On Optimal Zero-Delay Coding of Vector Markov Sources

Tamás Linder and Serdar Yüksel

Abstract

Optimal zero-delay coding (quantization) of a vector-valued Markov source driven by a noise process is considered. Using a stochastic control problem formulation, the existence and structure of optimal quantization policies are studied. For a finite-horizon problem with bounded per-stage distortion measure, the existence of an optimal zero-delay quantization policy is shown provided that the quantizers allowed are ones with convex codecells. The bounded distortion assumption is relaxed to cover cases that include the linear quadratic Gaussian problem. For the infinite horizon problem and a stationary Markov source the optimality of deterministic Markov coding policies is shown. The existence of optimal stationary Markov quantization policies is also shown provided randomization that is shared by the encoder and the decoder is allowed.

Index Terms

Real-time source coding, Markov source, quantization, stochastic control, Markov decision processes.

I. INTRODUCTION

A. Zero-delay coding

We consider a zero-delay (sequential) encoding problem where a sensor encodes an observed information source without delay. It is assumed that the information source \( \{x_t\}_{t \geq 0} \) is an \( \mathbb{R}^d \)-valued discrete-time Markov process. The encoder encodes (quantizes) the source samples and transmits the encoded versions to a receiver over a discrete noiseless channel with input and output alphabet \( M := \{1, 2, \ldots, M\} \), where \( M \) is a positive integer.

Formally, the encoder is specified by a quantization policy \( \Pi \), which is a sequence of Borel measurable functions \( \{\eta_t\}_{t \geq 0} \) with \( \eta_t : M^t \times (\mathbb{R}^d)^{t+1} \rightarrow M \). At time \( t \), the encoder transmits the \( M \)-valued message

\[
q_t = \eta_t(I_t)
\]

The authors are with the Department of Mathematics and Statistics, Queen’s University, Kingston, Ontario, Canada, K7L 3N6. Email: (yuksel,linder)@mast.queensu.ca.

This research was partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

The material of this paper was presented in part at the 51st IEEE Conference on Decision and Control (Maui, Hawaii, Dec. 10-13, 2012) and at the 2013 Workshop on Sequential and Adaptive Information Theory (Montreal, Quebec, Nov. 7-9, 2013).
with \( I_0 = x_0, I_t = (q_{[0,t-1]}, x_{[0,t]}) \) for \( t \geq 1 \), where we have used the notation \( q_{[0,t-1]} = (q_0, \ldots, q_{t-1}) \) and \( x_{[0,t]} = (x_0, x_1, \ldots, x_t) \). The collection of all such zero-delay policies is called the set of admissible quantization policies and is denoted by \( \Pi_A \).

Observe that for fixed \( q_{[0,t-1]} \) and \( x_{[0,t-1]} \), as a function of \( x_t \), the encoder \( \eta_t(q_{[0,t-1]}, x_{[0,t-1]}, \cdot) \) is a quantizer, i.e., a Borel measurable mapping of \( \mathbb{R}^d \) into the finite set \( M \). Thus any quantization policy at each time \( t \geq 0 \) selects a quantizer \( Q_t : \mathbb{R}^d \to M \) based on past information \( (q_{[0,t-1]}, x_{[0,t-1]}) \), and then “quantizes” \( x_t \) as \( q_t = Q_t(x_t) \).

Upon receiving \( q_t \), the receiver generates its reconstruction \( u_t \), also without delay. A zero-delay receiver policy is a sequence of measurable functions \( \gamma = \{\gamma_t\}_{t \geq 0} \) of type \( M^{t+1} \to U \), where \( U \) denotes the reconstruction alphabet (usually a Borel subset of \( \mathbb{R}^d \)). Thus

\[
    u_t = \gamma_t(q_{[0,t]}), \quad t \geq 0.
\]

For the finite horizon setting the goal is to minimize the average cumulative cost (distortion)

\[
    J_{\pi_0}(\Pi, \gamma, T) := E_{\pi_0}^{\Pi, \gamma} \left[ \frac{1}{T} \sum_{t=0}^{T-1} c_0(x_t, u_t) \right],
\]

for some \( T \geq 1 \), where \( c_0 : \mathbb{R}^d \times U \to \mathbb{R} \) is a nonnegative Borel measurable cost (distortion) function and \( E_{\pi_0}^{\Pi, \gamma} \) denotes expectation with initial distribution \( \pi_0 \) for \( x_0 \) and under the quantization policy \( \Pi \) and receiver policy \( \gamma \). We assume that the encoder and decoder know the initial distribution \( \pi_0 \).

We also consider the infinite-horizon average cost problem where the objective is to minimize

\[
    J_{\pi_0}(\Pi, \gamma) := \limsup_{T \to \infty} E_{\pi_0}^{\Pi, \gamma} \left[ \frac{1}{T} \sum_{t=0}^{T-1} c_0(x_t, u_t) \right].
\]

Our main assumption on the Markov source \( \{x_t\} \) is the following.

**Assumption 1.** The evolution of \( \{x_t\} \) is given by

\[
    x_{t+1} = f(x_t, w_t), \quad t = 0, 1, 2, \ldots,
\]

where \( f : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d \) is a Borel function and \( \{w_t\} \) is an independent and identically distributed (i.i.d.) vector noise sequence which is independent of \( x_0 \). It is assumed that for each fixed \( x \in \mathbb{R}^d \), the distribution of \( f(x, w_t) \) admits the (conditional) density function \( \phi(\cdot | x) \) (with respect to the \( d \)-dimensional Lebesgue measure) which is positive everywhere. Furthermore, \( \phi(\cdot | x) \) is bounded and Lipschitz uniformly in \( x \).

The above model includes the linear systems with Gaussian noise. Further conditions on \( f \) and the cost \( c_0 \), and the reconstruction alphabet \( U \) will be given in Sections III and IV for the finite-horizon problem (these include the case of a linear system and quadratic cost) and in Section V for the infinite-horizon problem.

Before proceeding further with formulating the results, we provide an overview of structural results for finite-horizon optimal zero-delay coding problems as well as a more general literature review.
B. Revisiting structural results for finite-horizon problems

Structural results for the finite horizon control problem described in the previous section have been developed in a number of important papers. Among these the classic works by Witsenhausen [34] and Walrand and Varaiya [32], using two different approaches, are of particular relevance. Teneketzis [31] extended these approaches to the more general setting of non-feedback communication and [35] extended these results to more general state spaces (including $\mathbb{R}^d$). The following two theorems summarize, somewhat informally, these two important structural results.

**Theorem 1** (Witsenhausen [34]). For the finite horizon problem, any zero-delay quantization policy $\Pi = \{\eta_t\}$ can be replaced, without any loss in performance, by a policy $\hat{\Pi} = \{\hat{\eta}_t\}$ which only uses $q_{[0,t-1]}$ and $x_t$ to generate $q_t$, i.e., such that $q_t = \hat{\eta}_t(q_{[0,t-1]}, x_t)$ for all $t = 1, \ldots, T - 1$.

For a complete and separable (Polish) metric space $X$ and its Borel sets $B(X)$, let $\mathcal{P}(X)$ denote the space of probability measures on $(X, B(X))$, endowed with the topology of weak convergence (weak topology). This topology is metrizable with the Prokhorov metric making $\mathcal{P}(X)$ itself a Polish space. Given a quantization policy $\Pi$, for all $t \geq 1$ let $\pi_t \in \mathcal{P}(\mathbb{R}^d)$ be the regular conditional probability defined by

$$\pi_t(A) := P(x_t \in A|q_{[0,t-1]})$$

for any Borel set $A \subset \mathbb{R}^d$.

The following result is due to Walrand and Varaiya [32] who considered sources taking values in a finite set. For the more general case of $\mathbb{R}^d$-valued sources the result appeared in [35].

**Theorem 2.** For the finite horizon problem, any zero-delay quantization policy can be replaced, without any loss in performance, by a policy which at any time $t = 1, \ldots, T - 1$ only uses the conditional probability measure $\pi_t = P(dx_t|q_{[0,t-1]})$ and the state $x_t$ to generate $q_t$. In other words, at time $t$ such a policy uses $\pi_t$ to select a quantizer $Q_t : \mathbb{R}^d \to M$ and then $q_t$ is generated as $q_t = Q_t(x_t)$.

As discussed in [35], the main difference between the two structural results above is the following: In the setup of Theorem 1, the encoder’s memory space is not fixed and keeps expanding as the encoding block length $T$ increases. In the setup of Theorem 2, the memory space of an optimal encoder is fixed. Of course, in general the space of probability measures is a very large one. However, it may be the case that different quantization outputs lead to the same conditional probabilities $\pi_t$, leading to a reduction in the required memory. More importantly, the setup of Theorem 2 allows one to apply the powerful theory of Markov Decision Processes on fixed state and action spaces, thus greatly facilitating the analysis.

In this paper, we show that under quite general assumptions on the Markov process, the cost function, and the admissible quantization policies there always exists a policy of the type suggested by Theorem 2 (a so-called Walrand-Varaiya-type policy) that minimizes the finite
horizon cost (1). For the infinite horizon problem (2), we show that there exists an optimal Walrand-Varaiya-type policy if the source is stationary. We also show that in general an optimal (possibly randomized) stationary quantization policy exists in the set of Walrand-Varaiya-type policies.

The rest of the paper is organized as follows. The next section gives a brief review of the literature. Section II contains background material on quantizers and the construction of a controlled Markov chain for our problem. Section III establishes the existence of optimal policies for the finite horizon case for bounded cost functions. Section IV considers the quadratic costs under conditions that cover linear systems. Section V considers the more involved infinite horizon case. Section VI contains concluding discussions. Most of the proofs are relegated to the Appendix.

C. Literature review and contributions

The existence of optimal quantizers for a one-stage \((T = 1)\) cost problem has been investigated in [1], [27], and [38], among other works.

An important inspiration for our work is Borkar et al. [10] which studied the optimal zero-delay quantization of Markov sources. For the infinite horizon setting, this paper provided a stochastic control formulation of the optimal quantization problem with a Lagrangian cost that combined squared distortion and instantaneous entropy, and gave an elegant proof for the existence of optimal policies.

It should be noted that [10] restricted the admissible quantizers \(Q_t\) at each time stage \(t\) to so-called nearest neighbor quantizers whose reconstruction values were also suboptimally constrained to lie within a fixed compact set. Furthermore, some fairly restrictive conditions were placed on the dynamics of the system. These include requirements on the system dynamics that rule out additive noise models with unbounded support such as the Gaussian noise (see p. 138 in [10]), and a uniform Lipschitz condition on the cost functions (see the condition on \(\hat{f}\) on p. 140 in [10]). These conditions made it possible to apply the discounted cost approach (see, e.g., [2]) to average cost optimization problems.

Furthermore, the encoder-decoder structure in [10] has been specified \textit{a priori}, whereas in this paper, we only relax global optimality when we restrict the quantizers to have convex codecells (to be defined later), which is a more general condition than assuming the nearest neighbor encoding rule. On the other hand, we are unable to claim the optimality of deterministic stationary quantization policies for the infinite-horizon problem, whereas [10] establishes optimality of such policies. However, as mentioned, the conditions on the cost functions, systems dynamics, and the uniform continuity condition over all quantizers are not required in our setting.

To our knowledge, the existence of optimal quantizers for a finite horizon setting has not been considered in the literature for the setup considered in this paper.

Other relevant work include [9] which considered optimization over probability measures for causal and non-causal settings, and [31], [22], [21] and [35] which considered zero-delay coding
of Markov sources in various setups. Structural theorems for zero-delay variable-rate coding of discrete Markov sources were studied in [19]. Recently [4] considered the average cost optimality equation for coding of discrete i.i.d. sources with limited lookahead and [18] studied real-time joint source-channel coding of a discrete Markov source over a discrete memoryless channel with feedback.

A different model for sequential source coding, called causal source coding, is studied in, e.g., [26], [33], [20]. In causal coding, the reconstruction depends causally on the source symbols, but in the information transmission process large delays are permitted, which makes this model less stringent (and one might argue less practical) than zero or limited-delay source coding.

For systems with control, structural results have also been investigated in the literature. In particular, for linear systems with quadratic cost criteria (known as LQG optimal control problems), it has been shown that the effect of the control policies can be decoupled from the estimation error without any loss. Under optimal control policies, [36] has shown the equivalence with the control-free setting considered in this paper (see also [5] and [25] for related results in different structural forms). We also note that the design results developed here can be used to establish the existence of optimal quantization and control policies for LQG systems [36].

**Contributions:** In view of the literature review, the main contributions of the paper can be summarized as follows.

(i) We establish a useful topology on the set of quantizers, building on [38], among other works, and show the existence of optimal coding policies for finite horizon optimization problems, under the assumption that the quantizers used have convex codecells. Notably, the set of sources considered includes LQG systems, i.e., linear systems driven by Gaussian noise under the quadratic cost criterion. The analysis requires the development of a series of technical results which facilitate establishing measurable selection criteria, reminiscent of those in [15].

(ii) We establish, for the first time to our knowledge, the optimality of Markov (i.e., Walrand-Varaiya type) coding policies for infinite-horizon sequential quantization problems, using a new approach. The prior work reviewed above strictly build on dynamic programming (which is only suitable for finite-horizon problems) or does not consider the question of global optimality of Markov policies.

