infty} \beta^n \left(\left(\frac{1}{k}\right)^{\frac{1}{d}} \alpha K_1\right) + \sum_{n=1}^{\infty} \beta^n \left(\left(\frac{1}{k}\right)^{\frac{1}{d}} (2n-1) K_2 \alpha M\right) \\
&= \left(\frac{1}{k}\right)^{\frac{1}{d}} \frac{\alpha}{1-\beta}\left(K_1 - \beta K_2 M + \frac{2\beta M K_2}{1-\beta}\right). \qquad (11)
\end{aligned}
$$

Here (11) follows from Assumption (j), Proposition 4.3, and Lemma 4.1, completing the proof.

E. Proof of Theorem 4.2

For any k and $x \in \mathsf{X}$, we have

$$\begin{aligned}
&\left|w_A(\pi,x) - w_A(\pi^k,x)\right| = \left|\nu_\pi(c_f) - \nu_{\pi^k}(c_{q_k})\right| \\
&\le \left|\nu_\pi(c_f) - \nu_\pi(c_{q_k})\right| + \left|\nu_\pi(c_{q_k}) - \nu_{\pi^k}(c_{q_k})\right| \\
&\le \sup_{x\in\mathsf{X}} K_1 d_{\mathsf{A}}\left(f(x), q_k(x)\right) + \|\nu_\pi - \nu_{\pi^k}\|_{TV} M \ \text{(by (j))} \\
&\le \left(\frac{1}{k}\right)^{\frac{1}{d}} K_1\alpha + \left(\|\nu_\pi - \lambda_n^{\pi,x}\|_{TV} + \left\|\lambda_n^{\pi,x} - \lambda_n^{\pi^k,x}\right\|_{TV}\right. \\
&\qquad\qquad \left. + \left\|\lambda_n^{\pi^k,x} - \nu_{\pi^k}\right\|_{TV}\right) M \\
&\le \left(\frac{1}{k}\right)^{\frac{1}{d}} K_1\alpha + \left(2C\kappa^n + \left(\frac{1}{k}\right)^{\frac{1}{d}} (2n-1)K_2\alpha\right) M \\
&= 2MC\kappa^n + \left((2n-1)K_2\alpha M + K_1\alpha\right)\left(\frac{1}{k}\right)^{\frac{1}{d}} \qquad (12)
\end{aligned}$$

where (12) follows from assumption (l) and Proposition 4.3.

REFERENCES

[1] V. Borkar, "Convex analytic methods in Markov decision processes," in *Handbook of Markov Decision Processes*, E. Feinberg and A. Shwartz, Eds. Norwell, MA: Kluwer Academic Publisher, 2002.

[2] D. Bertsekas and J. Tsitsiklis, *Neuro-Dynammic Programming*. Boston, MA: Athena Scientific, 1996.

[3] R. Ortner, "Pseudometrics for state aggregation in average reward Markov decision processes," in *Algorithmic Learning Theory*. Berlin, Germany: Springer-Verlag, 2007.

[4] R. Cavazos-Cadena, "Finite-state approximations for denumerable state discounted Markov decision processes," *Appl. Math. Optim.*, vol. 14, pp. 1–26, 1986.

[5] E. Feinberg, "On measurability and representation of strategic measures in Markov decision processes," *Stat., Prob. Game Theory*, vol. 30, pp. 29–43, pp. 29–43, 1996.

[6] O. Hernández-Lerma and J. Lasserre, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*. New York: Springer, 1996.

[7] O. Hernández-Lerma and J. Lasserre, *Markov Chains and Invariant Probabilities*. Boston, MA: Birkhauser, 2003.

[8] M. Schäl, "On dynamic programming: Compactness of the space of policies," *Stochastic Process. Appl.*, vol. 3, no. 4, pp. 345–364, 1975.

[9] R. Serfozo, "Convergence of Lebesgue integrals with varying measures," *Sankhya Ser. A*, pp. 380–402, 1982.

[10] O. Hernández-Lerma, R. Montes-De-Oca, and R. Cavazos-Cadena, "Recurrence conditions for Markov decision processes with Borel state space: A survey," *Annu. Oper. Res.*, vol. 28, no. 1, pp. 29–46, 1991.

[11] M. Hairer, "Ergodic properties of Markov processes," *Lecture Notes*, 2006.

[12] O. Hernández-Lerma and R. Romera, "Limiting discounted-cost control of partially observable stochastic systems," *SIAM J. Control Optim.*, vol. 40, no. 2, pp. 348–369, 2001.

[13] A. György and T. Linder, "Codecell convexity in optimal entropy-constrained vector quantization," *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1821–1828, Jul. 2003.

[14] E. A. Feinberg, P. O. Kasyanov, and N. V. Zadioanchuk, "Average cost Markov decision processes with weakly continuous transition probabilities," *Math. Oper. Res.*, vol. 37, no. 4, pp. 591–607, Nov. 2012.

[15] E. A. Feinberg and M. E. Lewis, "Optimality inequalities for average cost Markov decision processes and the stochastic cash balance problem," *Math. Oper. Res.*, vol. 32, no. 4, pp. 769–783, Nov. 2007.

[16] A. B. Piunovskiy, *Optimal Control of Random Sequences in Problems With Constraints*. New York: Kluwer, 1997.

[17] 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.

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

[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 2006.

[20] Y. Yamada, S. Tazaki, and R. M. Gray, "Asymptotic performance of block quantizers with difference distortion measures," *IEEE Trans. Inform. Theory*, vol. 26, pp. 6–14, Jan. 1980.

[21] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.

[22] N. Saldi, T. Linder, and S. Yüksel, Quantized Stationary Control Policies in Markov Decision Processes, 2014, arXiv:1310.5770.