(iii) We show the existence of optimal stationary, possibly randomized, policies which are globally optimal, for a large class of sources including LQG systems. As detailed above, the assumptions are weaker than those that have appeared in prior work.

**II. Quantizer actions and controlled Markov process construction**

In this section, we formally define the space of quantizers considered in the paper building on the construction in [38]. Recall the notation \( M = \{1, \ldots, M\} \).
**Definition 1.** An $M$-cell quantizer $Q$ on $\mathbb{R}^d$ is a (Borel) measurable mapping $Q : \mathbb{R}^d \to M$. We let $Q$ denote the collection of all $M$-cell quantizers on $\mathbb{R}^d$.

Note that each $Q \in Q$ is uniquely characterized by its quantization cells (or bins) $B_i = Q^{-1}(i) = \{x : Q(x) = i\}$, $i = 1, \ldots, M$ which form a measurable partition of $\mathbb{R}^d$.

**Remark 1.**

(i) We allow for the possibility that some of the cells of the quantizer are empty.

(ii) In source coding theory (see, e.g., [13]), a quantizer is a mapping $Q : \mathbb{R}^d \to \mathbb{R}^d$ with a finite range. In this definition, $Q$ is specified by a partition $\{B_1, \ldots, B_M\}$ of $\mathbb{R}^d$ and reconstruction values $\{c_1, \ldots, c_M\} \subset \mathbb{R}^d$ through the mapping rule $Q(x) = c_i$ if $x \in B_i$. In our definition, we do not include the reconstruction values.

In view of Theorem 2, any admissible quantization policy can be replaced by a Walrand-Varaiya-type policy. The class of all such policies is denoted by $\Pi_W$ and is formally defined as follows.

**Definition 2.** An (admissible) quantization policy $\Pi = \{\eta_t\}$ belongs to $\Pi_W$ if there exist a sequence of mappings $\{\hat{\eta}_t\}$ of the type $\hat{\eta}_t : \mathcal{P}(\mathbb{R}^d) \to Q$ such that for $Q_t = \hat{\eta}_t(\pi_t)$ we have $q_t = Q_t(x_t) = \eta_t(I_t)$.

Suppose we use a quantizer policy $\Pi = \{\hat{\eta}_t\}$ in $\Pi_W$. Let $P(dx_{t+1}|x_t)$ denote the transition kernel of the process $\{x_t\}$ determined by the system dynamics (3) and note that $P(q_t|\pi_t, x_t)$ is determined by the quantizer policy as $P(q_t|\pi_t, x_t) = 1_{\{Q_t(x_t) = q_t\}}$, where $Q_t = \hat{\eta}_t(\pi_t)$. Then standard properties of conditional probability can be used to obtain the following filtering equation for the evolution of $\pi_t$:

$$
\pi_{t+1}(dx_{t+1}) = \frac{\int_{\mathbb{R}^d} \pi_t(dx_t) P(q_t|\pi_t, x_t) P(dx_{t+1}|x_t)}{\int_{\mathbb{R}^d} \pi_t(dx_t) P(q_t|\pi_t, x_t) P(dx_{t+1}|x_t)} P(dx_{t+1}|x_t) \pi_t(dx_t).
$$

Clearly, given $\pi_t$ and $Q_t$, $\pi_{t+1}$ is conditionally independent of $\{\pi_{[0,t-1]}, Q_{[0,t-1]}\}$. Thus $\{\pi_t\}$ can be viewed as $\mathcal{P}(\mathbb{R}^d)$-valued controlled Markov process [15], [16] with $Q$-valued control $\{Q_t\}$ and average cost up to time $T - 1$ given by

$$
E_{\pi_0}^\Pi \left[ \frac{1}{T} \sum_{t=0}^{T-1} c(\pi_t, Q_t) \right] = \inf_\gamma J_{\pi_0}(\Pi, \gamma, T),
$$

where

$$
c(\pi_t, Q_t) := \sum_{i=1}^M \inf_{u \in U} \int_{Q^{-1}(i)} \pi_t(dx) c_0(x, u).
$$

In this context, $\Pi_W$ corresponds to the class of deterministic Markov control policies.
Recall that by Assumption 1 the density $\phi(\cdot | x)$ of $f(x, w_t)$ for fixed $x$ is bounded, positive, and Lipschitz, uniformly in $x$. By (3) and (4) $\pi_t$ admits a density, which we also denote by $\pi_t$, given by

$$\pi_t(z) = \int_{\mathbb{R}^d} \phi(z|x_{t-1})P(dx_{t-1}|q_{[0,t-1]}), \quad z \in \mathbb{R}^d, \quad t \geq 1.$$ 

Thus for any policy $\Pi$, with probability 1 we have $0 < \pi_t(z) \leq C$ for all $z$ and $t \geq 1$, where $C$ is an upper bound on $\phi$. Also, if $\phi(z|x)$ is Lipschitz in $z$ with constant $C_1$ for all $x$, then the bound

$$|\pi_t(z) - \pi_t(z')| \leq \int_{\mathbb{R}^d} |\phi(z|x_{t-1}) - \phi(z'|x_{t-1})|P(dx_{t-1}|q_{[0,t-1]}),$$

implies that $\{\pi_t\}_{t \geq 1}$ is uniformly Lipschitz with constant $C_1$. Let $S$ denote the set of all probability measures on $\mathbb{R}^d$ admitting densities that are bounded by $C$ and Lipschitz with constant $C_1$. Note that viewed as a class of densities, $S$ is uniformly bounded and equicontinuous. Lemma 3 in the Appendix shows that $S$ is closed in $\mathcal{P}(\mathbb{R}^d)$. Also, the preceding argument implies the following useful lemma.

**Lemma 1.** For any policy $\Pi \in \Pi_W$, we have $\pi_t \in S$ for all $t \geq 1$ with probability 1.

For technical reasons in most of what follows we restrict the set of quantizers by only allowing ones that have convex cells. Formally, this quantizer class $Q_c$ is defined by

$$Q_c = \{Q \in Q : Q^{-1}(i) \subset \mathbb{R}^d \text{ is convex for } i = 1, \ldots, M\},$$

where by convention we declare the empty set convex. Note that each nonempty cell of a $Q \in Q_c$ is a convex polytope in $\mathbb{R}^d$. The class of policies $\Pi^c_W$ is obtained by replacing $Q$ with $Q_c$ in Definition 2:

**Definition 3.** $\Pi^c_W$ denotes the set of all quantization policies $\Pi = \{\hat{\eta}_t\} \in \Pi_W$ such that $\hat{\eta}_t : \mathcal{P}(\mathbb{R}^d) \to Q_c$, i.e., $Q_t = \hat{\eta}_t(\pi_t) \in Q_c$ for all $t \geq 0$.

**Remark 2.**

(i) We note that the assumption of convex codecells is adopted for technical reasons and it likely results in a loss of system optimality. However, $Q_c$ is a fairly powerful class and it includes the class of nearest-neighbor quantizers considered in [10].

(ii) As opposed to general quantizers in $Q$, any $Q \in Q_c$ has a parametric representation. Let such a $Q$ have cells $\{B_1, \ldots, B_M\}$. As discussed in [14], by the separating hyperplane theorem, there exist pairs of complementary closed half spaces $\{(H_{i,j}, H_{j,i}) : 1 \leq i, j \leq M, i \neq j\}$ such that $B_i \subset \bigcap_{j \neq i} H_{i,j}$ for all $i$. Since $\overline{B}_i := \bigcap_{j \neq i} H_{i,j}$ is a closed convex polytope for each $i$, if $P \in \mathcal{P}(\mathbb{R}^d)$ admits a density, then $P(\overline{B}_i \setminus B_i) = 0$ for all $i$. We thus obtain a $P$-almost sure representation of $Q$ by the $M(M-1)/2$ hyperplanes $h_{i,j} = H_{i,j} \cap H_{j,i}$. One can represent such a hyperplane $h$ by a vector $(a_1, \ldots, a_d, b) \in \mathbb{R}^{d+1}$ with $\sum_k |a_k|^2 = 1$ such
that \( h = \{ x \in \mathbb{R}^d : \sum a_i x_i = b \} \), thus obtaining a parametrization over \( \mathbb{R}^{(d+1)M(M-1)/2} \) of all quantizers in \( Q_c \).

In order to facilitate the stochastic control analysis of the quantization problem we need an alternative representation of quantizers. As discussed in, e.g., [9] and [38], a quantizer \( Q \) with cells \( \{ B_1, \ldots, B_M \} \) can also be identified with the stochastic kernel (regular conditional probability), also denoted by \( Q \), from \( \mathbb{R}^d \) to \( M \) defined by

\[
Q(i|x) = \mathbb{1}_{\{ x \in B_i \}}, \quad i = 1, \ldots, M.
\]

We will endow the set of quantizers \( Q_c \) with a topology induced by the stochastic kernel interpretation. If \( P \) is a probability measure on \( \mathbb{R}^d \) and \( Q \) is a stochastic kernel from \( \mathbb{R}^d \) to \( M \), then \( PQ \) denotes the resulting joint probability measure on \( \mathbb{R}^d \times M \) defined through \( PQ(dx \, dy) = P(dx)Q(dy|x) \). For some fixed \( P \in \mathcal{P}(\mathbb{R}^d) \) let

\[
\Gamma_P := \{ PQ \in \mathcal{P}(\mathbb{R}^d \times M) : Q \in Q_c \}
\]

It follows from [38, Thm. 5.8] that \( \Gamma_P \) is a compact subset of \( \mathcal{P}(\mathbb{R}^d \times M) \) if \( P \) admits a density. If we introduce the equivalence relation \( Q \equiv Q' \) if and only if \( PQ = PQ' \), then the resulting set of equivalence classes, denoted by \( (Q_c)_P \), can be equipped with the quotient topology inherited from \( \Gamma_P \). In this topology \( Q_n \to Q \) if and only if (for representatives of the equivalence classes) \( PQ_n \to PQ \) weakly. Also, if \( PQ = PQ' \) for \( P \) admitting a positive density, then the (convex polytopal) cells of \( Q \) and \( Q' \) may differ only in their boundaries, and it follows that \( (Q_c)_P = (Q_c)_{P'} \) for any \( P' \) also admitting a positive density. From now on we will identify \( Q_c \) with \( (Q_c)_P \) and endow it with the resulting quotient topology, keeping in mind that this definition does not depend on \( P \) as long as it has a positive density. Lemma 3 in the Appendix shows that \( Q_c \) is compact.

For a given policy \( \Pi \in \Pi_W^C \), we will consider \( \{ (\pi_t, Q_t) \} \) as an \( S \times Q_c \)-valued process.

### III. Existence of optimal policies: Finite horizon setting

For any quantization policy \( \Pi \in \Pi_W \) and any \( T \geq 1 \) we define

\[
J_{\pi_0}(\Pi, T) := \inf_{\gamma} J_{\pi_0}(\Pi, \gamma, T) = \mathbb{E}_{\pi_0}^{\Pi} \left[ \frac{1}{T} \sum_{t=0}^{T-1} c(\pi_t, Q_t) \right],
\]

where \( c(\pi_t, Q_t) \) is defined in (5).

#### Assumption 2.

(i) The cost \( c_0 : \mathbb{R}^d \times U \to \mathbb{R} \) is nonnegative, bounded, and continuous.

(ii) \( U \) is compact.

#### Theorem 3. Under Assumptions 1 and 2 an optimal receiver policy always exists, i.e., for any \( \Pi \in \Pi_W \) there exist \( \gamma = \{ \gamma_t \} \) such that \( J_{\pi_0}(\Pi, \gamma, T) = J_{\pi_0}(\Pi, T) \).
Proof. At any \( t \geq 0 \) an optimal receiver has to minimize \( \int P(dx_t|q_{[0,t]})c_0(x_t, u) \) in \( u \). Under Assumption 2, the existence of a minimizer then follows from a standard argument, see, e.g., [38, Theorem 3.1]. □

The following result states the existence of optimal policies in \( Q_c \) for the finite horizon setting. The proof is given in Section VII-B of the Appendix.

**Theorem 4.** Suppose \( \pi_0 \) admits a density or it is a point mass \( \pi_0 = \delta_{x_0} \) for some \( x_0 \in \mathbb{R}^d \). For any \( T \geq 1 \), under Assumptions 1 and 2, there exists a policy \( \Pi \) in \( \Pi_C^W \) such that

\[
J_{\pi_0}(\Pi, T) = \inf_{\Pi' \in \Pi_C^W} J_{\pi_0}(\Pi', T). 
\]  

(6)

Letting \( J^T_T(\cdot) := 0 \) and

\[
J^T_0(\pi_0) := \min_{\Pi \in \Pi_C^W} J_{\pi_0}(\Pi, T),
\]

the dynamic programming recursion

\[
TJ^T_t(\pi) = \min_{Q \in Q_c} \left( c(\pi, Q) + TE\left[J^T_{t+1}(\pi_{t+1})|\pi_t = \pi, Q_t = Q\right]\right)
\]

(7)

holds for \( t = T - 1, T - 2, \ldots, 1, \pi \in \mathcal{S} \) and \( t = 0, \pi = \pi_0 \).

IV. THE FINITE HORIZON PROBLEM FOR QUADRATIC COST

Linear systems driven by Gaussian noise are important in many applications in control, estimation, and signal processing. For such linear systems with quadratic cost (known as LQG optimal control problems), it has been shown that the effect of the control policies can be decoupled from the estimation error without any loss (see [36], [30] and for a review [37]). In this section we consider the finite horizon problem under conditions that cover LQG systems. Let \( \|x\| \) denote the Euclidean norm of \( x \in \mathbb{R}^d \). We replace Assumption 2 of the preceding sections with the following.

**Assumption 3.**

(i) The function \( f \) in the system dynamics (3) satisfies \( \|f(x, w)\| \leq K(\|x\| + \|w\|) \) for some \( K > 0 \) and all \( x, w \in \mathbb{R}^d \).

(ii) \( U = \mathbb{R}^d \) and the cost is given by \( c_0(x, u) = \|x - u\|^2 \).

(iii) The common distribution \( \nu_w \) of the \( w_t \) satisfies \( \int z^2\nu_w(dz) < \infty \).

(iv) \( \pi_0 \) admits a density such that \( E_{\pi_0}[\|x_0\|^2] < \infty \) or it is a point mass \( \pi_0 = \delta_{x_0} \).

**Remark 3.**

(i) The above conditions cover the case of a linear-Gaussian system

\[
x_{t+1} = Ax_t + w_t, \quad t = 0, 1, 2, \ldots,
\]

where \( \{w_t\} \) is an i.i.d. Gaussian noise sequence with zero mean, \( A \) is a square matrix, and \( \pi_0 \) admits a Gaussian density having zero mean.
(ii) Assumption 3(i) implies
\[ \|x_t\|^2 \leq \hat{K} \left( \|x_0\|^2 + \sum_{i=0}^{t-1} \|w_i\|^2 \right) \]
for some \( \hat{K} \) that depends on \( t \) (see (39)). Together with Assumptions 3(iii) and (iv), this implies \( E_{\pi_0}[\|x_t\|^2] < \infty \) for all \( t \geq 0 \) under any quantization policy. Therefore
\[ \int_{\mathbb{R}^d} \|x_t\|^2 P(dx_t|q_{[0,t]}) \, dx < \infty \]
and an optimal receiver policy exists and is given by
\[ \gamma_t(q_{[0,t]}) = \int_{\mathbb{R}^d} x_t P(dx_t|q_{[0,t]}). \] (8)

The following is a restatement of Theorem 4 under conditions that allow unbounded cost. The proof is relegated to Section VII-C of the Appendix.

**Theorem 5.** Under Assumptions 1 and 3, for any \( T \geq 1 \) there exists an optimal policy in \( \Pi_{C_0} \) in the sense of (6) and the dynamic programming recursion (7) for \( J^T_t(\pi_t) \) also holds.

V. INFINITE HORIZON SETTING

For the infinite horizon setting, one may consider the discounted cost problem where the goal is to find policies that achieve
\[ V^\beta(\pi_0) = \inf_{\Pi \in \Pi_{C_0}^\infty} J^\beta_{\pi_0}(\Pi) \] (9)
for some \( \beta \in (0, 1) \), where
\[ J^\beta_{\pi_0}(\Pi) = \inf_{\gamma} \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \beta^t c_0(x_t, u_t). \]

The existence of optimal policies for this problem follows from the results in the previous section. In particular, it is well known that the value iteration algorithm (see, e.g., [23]) will converge to an optimal solution, since the cost function is bounded and the measurable selection hypothesis is applicable in view of Theorem 4. This leads to fixed point equation
\[ V^\beta(\pi) = \min_{Q \in \mathcal{Q}_c} \left( c(\pi, Q) + \beta \int_{\mathbb{R}^d} P(d\pi_{t+1}|\pi_t = \pi, Q_t = Q) V^\beta(\pi_{t+1}) \right). \]

The more challenging case is the average cost problem where one considers
\[ J_{\pi_0}(\Pi, \gamma) = \limsup_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} c_0(x_t, u_t) \] (10)
and the goal is to find an optimal policy attaining
\[ J_{\pi_0} := \inf_{\Pi \in \Pi_{A}} \inf_{\gamma} J_{\pi_0}(\Pi, \gamma). \] (11)
For the infinite horizon setting the structural results in Theorems 1 and 2 are not available in the literature, due to the fact that the proofs are based on dynamic programming which starts at a finite terminal time stage and optimal policies are computed backwards. However, we can prove an infinite-horizon analog of Theorem 2 assuming that an invariant measure \( \pi^* \) for \( \{x_t\} \) exists and \( \pi_0 = \pi^* \).

**Theorem 6.** Assume the cost \( c_0 \) is bounded and an invariant measure \( \pi^* \) exists. If \( \{x_t\} \) starts from \( \pi^* \), then there exists an optimal policy in \( \Pi_W \) solving the minimization problem (11), i.e.,

\[
\limsup_{T \to \infty} E_{\pi^*}^\Pi \left[ \frac{1}{T} \sum_{t=0}^{T} c(\pi_t, Q_t) \right] = J_{\pi^*}.
\]

The proof of the theorem relies on a construction that pieces together policies from \( \Pi_W \) that on time segments of appropriately large lengths increasingly well approximate the minimum infinite-horizon cost achievable by policies in \( \Pi_A \). Since the details are somewhat tedious, the proof is relegated to Section VII-D of the Appendix. We note that the condition that \( c_0 \) is bounded is not essential and, for example, the theorem holds for the quadratic cost if the invariant measure has a finite second moment.

The optimal policy constructed in the proof of Theorem 6 may not be stationary. In the next section we establish the existence of an optimal stationary policies in \( \Pi_W^C \) if randomization is allowed.

**A. Classes of randomized quantization policies**

We will consider two classes of randomized policies.

**Randomized Walrand-Varaiya-type (Markov) policies:** These policies, denoted by \( \bar{\Pi}_W^C \), are randomized over \( \Pi_W^C \), the Walrand-Varaiya-type Markov policies with quantizers having convex cells (Definition 3). Each \( \Pi \in \bar{\Pi}_W^C \) consists of a sequence of stochastic kernels \( \{\tilde{\eta}_t\} \) from \( \mathcal{P}(\mathbb{R}^d) \) to \( Q_c \). Thus, under \( \Pi \), for any \( t \geq 0 \),

\[
P^\Pi(Q_t(x_t) = q_t | q_{[0,t-1]}, Q_{[0,t-1]}, \pi_{[0,t]}) = \int_{Q_c} \left( \int_{\mathbb{R}^d} 1_{Q_t = q_t}(dx) \pi_t(dx) \right) \eta_t(dQ_t | \pi_t).
\]

(12)

It follows from, e.g., [12] or [28] that an equivalent model for randomization can be obtained by considering an i.i.d. randomization sequence \( \{r_t\} \), independent of \( \{x_t\} \) and uniformly distributed on \( [0,1] \), and a sequence of (measurable) randomized encoders \( \{\hat{\eta}_t\} \) of the form \( \hat{\eta}_t : \mathcal{P}(\mathbb{R}^d) \times [0,1] \to Q_c \) and \( Q_t \), such that \( Q_t = \hat{\eta}_t(\pi_t, r_t) \). In this case the induced stochastic kernel encoder \( \bar{\eta}_t \) is determined by

\[
\bar{\eta}_t(D | \pi_t) = u \{ r : \hat{\eta}(\pi_t, r) \in D \}
\]
for any Borel subset $D$ of $Q_c$, where $u$ denotes the uniform distribution on $[0, 1]$. For randomized policies we assume that all the randomization information is shared between the encoder and the decoder, that is

$$I_t^r := (q_{[0,t-1]}, r_{[0,t-1]})$$

is known at the decoder which can therefore track $\pi_t$ given by

$$\pi_t(A) := P(x_t \in A | q_{[0,t-1]}, r_{[0,t-1]})$$

for any Borel set $A \subset \mathbb{R}^d$.

We note that the cost $c(\pi_t, Q_t)$ is still defined by (5) since the decoder, having access to $I_t^r$ can also track $Q_t$. Also, in computing the cost $E_{\Pi}^{\Pi_0} \left[ \frac{1}{T} \sum_{t=0}^{T} c(\pi_t, Q_t) \right]$ of policy $\Pi \in \Pi_C^C$ after $T$ time stages, the expectation is also taken with respect to the randomization sequence $\{r_t\}$.

**Randomized stationary Walrand-Varaiya-type (Markov) policies:** Denoted by $\bar{\Pi}_{C,W,S}^C$, this class consists of all policies in $\bar{\Pi}_{C,W}^C$ that are stationary, i.e., the stochastic kernels $\bar{\eta}_t$ or the randomized encoders $\hat{\eta}_t$ do not depend on the time index $t$.

**B. Existence of optimal stationary policies**

1) **The bounded cost case:** In the infinite horizon setting, we add the following assumption, in addition to Assumptions 1 and 2.

**Assumption 4.** The chain $\{x_t\}$ is positive Harris recurrent (see [24]) with unique invariant measure $\pi^*$ such that for all $x_0 \in \mathbb{R}^d$,

$$\lim_{t \to \infty} E_{\delta_{x_0}} [\|x_t\|^2] = E_{\pi^*} [\|x\|^2] < \infty.$$

**Remark 4.**

(i) Under Assumption 1, Assumption 4 is implied by the condition that there exists some $x_0 \in \mathbb{R}^d$ so that $\lim_{t \to \infty} E_{\delta_{x_0}} [\|x_t\|^2] < \infty$. This follows from the following: The sequence of expected occupation measures for such a chain starting with initial condition $x_0$ is tight. Furthermore, this sequence of measures satisfies a uniform countable additivity condition, and as a result, has a converging subsequence in the setwise sense, the limit of which can be shown to be invariant. The uniqueness of the invariant measure $\pi^*$ follows from irreducibility, since there cannot be two disjoint absorbing sets by the positivity of the conditional density $\phi(\cdot | x)$ of $x_{t+1} = f(x_t, w_t)$ given $x_t = x$.

(ii) According to the preceding remark, a sufficient condition for Assumption 4 to hold is that $f$ in (3) satisfies $\|f(x, w)\| \leq K (\|x\| + \|w\|)$ for some $K < 1$ and $w_t$ has zero mean and second moment $E[\|w_t\|^2] < \infty$.

To show the existence of an optimal stationary policy, we adopt the convex analytic approach of [7] (see [2] for a detailed discussion). Here we only present the essential steps.
Fix a policy $\Pi \in \Pi_W^C$ and an initial distribution $\pi_0$. Let $v_t \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d) \times Q_c)$ be the sequence of expected occupation measures determined by

$$v_t(D) = E_{\pi_0}^\Pi \left( \frac{1}{t} \sum_{i=0}^{t-1} 1_{\{\pi_i, Q_i \in D\}} \right)$$

for any Borel subset $D$ of $\mathcal{P}(\mathbb{R}^d) \times Q_c$.

Let $P(d\pi_{t+1}|\pi_t, Q_t) = P^\Pi(d\pi_{t+1}|\pi_t, Q_t)$ be the transition kernel determined by the filtering equation (4) and note that it does not depend on $\Pi$ and $t$. Also note that $P(S|\pi, Q) = 1$ for any $\pi$ and $Q$, where $S \subset \mathcal{P}(\mathbb{R}^d)$ is the set of probability measures, defined in Section II, admitting densities that satisfy the same bound and Lipschitz condition as the density of the additive noise $w_t$ ($S$ contains the set of reachable states for $\{\pi_t\}$ under any quantization policy).

If $X$ is a topological space, let $\mathcal{C}_b(X)$ denote the set of all bounded and continuous real-valued functions on $X$. Let $\mathcal{G}$ be the set of so-called ergodic occupation measures on $\mathcal{P}(\mathbb{R}^d) \times Q_c$, defined by

$$\mathcal{G} = \left\{ v \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d) \times Q_c) : \int f(\pi) v(d\pi dQ) = \int \int f(\pi') P(d\pi' | \pi, Q) v(d\pi' dQ) \text{ for all } f \in \mathcal{C}_b(\mathcal{P}(\mathbb{R}^d)) \right\}.$$

Note that any $v \in \mathcal{G}$ is supported on $S \times Q_c$.

Any $v \in \mathcal{G}$ can be disintegrated as $v(d\pi dQ) = \check{v}(d\pi) \check{\eta}(dQ | \pi)$, where $\check{\eta}$ is a stochastic kernel from $\mathcal{P}(\mathbb{R}^d)$ to $Q_c$ which corresponds to the randomized stationary policy $\Pi = \{\check{\eta}_t\}$ in $\Pi_W^C$ such that $\check{\eta}_t = \check{\eta}$ for all $t$. Then the transition kernel of the process $\{(\pi_t, Q_t)\}$ induced by $\Pi$ does not depend on $t$ and is given by

$$P^\Pi(d\pi_{t+1} dQ_{t+1} | \pi_t, Q_t) = P(d\pi_{t+1} | \pi_t, Q_t) \check{\eta}(dQ_{t+1} | \pi_t).$$

In fact, it directly follows from the definition of $\mathcal{G}$ that

$$\int g(\pi, Q) v(d\pi dQ) = \int \int g(\pi', Q') P^\Pi(d\pi' dQ' | \pi, Q) v(d\pi' dQ) \quad (13)$$

for all $g \in \mathcal{C}_b(\mathcal{P}(\mathbb{R}^d) \times Q_c)$, i.e., $v$ is an invariant measure for the transition kernel $P^\Pi$.

The following proposition, proved in Section VII-E, will imply the existence of optimal stationary policies.

**Proposition 1.** (a) For any initial distribution $\pi_0$ and policy $\Pi \in \Pi_W^C$, if $\{v_{t_n}\}$ is a subsequence of the expected occupation measures $\{v_t\}$ such that $v_{t_n} \to \bar{v}$ weakly, then $\bar{v} \in \mathcal{G}$.

Furthermore

$$\lim_{n \to \infty} \int_{\mathcal{P}(\mathbb{R}^d) \times Q_c} c(\pi, Q) v_{t_n}(d\pi dQ) = \int_{\mathcal{P}(\mathbb{R}^d) \times Q_c} c(\pi, Q) \bar{v}(d\pi dQ). \quad (14)$$

(b) For any $x_0 \in \mathbb{R}^d$, initial distribution $\pi_0 = \delta_{x_0}$, and policy $\Pi \in \Pi_W^C$, $\{v_t\}$ is relatively compact.

May 7, 2014
(c) \( G \) is compact.

For any initial distribution \( \delta_{x_0} \) and policy \( \Pi \in \bar{\Pi}_W \), we have

\[
\liminf_{T \to \infty} E^\Pi_{\delta_{x_0}} \left[ \frac{1}{T} \sum_{t=0}^{T-1} c(\pi_t, Q_t) \right] = \liminf_{T \to \infty} \int_{\mathcal{P}(\mathbb{R}^d) \times \mathcal{Q}_c} c(\pi, Q) v_T(d\pi dQ).
\]

Let \( \{v_{T_n}\} \) be a subsequence of \( \{v_T\} \) such that

\[
\liminf_{T \to \infty} \int_{\mathcal{P}(\mathbb{R}^d) \times \mathcal{Q}_c} c(\pi, Q) v_T(d\pi dQ) = \lim_{n \to \infty} \int_{\mathcal{P}(\mathbb{R}^d) \times \mathcal{Q}_c} c(\pi, Q) v_{T_n}(d\pi dQ).
\]

By Proposition 1(b) there exists a subsequence of \( \{v_{T_n}\} \), which we also denote by \( \{v_{T_n}\} \), weakly converging to some \( \bar{v} \). By Proposition 1(a) we have \( \bar{v} \in G \) and \( \int c d\bar{v}_{T_n} \to \int c d\bar{v} \). Therefore

\[
\liminf_{T \to \infty} E^\Pi_{\delta_{x_0}} \left[ \frac{1}{T} \sum_{t=0}^{T-1} c(\pi_t, Q_t) \right] = \int_{\mathcal{P}(\mathbb{R}^d) \times \mathcal{Q}_c} c(\pi, Q) \bar{v}(d\pi dQ)
\geq \inf_{v \in \mathcal{G}} \int_{\mathcal{P}(\mathbb{R}^d) \times \mathcal{Q}_c} c(\pi, Q) v(d\pi dQ).
\]

In addition, since \( c \) is continuous on \( S \times \mathcal{Q}_c \) (by Lemma 4) and each \( v \in \mathcal{G} \) is supported on \( S \times \mathcal{Q}_c \), the mapping \( v \mapsto \int c d\bar{v} \) is continuous on \( \mathcal{G} \). Since \( \mathcal{G} \) is compact by Proposition 1(c), there exists \( v^* \in \mathcal{G} \) achieving the above infimum. Hence

\[
c^* := \int_{\mathcal{P}(\mathbb{R}^d) \times \mathcal{Q}_c} c(\pi, Q) v^*(d\pi dQ) = \min_{v \in \mathcal{G}} \int_{\mathcal{P}(\mathbb{R}^d) \times \mathcal{Q}_c} c(\pi, Q) v(d\pi, dQ)
\tag{15}
\]

provides an ultimate lower bound on the infinite-horizon cost of any policy.

The following theorem shows the existence of a stationary policy achieving this lower bound if we consider the initial distribution \( \pi_0 \) as “design parameter” we can freely choose.

**Theorem 7.** Under Assumptions 1, 2 and 4, there exists a stationary policy \( \Pi^* \in \bar{\Pi}_{W,S}^C \) that is optimal in the sense that with an appropriately chosen initial distribution \( \pi_0^* \),

\[
\lim_{T \to \infty} E^\Pi_{\pi_0^*} \left[ \frac{1}{T} \sum_{t=0}^{T-1} c(\pi_t, Q_t) \right] \leq \liminf_{T \to \infty} E^\Pi_{\delta_{x_0}} \left[ \frac{1}{T} \sum_{t=0}^{T-1} c(\pi_t, Q_t) \right]
\]

for any \( x_0 \in \mathbb{R}^d \) and \( \Pi \in \bar{\Pi}_{W}^C \).

**Proof:** We must prove the existence of \( \Pi^* \in \bar{\Pi}_{W,S}^C \) which achieves infinite horizon cost \( c^* \) for some initial distribution \( \pi_0^* \). Consider \( v^* \) achieving the minimum in (15), disintegrate it as \( v^*(d\pi dQ) = \bar{v}^*(d\pi) \bar{\eta}^*(dQ|\pi) \), and let \( \Pi^* \in \bar{\Pi}_{W,S}^C \) be the policy corresponding to \( \bar{\eta}^* \). Since \( v^* \) is an invariant measure for the transition kernel \( P^{\Pi^*} \) (see (13)), for any \( T \geq 1 \),

\[
c^* = E^*_{\delta_{x_0}} \left[ \frac{1}{T} \sum_{t=0}^{T-1} c(\pi_t, Q_t) \right],
\]
where the notation $E_{\hat{v}^*}^{\Pi^*}$ signifies that the initial distribution $\pi_0$ is picked randomly with distribution $\hat{v}^*$. Thus

$$c^* = \lim_{T \to \infty} E_{\hat{v}^*}^{\Pi^*} \left[ \frac{1}{T} \sum_{t=0}^{T-1} c(\pi_t, Q_t) \right]. \quad (16)$$

From the individual ergodic theorem (see [17]) the limit

$$f(\pi_0) := \lim_{T \to \infty} E_{\pi_0}^{\Pi^*} \left[ \frac{1}{T} \sum_{t=0}^{T-1} c(\pi_t, Q_t) \right]$$

exists for $\hat{v}^*$-a.e. $\pi_0$ and

$$\int_{\mathcal{P}(\mathbb{R}^d)} f(\pi_0) \hat{v}^*(d\pi_0) = c^*.$$

Hence for some $\pi_0$ in the support of $\hat{v}^*$ we must have

$$\lim_{T \to \infty} E_{\pi_0}^{\Pi^*} \left[ \frac{1}{T} \sum_{t=0}^{T-1} c(\pi_t, Q_t) \right] \leq c^*$$

Any such $\pi_0$ can be picked as $\pi_0^*$ so that the claim of the theorem holds.

In the preceding theorem the initial state distribution $\pi_0$ is a design parameter which is chosen along with the quantization policy to optimize the cost. This assumption may be unrealistic. However, consider the fictitious optimal stationary policy in (16) which is allowed to pick the initial distribution $\pi_0$ according to $\hat{v}^*$. It follows from the analysis in the proof of Proposition 1 (see (55)) that the expectation of $\pi_0$ according to $\hat{v}^*$ is precisely the invariant distribution $\pi^*$ for $\{x_t\}$. Based on this, one can prove the following, more realistic version of the optimality result. The proof, which is not given here, is an expanded and more refined version of the proof of Theorem 7.

**Theorem 8.** Under the setup of Theorem 7, assume that $\{x_t\}$ is started from the invariant distribution $\pi^*$. If the optimal stationary policy $\Pi^* \in \Pi_{\bar{W},S}^C$ is used in such a way that the encoder and decoder’s initial belief $\pi_0$ is picked randomly according to $\hat{v}^*$ (but independently of $\{x_t\}$), then $\Pi^*$ is still optimal in the sense of Theorem 7.

**Remark 5.** We have not shown that an optimal stationary policy is deterministic. In the convex analytic approach, the existence of an optimal deterministic stationary policy directly follows if one can show that the extreme points of ergodic occupation measures satisfy the following: (i) They are induced by deterministic policies; and (ii) under these policies the state invariant measures are ergodic. This property of the extreme points of the set of ergodic occupation measures has been proved by Meyn in [23, Proposition 9.2.5] for countable state spaces and by Borkar in [7] and [3] for a specific case involving $\mathbb{R}^d$ as the state space and a non-degeneracy condition which amounts to having a density assumption on the one-stage transition kernels. Unfortunately, these approaches do not seem to apply in our setting.
2) The quadratic cost case: In the infinite horizon setting for the important case of the (unbounded) quadratic cost function, we add the following assumption, in addition to Assumption 3.

Assumption 5. The chain \( \{x_t\} \) is positive Harris recurrent with unique invariant measure \( \pi^* \) such that for some \( \epsilon > 0 \) and all \( x_0 \in \mathbb{R}^d \),

\[
\lim_{t \to \infty} E_{\delta x_0} [\|x_t\|^{2+\epsilon}] = E_{\pi^*} [\|x\|^{2+\epsilon}] < \infty.
\]

Remark 6. A sufficient condition for Assumption 5 to hold is that \( f \) in (3) satisfies

\[
\|f(x, w)\| \leq K (\|x\| + \|w\|) \quad \text{for some} \quad K < 1 \quad \text{and} \quad w_t \text{ has zero mean and finite } (2+\epsilon)\text{th moment } E [\|w_t\|^{2+\epsilon}] < \infty.
\]

In particular, the assumption holds for the LQG case \( x_{t+1} = Ax_t + w_t \), with \( A \) a \( d \times d \) matrix having eigenvalues of absolute value less than 1 and \( w_t \) having a nondegenerate Gaussian distribution with zero mean.

Theorem 9. Under Assumptions 3 and 5, there exists a stationary policy \( \Pi^* \) in \( \Pi_{W,S}^C \) that is optimal in the sense that with an appropriately chosen initial distribution \( \pi^*_0 \),

\[
\lim_{T \to \infty} E^{\Pi^*_0}_{\Pi_{W,S}^C} \left[ \frac{1}{T} \sum_{t=0}^{T-1} c(\pi_t, Q_t) \right] \leq \lim_{T \to \infty} \inf_{\Pi_{W,S}^C} E_{\delta x_0} \left[ \frac{1}{T} \sum_{t=0}^{T-1} c(\pi_t, Q_t) \right]
\]

for any \( x_0 \in \mathbb{R}^d \) and \( \Pi \in \Pi_{W,S}^C \). Furthermore, if \( \{x_t\} \) is started from the invariant distribution \( \pi^* \) and the optimal stationary policy \( \Pi^* \in \Pi_{W,S}^C \) is used in such a way that the encoder and decoder’s initial belief \( \pi_0 \) is picked randomly according to \( \hat{v}^* \) (but independently of \( \{x_t\} \)), then \( \Pi^* \) is still optimal in the above sense (with \( \pi^* \) replacing \( \pi^*_0 \)).

Proof: The proof is almost identical to that of Theorems 7 and 8, with the following minor adjustments, which are needed to accommodate the unboundedness of the quadratic cost function.

This modification is facilitated by Assumption 5 which implies that, similar to (53) and (56) in the proof of Proposition 1, for the sequence of expected occupation measures \( \{v_t\} \) corresponding to any initial distribution \( \delta x_0 \), we have

\[
\sup_{t \geq 0} \int_{P(\mathbb{R}^d) \times \mathcal{Q}_c} \left( \int_{\mathbb{R}^d} \|x\|^{2+\epsilon} \pi (dx) \right) v_t (d\pi dQ) < \infty,
\]

as well as for all \( v \in \mathcal{G} ,

\[
\int_{P(\mathbb{R}^d) \times \mathcal{Q}_c} \left( \int_{\mathbb{R}^d} \|x\|^{2+\epsilon} \pi (dx) \right) v (d\pi dQ) = \int_{\mathbb{R}^d} \|x\|^{2+\epsilon} \pi^* (dx) < \infty.
\]

These uniform integrability properties of \( \{v_t\} \) and \( \mathcal{G} \) allow us to use the continuity result Lemma 8 for \( c(\pi, Q) \). All other parts of the proof remain unchanged.
VI. CONCLUDING REMARKS

In this paper we established structural and existence results concerning optimal quantization policies for Markov sources. The key ingredient of our analysis was the characterization of quantizers as a subset of the space of stochastic kernels. This approach allows one to introduce a useful topology with respect to which the set of quantizers with a given number of convex codecells is compact, facilitating the proof of existence results. We note that both our assumption of convex-codecell quantizers and the more restrictive assumption of nearest neighbor-type quantizers in Borkar et al. [10] may preclude global optimality over all zero-delay quantization policies. The existence and finer structural characterization of such globally optimal policies are still open problems.

The existence and the structural results can be useful for the design of networked control systems where decision makers have imperfect observation of a plant to be controlled. The machinery presented here is particularly useful in the context of optimal quantized control of a linear system driven by unbounded noise: For LQG optimal control problems it has been shown that the effect of the control policies can be decoupled from the estimation error and the design results here can be used to establish existence of optimal quantization and control policies for LQG systems.

A further research direction is the formulation of the communication problem over a channel with feedback. The tools and the topological analysis developed in this paper could be useful in establishing optimal coding and decoding policies and the derivation of error-exponents with feedback. Relevant efforts in the literature on this topic include [29].

VII. APPENDIX

A. Auxiliary results

Recall that a sequence of probability measures \( \{\mu_n\} \) in \( \mathcal{P}(X) \) converges to \( \mu \in \mathcal{P}(X) \) weakly if \( \int_X c(x)\mu_n(dx) \to \int_X c(x)\mu(dx) \) for every continuous and bounded \( c : X \to \mathbb{R} \). For \( \mu, \nu \in \mathcal{P}(X) \) the total variation metric is defined by

\[
d_{TV}(\mu, \nu) := 2 \sup_{B \in \mathcal{B}(X)} |\mu(B) - \nu(B)|
\]

\[
= \sup_{g : \|g\|_{\infty} \leq 1} \left| \int g(x)\mu(dx) - \int g(x)\nu(dx) \right|, \tag{17}
\]

where the second supremum is over all measurable real functions \( g \) such that \( \|g\|_{\infty} := \sup_{x \in X} |g(x)| \leq 1 \).

**Definition 4** ([38]). Let \( P \in \mathcal{P}(\mathbb{R}^d) \). A quantizer sequence \( \{Q_n\} \) converges to \( Q \) weakly at \( P \) \((Q_n \to Q \text{ weakly at } P)\) if \( PQ_n \to PQ \text{ weakly} \). Similarly, \( \{Q_n\} \) converges to \( Q \) in total variation at \( P \) \((Q_n \to Q \text{ in total variation at } P)\) if \( PQ_n \to PQ \text{ in total variation} \).

The following lemma will be very useful in the upcoming optimality proofs.
Lemma 2. (a) Let \( \{\mu_n\} \) be a sequence of probability density functions on \( \mathbb{R}^d \) which are uniformly equicontinuous and uniformly bounded and assume \( \mu_n \to \mu \) weakly. Then \( \mu_n \to \mu \) in total variation.

(b) Let \( \{Q_n\} \) be a sequence in \( Q_c \) such that \( Q_n \to Q \) weakly at \( P \) for some \( Q \in Q_c \). If \( P \) admits a density, then \( Q_n \to Q \) in total variation at \( P \). If the density of \( P \) is positive, then \( Q_n \to Q \) in total variation at any \( P' \) admitting a density.

(c) Let \( \{Q_n\} \) be a sequence in \( Q_c \) such that \( Q_n \to Q \) weakly at \( P \) for some \( Q \in Q_c \) where \( P \) admits a positive density. Suppose further that \( P_n' \to P' \) in total variation where \( P_n' \) admits a density. Then \( P_n'Q_n \to P'Q \) in total variation.

Proof. (a) We will denote a density and its induced probability measure by the same symbol.

By the Arzelà-Ascoli theorem the sequence of densities \( \{\mu_n\} \), when restricted to a given compact subset of \( \mathbb{R}^d \), is relatively compact with respect to the supremum norm. Considering the sequence of increasing closed balls \( K_i = \{x : \|x\| \leq i\} \) of radius \( i = 1, 2, \ldots \), one can use Cantor’s diagonal argument as in [38, Lemma 4.3] to obtain a subsequence \( \{\mu_{nk}\} \) and a nonnegative continuous function \( \hat{\mu} \) such that \( \mu_{nk}(x) \to \hat{\mu}(x) \) for all \( x \), where the convergence is uniform over compact sets. Since \( \int_B |\mu_{nk}(x) - \hat{\mu}(x)| \, dx \to 0 \) for any bounded Borel set \( B \), and since \( \{\mu_n\} \) is tight by weak convergence, it follows that \( \hat{\mu} \) is a probability density. Since \( \mu_{nk} \) converges to \( \hat{\mu} \) pointwise, by Scheffe’s theorem [6] \( \mu_{nk} \) converges to \( \hat{\mu} \) in the \( L_1 \) norm, which is equivalent to convergence in total variation. Since \( \mu_n \to \mu \) weakly, we must have \( \mu = \hat{\mu} \).

The preceding argument implies that any subsequence of \( \{\mu_n\} \) has a further subsequence that converges to \( \mu \) in the (metric of) total variation. This implies that \( \mu_n \to \mu \) in total variation.

(b) It was shown in the proof of Theorem 5.7 of [38] that

\[
    d_{TV}(PQ_n, PQ) \leq \sum_{i=1}^M P(B_i^n \triangle B_i), \tag{18}
\]

where \( B_1^n, \ldots, B_M^n \) and \( B_1, \ldots, B_M \) are the cells of \( Q_n \) and \( Q \), respectively, and \( B_i^n \triangle B_i := (B_i^n \setminus B_i) \cup (B_i \setminus B_i^n) \). Since \( Q \) has convex cells, the boundary \( \partial B_i \) of each cell \( B_i \) has zero Lebesgue measure, so \( P(\partial B_i) = 0 \) because \( P \) has a density. Since \( \partial(B_i \times \{j\}) = \partial B_i \times \{j\} \), and \( PQ(A \times \{j\}) = P(A \cap B_j) \), we have

\[
PQ(\partial(B_i \times \{j\})) = P(\partial B_i \cap B_j) = 0,
\]

for all \( i \) and \( j \). Thus if \( PQ^n \to PQ \) weakly, then \( PQ^n(B_i \times \{j\}) \to PQ(B_i \times \{j\}) \) by the Portmanteau theorem, which is equivalent to

\[
P(B_i \cap B_j^n) \to P(B_i \cap B_j)
\]

for all \( i \) and \( j \). Since \( \{B_1^n, \ldots, B_M^n\} \) and \( \{B_1, \ldots, B_M\} \) are both partitions of \( \mathbb{R}^d \), this implies \( P(B_i^n \triangle B_i) \to 0 \) for all \( i \), which in turns proves that \( PQ^n \to PQ \) in total variation via (18).
Lemma 3. (a) \( P \) has a positive density and \( P' \) admits a density, then \( P' \) is absolutely continuous with respect to \( P \) and so \( P(B_i^n \cap B_i) \to 0 \) implies \( P'(B_i^n \cap B_i) \to 0 \). Combined with the preceding argument this proves the second statement in part (b).

(c) For any \( A \in \mathcal{B}(X \times M) \) let \( A(x) \coloneqq \{ y : (x, y) \in A \} \). Then

\[
|P_n'Q_n(A) - P'Q_n(A)| = \left| \int_{\mathbb{R}^d} Q_n(A(x)|x)P_n'(dx) - \int_{\mathbb{R}^d} Q_n(A(x)|x)P'(dx) \right|
\leq d_{TV}(P_n', P'),
\]

where the inequality is due to (17). Taking the supremum over all \( A \) yields

\[
d_{TV}(P_n'Q_n, P'Q_n) \leq d_{TV}(P_n', P').
\]

Hence

\[
d_{TV}(P_n'Q_n, P'Q) \leq d_{TV}(P_n'Q_n, P'Q_n) + d_{TV}(P'Q_n, P'Q)
\leq d_{TV}(P_n', P) + d_{TV}(P'Q_n, P'Q).
\]

From part (b) we know that \( Q_n \to Q \) in total variation at \( P' \). Since \( P_n' \to P \) in total variation, we obtain \( d_{TV}(P_n'Q_n, P'Q) \to 0 \).

Recall from Section II the definition of \( S \subset \mathcal{P}(\mathbb{R}^d) \) as the set of probability measures admitting densities that are uniformly bounded and uniformly Lipschitz (with constants determined by the conditional density \( \phi(\cdot|x) \) of \( x_{t+1} = f(x_t, w_t) \) given \( x_t = x \)). In Lemma 1 we showed that \( S \) contains all reachable states, i.e., \( \pi_t \in S \) for all \( t \geq 1 \) with probability 1 under any policy \( \Pi \in \Pi_{TV} \).

Lemma 2(a) immediately implies that for any sequence \( \{\mu_n\} \) in \( S \) and \( \mu \in S \), \( \mu_n \to \mu \) weakly if and only if \( \mu_n \to \mu \) in total variation. In this case we simply say that \( \{\mu_n\} \) converges to \( \mu \) in \( S \).

As discussed in Section II, we can define the (quotient) topology on \( Q_c \) induced by weak convergence of sequences at a given \( P \) admitting a positive density. Lemma 2(b) implies that any sequence in \( Q_c \) converging in this topology will converge both weakly and in total variation at any \( P' \) admitting a density. To say that \( \{Q_n\} \) converges in \( Q_c \) will mean convergence in this topology. As well, we equip \( S \times Q_c \) with the corresponding product topology and continuity of any \( F : S \times Q_c \to \mathbb{R} \) will be meant in this sense.

**Lemma 3.** (a) \( S \) is closed in \( \mathcal{P}(\mathbb{R}^d) \).
(b) \( Q_c \) is compact.
(c) If \( \{\mu_n, Q_n\} \) converges in \( S \times Q_c \) to \( (\mu, Q) \in S \times Q_c \), then \( \mu_nQ_n \to \mu Q \) in total variation.

Thus any \( F : S \times Q_c \to \mathbb{R} \) is continuous if \( F(\mu_n, Q_n) \to F(\mu, Q) \) whenever \( \mu_nQ_n \to \mu Q \) in total variation.

**Proof:** (a) Recall that \( S \) is a uniformly bounded and uniformly equicontinuous family of densities. Lemma 2(a) shows that if \( \{\mu_n\} \) is a sequence in \( S \) and \( \mu_n \to \mu \) weakly, then \( \mu \) has
a density. The proof also shows that some subsequence of (the densities of) \( \{\mu_n\} \) converges to (the density of) \( \mu \). Thus \( \mu \) must admit the same uniform upper bound and Lipschitz constant as all densities in \( S \), proving that \( \mu \in S \).

(b) The compactness of \( Q_c \) was shown in [38, Thm. 5.8].

(c) If \( \{(\mu_n, Q_n)\} \) converges in \( S \times Q_c \) to \( (\mu, Q) \in S \times Q_c \) then \( \mu_n \to \mu \) in total variation. Since \( \mu \) has a density, \( Q_n \to Q \) in \( Q_c \) implies that \( Q_n \to Q \) in total variation at \( \mu \). Thus \( \mu_n Q_n \to \mu Q \) in total variation by Lemma 2(c).

B. Proof of Theorem 4

The first statement of the following theorem immediately implies Theorem 4.

**Theorem 10.** For \( t = T - 1, \ldots, 0 \) define the value function \( J_T^t \) at time \( t \) recursively by

\[
J_T^t(\pi) = \inf_{Q \in Q_c} \left( \frac{1}{T} c(\pi, Q) + E[J_{t+1}^T(\pi_{t+1}) | \pi_t = \pi, Q_t = Q] \right)
\]

with \( J_T^T := 0 \) and \( c(\pi, Q) \) defined in (5). Then for any \( t \geq 1 \) and \( \pi \in S \) or \( t = 0 \) and \( \pi \in S \cup \{\pi_0\} \), the infimum is achieved by some \( Q \) in \( Q_c \). Moreover, \( J_T^t(\pi) \) is continuous on \( S \).

The rest of this section is devoted to proving Theorem 10. The proof is through a series of lemmas that show the continuity of both \( c(\pi, Q) \) and \( E[J_{t+1}^T(\pi_{t+1}) | \pi_t = \pi, Q_t = Q] \) in \( (\pi, Q) \).

**Lemma 4.** \( c(\pi, Q) \) is continuous on \( S \times Q_c \).

**Proof:** If \( \{(\pi_n, Q_n)\} \) converges in \( S \times Q_c \) then \( \pi_n Q_n \to \pi Q \) in total variation by Lemma 3(c). We have to show that in this case

\[
c(\pi_n, Q_n) = \inf_{\gamma} \int_{\mathbb{R}^d} \pi_n(d\gamma) \sum_{i=1}^M Q_n(i|x) c_0(x, \gamma(i))
\]

\[
\to \inf_{\gamma} \int_{\mathbb{R}^d} Q(d\gamma) \sum_{i=1}^M Q(i|x) c_0(x, \gamma(i)) = c(\pi, Q).
\]

This follows verbatim from the proof of [38, Thm. 3.4] where for any bounded \( c_0 \) the convergence for a fixed \( \pi \) and \( Q_n \to Q \) was shown.

We now start proving Theorem 10. At \( t = T - 1 \) we have

\[
J_{T-1}^T(\pi) = \inf_{Q \in Q_c} c(\pi, Q).
\]

By Lemma 4 and the compactness of the set of quantizers \( Q_c \) (Lemma 3(b)) there exists an optimal quantizer that achieves the infimum. The following lemma will be useful.

**Lemma 5.** If \( F : S \times Q_c \to \mathbb{R} \) is continuous then \( \inf_{Q \in Q_c} F(\pi, Q) \) is achieved by some \( Q \) in \( Q_c \) and \( \min_Q F(\pi, Q) \) is continuous in \( \pi \) on \( S \).
Proof. The existence of an optimal \( Q \) in \( Q_c \) achieving \( \inf_{Q \in \mathcal{Q}} F(\pi, Q) \) is a consequence of the continuity of \( F \) and the compactness of \( Q_c \). Assume \( \pi_n \to \pi \) in \( \mathcal{S} \) and let \( Q_n \) be optimal for \( \pi_n \) and \( Q \) optimal for \( \pi \). Then
\[
\left| \min_{Q'} F(\pi_n, Q') - \min_{Q'} F(\pi, Q') \right| \\
\leq \max \left( F(\pi_n, Q) - F(\pi, Q), F(\pi, Q_n) - F(\pi_n, Q_n) \right).
\] (19)
The first term in the maximum converges to zero since \( F \) is continuous. To show that the second converges to zero, suppose to the contrary that for some \( \epsilon > 0 \) and for a subsequence \( \{ (\pi_{n_k}, Q_{n_k}) \} \),
\[
|F(\pi, Q_{n_k}) - F(\pi_{n_k}, Q_{n_k})| \geq \epsilon.
\] (20)
By Lemma 3(b), there is a further subsequence \( \{ Q_{n_k'} \} \) converging to some \( Q' \) in \( Q_c \). Then \( \{ (\pi, Q_{n_k'}) \} \) and \( \{ (\pi_{n_k'}, Q_{n_k'}) \} \) both converge to \( (\pi, Q') \), which contradicts (20) since \( F \) is continuous.

As a consequence of Lemmas 4 and 5, \( J^T_{T-1}(\pi) \) is continuous on \( \mathcal{S} \).

Let \( t = T - 2 \). We want to show that the minimization problem
\[
J^T_{T-2}(\pi) = \min_{Q \in \mathcal{Q}_c} \left( \frac{1}{T} c(\pi, Q) + E\left[J^T_{T-1}(\pi_{T-1})|\pi_{T-2} = \pi, Q_{T-2} = Q\right] \right)
\] (21)
has a solution and \( J^T_{T-2}(\pi) \) is continuous on \( \mathcal{S} \).

Consider the conditional probability distributions given by
\[
\hat{\pi}(m, \pi, Q)(C) := P(x_{t+1} \in C|\pi_t = \pi, Q_t = Q, q_t = m) \\
= \frac{1}{\pi(Q^{-1}(m))} \int_{\mathcal{C}} \left( \int_{\mathbb{R}^d} \pi(dx)1_{\{x \in B_m\}} \phi(z|x) \right) dz
\] (22)
(if \( \pi(Q^{-1}(m)) = 0 \), then \( \hat{\pi}(m, \pi, Q) \) is set arbitrarily). Note that
\[
E[J^T_{T-1}(\pi_{T-1})|\pi_{T-2} = \pi, Q_{T-2} = Q] \\
= \sum_{m=1}^{M} J^T_{T-1}(\hat{\pi}(m, \pi, Q)) \pi(Q^{-1}(m)),
\] (23)
where
\[
\pi(Q^{-1}(m)) = P(q_{T-2} = m|\pi_{T-2} = \pi, Q_{T-2} = Q).
\]

The following lemma will imply that if \( (\pi_n, Q_n) \to (\pi, Q) \) in \( \mathcal{S} \times Q_c \), then
\[
J^T_{T-1}(\hat{\pi}(m, \pi_n, Q_n)) \pi_n(Q^{-1}_n(m)) \to J^T_{T-1}(\hat{\pi}(m, \pi, Q)) \pi(Q^{-1}(m))
\] (24)
for all \( m \).

Lemma 6. If \( \pi_nQ_n \to \pi Q \) in total variation, then \( \hat{\pi}(m, \pi_n, Q_n) \to \hat{\pi}(m, \pi, Q) \) in total variation for every \( m = 1, \ldots, M \) with \( \pi(Q^{-1}(m)) > 0 \).
Proof. Let \( B_1, \ldots, B_M \) and \( B_1^n, \ldots, B_M^n \) denote the cells of \( Q \) and \( Q_n \), respectively. Since for any Borel set \( A \), \( \pi_n Q_n(A \times \{ j \}) = \pi_n(A \cap B_j^n) \), the convergence of \( \pi_n Q_n \) to \( \pi Q \) implies that \( \pi_n(A \cap B_j^n) \to \pi(A \cap B_j) \). This implies \( \pi_n(B_i \cap B_j^n) \to \pi(B_i \cap B_j) \) for all \( i \) and \( j \), from which we obtain

\[
\pi_n(B_m^n) \to \pi(B_m), \quad \pi_n(B_m^n \triangle B_m) \to 0, \quad m = 1, \ldots, M. \tag{25}
\]

If \( \pi(B_m) > 0 \), the probability distribution \( \hat{\pi}(m, \pi, Q) \) has density

\[
\hat{\pi}(m, \pi, Q)(z) = \frac{1}{\pi(B_m)} \int_{B_m} \pi(dx) \phi(z|x)
\]

so by Scheffé’s theorem [6] it suffices to show that \( \hat{\pi}(m, \pi_n, Q_n)(z) \to \hat{\pi}(m, \pi, Q)(z) \) for all \( z \).

As \( \pi(B_m) > 0 \) by assumption and \( \pi_n(B_m^n) \to \pi(B_m) \), it is enough to establish the convergence of \( v_n^m(z) := \int_{B_m^n} \pi_n(dx) \phi(z|x) \) to \( v^m(z) := \int_{B_m} \pi(dx) \phi(z|x) \).

For any \( z \in \mathbb{R}^d \) we have

\[
|v_n^m(z) - v^m(z)| \leq \left| \int_{\mathbb{R}^d} \pi_n(dx) \left( 1_{\{x \in B_m^n\}} - 1_{\{x \in B_m\}} \right) \phi(z|x) \right|
+ \left| \int_{\mathbb{R}^d} 1_{\{x \in B_m\}} \left( \pi_n(x) - \pi(x) \right) \phi(z|x) dx \right|
= \int_{B_m^n \triangle B_m} \pi_n(dx) \phi(z|x) + \int_{\mathbb{R}^d} \pi_n(x) - \pi(x) \phi(z|x) dx
\leq C \left[ \pi_n(B_m^n \triangle B_m) + d_{TV}(\pi_n, \pi) \right], \tag{26}
\]

where \( C \) is a uniform upper bound on \( \phi \). Since both terms in the brackets converge to zero as \( n \to \infty \), the proof is complete.

Now if \( (\pi_n, Q_n) \to (\pi, Q) \) in \( S \times Q_c \), then by Lemma 3 and (25) we have \( \pi_n(Q_n^{-1}(m)) \to \pi(Q^{-1}(m)) \) for all \( m \). If \( \pi(m, \pi, Q) > 0 \) for some \( m \), then Lemma 6 implies that \( \hat{\pi}(m, \pi_n, Q_n) \to \hat{\pi}(m, \pi, Q) \) in total variation; hence (24) holds in this case by the continuity of \( J_{T-1}^T \). If \( \pi(Q^{-1}(m)) = 0 \), then by (25) and the boundedness of the cost (24) holds again. In view of (23), we obtain that \( E[J_{T-1}^T(\pi_{T-1})|\pi_{T-2} = \pi, Q_{T-2} = Q] \) is continuous on \( S \times Q_c \).

We have shown that both expressions on the right side of (21) are continuous on \( S \times Q_c \). By Lemma 5 the minimization problem (21) has a solution and \( J_{T-2}^T(\pi) \) is continuous on \( S \). The recursion applies for all further stages \( t = T - 3, \ldots, 1 \), without change since \( \pi_t \in S \) for all \( t \geq 1 \) under any policy. If \( \pi_0 \) admits a density, then at the last stage \( t = 0 \) there exists a minimizing \( Q \) for

\[
J_0^T(\pi_0) = \inf_{Q \in Q_c} \left( \frac{1}{T} c(\pi_0, Q) + E[J_1^T(\pi_1)|\pi_0, Q_t = Q] \right)
\]

since the preceding proofs readily imply that both \( c(\pi_0, Q) \) and \( E[J_1^T(\pi_1)|\pi_0, Q_t = Q] \) are continuous in \( Q \) as long as \( \pi_0 \) admits a density. If \( \pi_0 \) is a point mass on \( x_0 \), then any \( Q \) is optimal. This establishes Theorem 10. \qed
C. Proof of Theorem 5

The first statement of the following counterpart of Theorem 10 immediately implies Theorem 5.

**Theorem 11.** Consider Assumption 3. For \( t = T - 1, \ldots, 0 \) define the value function \( J_t^T \) at time \( t \) recursively by

\[
J_t^T(\pi) = \inf_{Q \in Q_c} \left( \frac{1}{T} c(\pi, Q) + E\left[ J_{t+1}^T(\pi_{t+1}) | \pi_t = \pi, Q_t = Q \right] \right)
\]

with \( J_T^T := 0 \) and \( c(\pi, Q) \) defined in (5). Then for any \( t \geq 1 \) and \( \pi \in S \) or \( t = 0 \) and \( \pi \in S \cup \{ \pi_0 \} \) the infimum is achieved by some \( Q \) in \( Q_c \).

Moreover, \( J_t^T(\pi) \) is continuous on \( S \) in the sense that if \( \pi_n \to \pi \) and \( \{ \pi_n \} \) satisfies the uniform integrability condition

\[
\lim_{L \to \infty} \sup_{n \geq 1} \int_{\{ \|x\|^2 \geq L \}} \|x\|^2 \pi_n(dx) = 0,
\]

then \( J_t^T(\pi_n) \to J_t^T(\pi) \).

To prove Theorem 11 we need to modify the proof of Theorem 10 only in view of the unboundedness of the cost, which affects the proof of the continuity of \( c(\pi, Q) \) and \( E\left[ J_{t+1}^T(\pi_{t+1}) | \pi_t = \pi, Q_t = Q \right] \).

We first establish the continuity of \( c(\pi, Q) \) in a more restricted sense than in Lemma 4. We know from (8) that given \( \pi_t = \pi \) and \( Q_t = Q \) with cells \( B_1, \ldots, B_M \), the unique optimal receiver policy is given, for any \( m \) such that \( \pi(B_m) > 0 \), by

\[
\gamma(m) = \int_{B_m} x \pi(dx).
\]

If \( \pi(B_m) = 0 \), then \( \gamma(m) \) is arbitrary. Using this optimal receiver policy, define \( \bar{Q} : \mathbb{R}^d \to \mathbb{R}^d \) by

\[
\bar{Q}(x) = \gamma(Q(x)).
\]

Note that \( c(\pi, Q) = E_\pi \left[ \|x - \bar{Q}(x)\|^2 \right] \) and that for all \( m \),

\[
E_\pi \left[ \|x - \bar{Q}(x)\|^2 1_{\{x \in B_m\}} \right] = \int_{B_m} \|x - \gamma(m)\|^2 \pi(dx) \leq \int_{B_m} \|x\|^2 \pi(dx)
\]

which implies

\[
c(\pi, Q) \leq \int \|x\|^2 \pi(dx) = E_\pi \left[ \|x\|^2 \right]. \tag{29}
\]

**Lemma 7.** Assume \((\pi_n, Q_n) \to (\pi, Q)\) in \( S \times Q_c \) and \( \{ \pi_n \} \) satisfies the uniform integrability condition (27). Then \( c(\pi_n, Q_n) \to c(\pi, Q) \).
We have $\pi_n(B_m^n) \rightarrow \pi(B_m)$ and $\pi_n(B_m^n \triangle B_m) \rightarrow 0$. Let $I = \{m \in \{1, \ldots, M\} : \pi(B_m) > 0\}$. We have for any $L > 0$ and $m \in I$,

$$
\int_{B_m^n} x1_{\{\|x\|^2 < L\}} \pi_n(dx) \rightarrow \int_{B_m} x1_{\{\|x\|^2 < L\}} \pi(dx).
$$

This and a standard truncation argument that makes use of (27) imply

$$
\int_{B_m^n} x \pi_n(dx) \rightarrow \int_{B_m} x \pi(dx), \quad m \in I
$$

so the optimal receiver policy $\gamma_n$ for $Q_n$ satisfies $\gamma_n(m) \rightarrow \gamma(m)$ for all $m \in I$. In particular, this implies that for all $m \in I$,

$$
D_m := \sup_{n \geq 1} \sup_{x \in B_m^n} \|\bar{Q}_n(x)\|^2 < \infty.
$$

In turn, the parallelogram law gives for $m \in I$ and $x \in B_m^n$,

$$
\|x - \bar{Q}_n(x)\|^2 \leq 2\|x\|^2 + 2\|\bar{Q}_n(x)\|^2 \leq 2\|x\|^2 + 2D_m
$$

so for $m \in I$ we obtain

$$
\lim_{L \rightarrow \infty} \sup_{n \geq 1} \int_{B_m^n} \pi_n(dx) \|x - \bar{Q}_n(x)\|^2 1_{\{\|x\|^2 \geq L\}} \leq \lim_{L \rightarrow \infty} \sup_{n \geq 1} \int_{\mathbb{R}^d} \pi_n(dx) \left(2\|x\|^2 + 2D_m\right) 1_{\{\|x\|^2 \geq L\}} = 0,
$$

(30)

where the second limit is zero due to (27).

Since $\pi_n \rightarrow \pi$ in total variation, $\pi_n(B_m^n) \rightarrow \pi(B_m)$ and $\pi_n(B_m^n \triangle B_m) \rightarrow 0$, and since $\|x - \bar{Q}_n(x)\|^2$ is uniformly bounded if $\|x\|^2 < L$, we have

$$
\int_{B_m^n} \pi_n(dx) \|x - \bar{Q}_n(x)\|^2 1_{\{\|x\|^2 < L\}} \rightarrow \int_{B_m} \pi(dx) \|x - \bar{Q}(x)\|^2 1_{\{\|x\|^2 < L\}}.
$$

Then uniform integrability (30) and a standard truncation argument yield for $m \in I$

$$
\int_{B_m^n} \pi_n(dx) \|x - \bar{Q}_n(x)\|^2 \rightarrow \int_{B_m} \pi(dx) \|x - \bar{Q}(x)\|^2.
$$

(31)

Assume $m \notin I$. Then we have

$$
\int_{B_m^n} \|x\|^2 \pi_n(dx) \leq \int_{\{\|x\|^2 \geq L\}} \|x\|^2 \pi_n(dx) + L\pi_n(B_m^n) \rightarrow 0
$$

from (27) and since $\pi_n(B_m^n) \rightarrow 0$. In view of (28) we obtain

$$
\int_{B_m^n} \pi_n(dx) \|x - \bar{Q}_n(x)\|^2 \rightarrow 0.
$$

This and (31) give

$$
c(\pi_n, Q_n) = \int_{\mathbb{R}^d} \pi_n(dx) \|x - \bar{Q}_n(x)\|^2 \rightarrow \int_{\mathbb{R}^d} \pi(dx) \|x - \bar{Q}(x)\|^2 = c(\pi, Q)
$$

May 7, 2014
Lemma 9. Assume $F : S \times Q_c \to \mathbb{R}$ is continuous in the sense that if $(\pi_n, Q_n) \to (\pi, Q)$ in $S \times Q_c$ and $\{\pi_n\}$ satisfies the uniform integrability condition (27), then $F(\pi_n, Q_n) \to F(\pi, Q)$. Then

$$\inf_{Q \in Q_c} F(\pi, Q)$$

is achieved by some $Q$ in $Q_c$ and $\min_Q F(\pi, Q)$ is continuous in $\pi$ in the sense that if $\pi_n \to \pi$ in $S$ and $\{\pi_n\}$ is uniformly integrable, then $\min_Q F(\pi_n, Q) \to \min_Q F(\pi, Q)$.

**Proof.** The existence of an optimal $Q$ for any $\pi \in S$ is a consequence of the compactness of $Q_c$. The rest of the proof follows verbatim the proof of Lemma 5 with the convergence sequence $\{\pi_n\}$ also assumed to be uniformly integrable.

Lemmas 7 and 8 prove Theorem 11 for $t = T - 1$. To prove the theorem for all $t$, we apply backward induction. Assume that the both statements of the theorem hold for $t' = T-1, \ldots, t+1$.

Recall the conditional distribution $\hat{\pi}(m, \pi, Q)$ defined in (22). The following lemma shows that the uniform integrability condition is inherited in the induction step.

**Lemma 9.** Assume $(\pi_n, Q_n) \to (\pi, Q)$ in $S \times Q_c$ and $\{\pi_n\}$ satisfies the uniform integrability condition (27). If cell $B_m$ of $Q$ satisfies $\pi(B_m) > 0$, then $\{\hat{\pi}(m, \pi_n, Q_n)\}$ is uniformly integrable in the sense of (27).

**Proof.** Let $B_m^n$ denote the $n$th cell of $Q_n$. Since $\pi(B_m) > 0$, we have $\pi_n(B_m^n) > 0$ for $n$ large enough, so

$$\int_{\mathbb{R}^d} \|z\|^2 \hat{\pi}(m, \pi_n, Q_n)(z) 1\{\|z\|^2 \geq L\} \, dz$$

$$\leq \frac{1}{\pi(B_m^n)} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \|x\|^2 + \|w\|^2 \left(1\{\|x\|^2 \geq L/(4K^2)\} + 1\{\|w\|^2 \geq L/(4K^2)\}\right) \pi_n(dx) \nu_w(dw).$$

Since $\|f(x, w)\|^2 \leq 2K^2(\|x\|^2 + \|w\|^2)$ by Assumption 3(i), we have

$$1\{\|f(x, w)\|^2 \geq L\} \leq 1\{\|x\|^2 \geq L/(4K^2)\} + 1\{\|w\|^2 \geq L/(4K^2)\}$$

and so

$$\int_{\mathbb{R}^d} \|z\|^2 \hat{\pi}(m, \pi_n, Q_n)(z) 1\{\|z\|^2 \geq L\} \, dz$$

$$\leq \frac{1}{\pi(B_m^n)} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} 2K^2(\|x\|^2 + \|w\|^2) \left(1\{\|x\|^2 \geq L/(4K^2)\} + 1\{\|w\|^2 \geq L/(4K^2)\}\right) \pi_n(dx) \nu_w(dw)$$

$$= \frac{2K^2}{\pi(B_m^n)} \int_{\mathbb{R}^d} \|x\|^2 1\{\|x\|^2 \geq L/(4K^2)\} \pi_n(dx)$$

$$+ \frac{2K^2}{\pi(B_m^n)} \left(\int_{\mathbb{R}^d} \|x\|^2 \pi_n(dx)\right) \left(\int_{\mathbb{R}^d} 1\{\|w\|^2 \geq L/(4K^2)\} \nu_w(dw)\right)$$

$$+ \frac{2K^2}{\pi(B_m^n)} \left(\int_{\mathbb{R}^d} \|w\|^2 \nu_w(dw)\right) \left(\int_{\mathbb{R}^d} 1\{\|x\|^2 \geq L/(4K^2)\} \pi_n(dx)\right)$$

(32)
\[ + \frac{2K^2}{\pi(B^*_m)} \int \mathbb{R}^d \|w\|^2 \mathbb{1}_{\{\|w\|^2 \geq L/(4K^2)\}} \nu_w(dw). \] (35)

Recall that \( \pi_n(B^*_m) \to \pi(B_m) \). Thus (32) converges to zero as \( L \to \infty \) uniformly in \( n \) by (27). The uniform convergence to zero of each of (33), (34), and (35) follows since \( \int \|w\|^2 \nu_w(dw) < \infty \) and \( \sup_{n \geq 1} \int \|x\|^2 \pi_n(x) \, dx < \infty \) by (27). This proves
\[
\lim_{L \to \infty} \sup_{n \geq 1} \int_{\mathbb{R}^d} \|z\|^2 \hat{\pi}(m, \pi_n, Q_n)(z) 1_{\{\|z\|^2 \geq L\}} \, dz = 0
\]
as claimed. \( \square \)

The next lemma shows the continuity of \( E[J_{t+1}^T(\pi_t) | \pi = \pi_t, Q_t = Q] \).

**Lemma 10.** \( E[J_{t+1}^T(\pi_t) | \pi_t = \pi, Q_t = Q] \) is continuous on \( S \times Q_c \) in the sense of Lemma 8.

**Proof.** Assume \((\pi_n, Q_n) \to (\pi, Q)\) in \( S \times Q_c \) and \( \{\pi_n\} \) satisfies the uniform integrability condition (27). Let \( B_1, \ldots, B_M \) and \( B^n_1, \ldots, B^n_M \) denote the cells of \( Q \) and \( Q_n \), respectively. In view of (23) and the fact that \( \pi_n(B^*_m) \to \pi(B_m) \), we need to prove that for all \( m \) with \( \pi(B_m) > 0 \),
\[
J_{t+1}^T(\hat{\pi}(m, \pi_n, Q_n)) \to J_{t+1}^T(\hat{\pi}(m, \pi, Q)) \quad (36)
\]
and for \( m \) with \( \pi(B_m) = 0 \),
\[
J_{t+1}^T(\hat{\pi}(m, \pi_n, Q_n)) \pi_n(B^*_m) \to 0. \quad (37)
\]

The convergence in (36) follows from Lemmas 6 and 9, and the induction hypothesis that \( J_{t+1}^T(\cdot) \) is continuous along convergent and uniformly integrable sequences in \( S \).

To prove (37) first note that from (29) we have
\[
J_{t+1}^T(\pi_{t+1}) \leq E \left[ \frac{1}{T} \sum_{i=t+1}^{T-1} \|x_i\|^2 \right],
\]
where \( x_{t+1} \) has distribution \( \pi_{t+1} \) and \( x_i = f(x_{i-1}, w_{i-1}) \), where \( w_{t+1}, \ldots, w_{T-1} \) are independent of \( x_{t+1} \). Accordingly,
\[
J_{t+1}^T(\hat{\pi}(m, \pi_n, Q_n)) \pi_n(B^*_m) \leq E \left[ \frac{1}{T} \left( \sum_{i=t+1}^{T-1} \|x_{i,n}\|^2 \right) 1_{\{x_{i,n} \in B^*_m\}} \right], \quad (38)
\]
where \( x_{i,n} \) has distribution \( \pi_n \).

Now note that the assumption \( \|f(x, w)\| \leq K(\|x\| + \|w\|) \) and the inequality \( \|x + y\|^2 \leq 2\|x\|^2 + 2\|y\|^2 \) imply the upper bound
\[
\|x_{t+j,n}\|^2 \leq (2K^2)^j \|x_{t,n}\|^2 + \sum_{i=0}^{j-1} (2K^2)^{j-i} \|w_{t+i}\|^2. \quad (39)
\]

Thus for any \( j = 1, \ldots, T - t - 1 \) we have
\[
E[\|x_{t+j,n}\|^2 1_{\{x_{i,n} \in B^*_m\}}]
\]
to zero, finishing the proof of the lemma. For the last step

\[ \lim_{t \to \infty} E \left[ \left\| x_{t,n} \right\|^2 1_{\{x_{t,n} \in B_m^0\}} \right] = \left(2K^2\right)^j E \left[ \left\| x_{t,n} \right\|^2 1_{\{x_{t,n} \in B_m^0\}} \right] + \sum_{i=0}^{j} (2K^2)^{1-i} \left[ \left\| w_{t+i-1} \right\|^2 \right] \pi_n(B_m^0), \]  

(40)

where we used the independence of \( w_t, \ldots, w_{T-1} \) and \( x_{t,n} \). The first expectation in (40) converges to zero as \( n \to \infty \) since \( \{\pi_n\} \) is uniformly integrable and \( \pi_n(B_m^0) \to \pi(B_m) = 0 \), while the second one converges to zero since \( \pi_n(B_m^0) \to 0 \). This proves that the right side of (38) converges to zero, finishing the proof of the lemma. \( \square \)

Lemmas 7 and 10 show that

\[ F_t(\pi, Q):= \frac{1}{T} c(\pi, Q) + E \left[ J_{t+1}^T(\pi_{t+1}) \big| \pi_t = \pi, Q_t = Q \right] \]

satisfies the conditions of Lemma 8, which in turn proves the induction hypothesis for \( t' = t \). For the last step \( t = 0 \) a similar argument as in the proof of Theorem 10 applies (but here we also need the condition \( E_{\pi_0} \left[ \left\| x_0 \right\|^2 \right] < \infty \)). This finishes the proof of Theorem 11. \( \square \)

D. Proof of Theorem 6

Define

\[ J^{\pi^*}(T) := \inf_{\Pi \in \Pi_A} \inf_{\gamma} E_{\pi^*}^\Pi \left[ \frac{1}{T} \sum_{t=0}^{T-1} c_0(x_t, u_t) \right] \]

and note that \( \limsup_{T \to \infty} J^{\pi^*}(T) \leq J^{\pi^*} \). Thus there exists an increasing sequence of time indices \( \{T_k\} \) such that for all \( k = 1, 2, \ldots \),

\[ J^{\pi^*}(T_k) \leq J^{\pi^*} + \frac{1}{k}. \]  

(41)

A key observation is that by Theorem 2 for all \( k \) there exists \( \Pi_k = \{\hat{\eta}_t^{(k)}\} \in \Pi_W \) (a Markov policy) such that

\[ J^{\pi^*}(\Pi_k, T_k) := E_{\pi^*}^\Pi \left[ \frac{1}{T_k} \sum_{t=0}^{T_k-1} c(\pi_t, Q_t) \right] \leq J^{\pi^*}(T_k) + \frac{1}{k}. \]  

(42)

Now let \( n_1 = 1 \) and for \( k = 2, 3, \ldots \), choose the positive integers \( n_k \) inductively as

\[ n_k = \left[ k \cdot \max \left( \frac{T_{k+1}}{T_k}, \frac{n_{k-1}T_k}{T_{k-1}} \right) \right], \]

(43)

where \( \left[ x \right] \) denotes the smallest integer greater than equal to \( x \). Letting \( T_k' = n_kT_k \) for all \( k \), we have \( T_k' \geq kT'_{k-1} \), and hence

\[ \lim_{k \to \infty} \sum_{l=0}^{k-1} T_k' = 1. \]  

(44)

Now let \( N_0 = 0, N_k = \sum_{t=1}^{k} T_k' \) for \( k \geq 1 \), and define the policy \( \Pi = \{\hat{\eta}_t\} \in \Pi_W \) by piecing together, in a periodic fashion, the initial segments of \( \Pi_k \) as follows:
(1) For \( t = N_{k-1} + jT_k \), where \( k \geq 1 \) and \( 0 \leq j < n_k \), let \( \hat{\eta}_t (\cdot) \equiv \hat{\eta}^{(k)}_0 (\pi^*) \).

(2) For \( t = N_{k-1} + jT_k + i \), where \( k \geq 1 \), \( 0 \leq j < n_k \), and \( 1 \leq i < T_k \), let \( \hat{\eta}_t = \hat{\eta}^{(k)}_i \).

In the rest of the proof we show that \( \Pi \) is optimal. First note that by the stationarity of \( \{x_t\} \) we have, for all \( k \geq 1 \) and \( j = 0, \ldots, n_k - 1 \),

\[
E_{\pi^*}[\sum_{t=N_{k-1}+jT_k}^{N_{k-1}+(j+1)T_k-1} c(\pi_t, Q_t)] = T_kJ_{\pi^*}(\Pi_k, T_k).
\]

Hence, for \( T = N_{k-1} + jT_k + i \), where \( k \geq 3 \), \( 0 \leq j < n_k \), and \( 0 \leq i < T_k \), we have

\[
E_{\pi^*}[\frac{1}{T} \sum_{t=0}^{T-1} c(\pi_t, Q_t)] = E_{\pi^*}[\frac{1}{T} \sum_{t=0}^{N_{k-2}-1} c(\pi_t, Q_t)] + E_{\pi^*}[\frac{1}{T} \sum_{t=N_{k-2}}^{T-1} c(\pi_t, Q_t)]
\]

\[
= \frac{1}{T} \sum_{t=1}^{k-2} T_t' J_{\pi^*}(\Pi_t, T_t) + \frac{1}{T} \left( T'_k \sum_{t=1}^{k-1} T_t' J_{\pi^*}(\Pi_{k-1}, T_{k-1}) + jT_k J_{\pi^*}(\Pi_k, T_k) \right) + E_{\pi^*}[\frac{1}{T} \sum_{t=N_{k-1}+jT_k}^{T-1} c(\pi_t, Q_t)]
\]

(45)

(46)

(the last sum is empty if \( i = 0 \)).

Let \( \hat{C} \) be a uniform upper bound on the cost \( c_0 \). Since \( T \geq N_{k-1} \), the first term in (45) can be bounded as

\[
\frac{1}{T} \sum_{t=1}^{k-2} T_t' J_{\pi^*}(\Pi_t, T_t) \leq \hat{C} \frac{1}{N_{k-1}} \sum_{t=1}^{k-2} T_t' = \hat{C} \frac{N_{k-2}}{N_{k-1}}
\]

\[
= \hat{C} \frac{N_{k-2}}{T_{k-2}} + \frac{T_{k-1}}{T_{k-2}} \to 0 \quad \text{as} \quad k \to \infty
\]

(47)

since \( \frac{N_{k-2}}{T_{k-2}^2} \to 1 \) from (44) and \( \frac{T_{k-1}}{T_{k-2}^2} \geq k - 1 \) from (43).

Since \( T_{k-1} + jT_k \leq T \), the second term in (45) can be upper bounded as

\[
\frac{1}{T} \left( T'_k \sum_{t=1}^{k-1} T_t' J_{\pi^*}(\Pi_{k-1}, T_{k-1}) + jT_k J_{\pi^*}(\Pi_k, T_k) \right)
\]

\[
\leq \max \left( J_{\pi^*}(\Pi_{k-1}, T_{k-1}), J_{\pi^*}(\Pi_k, T_k) \right).
\]

(48)

Finally, the expectation in (46) is upper bounded as

\[
E_{\pi^*}[\frac{1}{T} \sum_{t=N_{k-1}+jT_k}^{T-1} c(\pi_t, Q_t)] \leq \hat{C} \frac{T_k}{T} \leq \hat{C} \frac{T_k}{T_{k-1}} \leq \hat{C} \frac{1}{k-1} \to 0 \quad \text{as} \quad k \to \infty,
\]

(49)

where the last inequality is due to (43).
Combining (45)–(48) we obtain
\[
\limsup_{T \to \infty} E^\Pi \left[ \frac{1}{T} \sum_{t=0}^{T-1} c(\pi_t, Q_t) \right] \leq \limsup_{k \to \infty} J_{\pi^*}(\Pi_k, T_k) \leq J_{\pi^*}
\]
which proves the optimality of \( \Pi \).

E. Proof of Proposition 1

Proof of (a). Here we show that any weak limit of \( \{v_t\} \) must belong to \( G \). For \( v \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d) \times \mathcal{Q}_c) \) and \( g \in C_b(\mathcal{P}(\mathbb{R}^d) \times \mathcal{Q}_c) \) or \( f \in C_b(\mathcal{P}(\mathbb{R}^d)) \) define
\[
\langle v, g \rangle := \int g(\pi, Q) v(d\pi dQ), \quad \langle v, f \rangle := \int f(\pi) v(d\pi dQ).
\]
Also define \( vP \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d)) \) by
\[
vP(A) := \int P(\pi_{t+1} \in A | \pi_t = \pi, Q_t = Q) v(d\pi dQ)
\]
for any measurable \( A \subset \mathcal{P}(\mathbb{R}^d) \). Note that \( v \in G \) is equivalent to
\[
\langle vP, f \rangle = \langle v, f \rangle \quad \text{for all } f \in C_b(\mathcal{P}(\mathbb{R}^d)).
\]
From the definition of \( v_tP \), we have for any \( f \in C_b(\mathcal{P}(\mathbb{R}^d)) \),
\[
\langle v_t, f \rangle - \langle v_tP, f \rangle = \frac{1}{t} E_{\pi_0} \left[ \sum_{i=0}^{t-1} f(\pi_i) - \sum_{i=1}^{t} f(\pi_i) \right] = \frac{1}{t} E_{\pi_0} [ f(\pi_0) - f(\pi_t) ] \to 0 \text{ as } t \to \infty.
\]
Now suppose that \( v_{t_k} \to \bar{v} \) weakly along a subsequence of \( \{v_t\} \). Then \( \langle v_{t_k}, f \rangle \to \langle \bar{v}, f \rangle \) for all \( f \in C_b(\mathcal{P}(\mathbb{R}^d)) \), and (51) implies
\[
\langle v_{t_k}P, f \rangle \to \langle \bar{v}, f \rangle.
\]

The following lemma is proved at the end of this section.

Lemma 11. The transition kernel \( P(d\pi_{t+1} | \pi_t, Q_t) \) is continuous in the weak-Feller sense, i.e., for any \( f \in C_b(\mathcal{P}(\mathbb{R}^d)) \),
\[
P f(\pi, Q) := \int_{\mathcal{P}(\mathbb{R}^d) \times \mathcal{Q}_c} f(\pi') P(d\pi' | \pi, Q)
\]
is continuous on \( S \times \mathcal{Q}_c \).

The lemma implies that \( P f \in C_b(S \times \mathcal{Q}_c) \), so \( \langle v_{t_k}, P f \rangle \to \langle \bar{v}, P f \rangle \). However, since for all \( v \),
\[
\langle vP, f \rangle = \int_{\mathcal{P}(\mathbb{R}^d) \times \mathcal{Q}_c} f(\pi) P(d\pi' | \pi, Q) v(d\pi dQ) = \langle v, P f \rangle,
\]
this is equivalent to \( \langle v_{t_k}P, f \rangle \to \langle \bar{v}P, f \rangle \). Combining this with (52) yields \( \langle \bar{v}P, f \rangle = \langle \bar{v}, f \rangle \) which finishes the proof that \( \bar{v} \in G \).
Although \( c(\pi, Q) \) is continuous on \( S \times Q_c \) by Lemma 4, the limit relation (14) does not follow immediately since \( \pi_0 \) may not be in \( S \) and thus \( v_t \) may not be supported on \( S \times Q_c \). However, since \( \pi_t \in S \) for all \( t \geq 1 \) with probability 1, we have \( v_t(S \times Q_c) \geq 1 - 1/t \), and we can proceed as follows: Recall that \( S \times Q_c \) is a closed subset of \( \mathcal{P}([0,1]) \times Q_c \) by Lemma 3 and the topology on \( \mathcal{P}([0,1]) \times Q_c \) is metrizable. Thus by the Tietze-Urysohn extension theorem [11] there exists \( \tilde{c} \in C_b(\mathcal{P}([0,1]) \times Q_c) \) which coincides with \( c \) on \( S \times Q_c \). Then since \( v_t(S \times Q_c) \geq 1 - 1/t \) and both \( c \) and \( \tilde{c} \) are bounded,

\[
\lim_{n \to \infty} \int_{\mathcal{P}([0,1]) \times Q_c} \left| \tilde{c}(\pi, Q) - c(\pi, Q) \right| v_t(d\pi dQ) = 0.
\]

On the other hand, \( v_t \to \bar{v} \) implies

\[
\lim_{n \to \infty} \int_{\mathcal{P}([0,1]) \times Q_c} \tilde{c}(\pi, Q)v_t(d\pi dQ) = \int_{\mathcal{P}([0,1]) \times Q_c} \tilde{c}(\pi, Q)\bar{v}(d\pi dQ) = \int_{\mathcal{P}([0,1]) \times Q_c} c(\pi, Q)\bar{v}(d\pi dQ),
\]

where the last equality holds since \( \bar{v} \in G \) is supported on \( S \times Q_c \). This proves (14).

**Proof of (b).** We need the following simple lemma.

**Lemma 12.** Let \( H \) be a collection of probability measures on \( \mathcal{P}([0,1]) \times Q_c \) such that

\[
R := \sup_{v \in H} \int_{\mathcal{P}([0,1]) \times Q_c} \left( \int_{\mathbb{R}^d} \|x\|^2 \pi(dx) \right) v(d\pi dQ) < \infty
\]

Then \( H \) is tight and is thus relatively compact.

**Proof:** For any \( \alpha > 0 \) let

\[
K_{\alpha} := \left\{ \pi \in \mathcal{P}([0,1]) : \int_{\mathbb{R}^d} \|x\|^2 \pi(dx) \leq \alpha \right\}.
\]

Then \( \pi(\{x : \|x\|^2 > L\}) \leq \alpha/L \) for all \( \pi \in K_{\alpha} \) by Markov’s inequality. Hence \( K_{\alpha} \) is tight and thus relatively compact. A standard truncation argument shows that if \( \pi_k \to \pi \) (weakly) for a sequence \( \{\pi_k\} \) in \( K_{\alpha} \), then

\[
\alpha \geq \limsup_{k \to \infty} \int_{\mathbb{R}^d} \|x\|^2 \pi_k(dx) \geq \int_{\mathbb{R}^d} \|x\|^2 \pi(dx)
\]

so \( K_{\alpha} \) is also closed. Thus \( K_{\alpha} \) is compact.

Let \( f(\pi) := \int_{\mathbb{R}^d} \|x\|^2 \pi(dx) \). Then

\[
\int_{\mathcal{P}([0,1]) \times Q_c} f(\pi) v(d\pi dQ) \leq R \quad \text{for all } v \in H
\]

Again by Markov’s inequality,

\[
\int_{\mathcal{P}([0,1]) \times Q_c} f(\pi) v(d\pi dQ) \geq \alpha v((K_{\alpha})^c \times Q_c)
\]
implying, for all $v \in H$,

$$v(K_\alpha \times Q_c) \geq 1 - \frac{R}{\alpha}.$$  

Since $Q_c$ is compact and $K_\alpha$ is compact for all $\alpha > 0$, we obtain that $H$ is tight.

Let $\Pi$ be an arbitrary fixed policy in $\bar{\Pi}^C_W$, fix the initial distribution $\delta_{x_0}$, and consider the corresponding sequence of expected occupation measures $\{v_t\}$. Then

$$\int_{P(\mathbb{R}^d) \times Q_c} \left( \int_{\mathbb{R}^d} \|x\|^2 \pi(dx) \right) v_t(d\pi dQ) = E_{\delta_{x_0}} \left[ \frac{1}{t} \sum_{k=0}^{t-1} \|x_k\|^2 \right] \rightarrow E_{\hat{\pi}} [\|x\|^2] < \infty$$

by Assumption 4. Hence

$$\sup_{t \geq 0} \int_{P(\mathbb{R}^d) \times Q_c} \left( \int_{\mathbb{R}^d} \|x\|^2 \pi(dx) \right) v_t(d\pi dQ) < \infty.$$  

Thus $\{v_t\}$ is relatively compact by Lemma 12, proving part (b) of the proposition.

Proof of (c) We will show that $G$ is closed and relatively compact. To show closedness, let $\{v_n\}$ be a sequence in $G$ such that $v_n \rightarrow \bar{v}$. Using the notation introduced in the proof of part (a), we have for any $f \in C_b(P(\mathbb{R}^d))$ by (50),

$$\langle v_n P, f \rangle = \langle v_n, f \rangle \rightarrow \langle \bar{v}, f \rangle.$$  

But we also have

$$\langle v_n P, f \rangle = \langle v_n, Pf \rangle \rightarrow \langle \bar{v} P, f \rangle = \langle \bar{v} P, f \rangle,$$

where the limit holds by the weak-Feller property of $P$ (Lemma 11). Thus $\langle \bar{v} P, f \rangle = \langle \bar{v}, f \rangle$, showing that $\bar{v} \in G$. Hence $G$ is closed.

To show relative compactness, recall from (22) the conditional distributions

$$\hat{\pi}(m, \pi, Q)(dx_{t+1}) = P(dx_{t+1}|x_t = \pi, Q_t = Q, q_t = m), \quad m = 1, \ldots, M.$$  

For any $(\pi, Q)$ and Borel set $A \subset \mathbb{R}^d$,

$$\int_{P(\mathbb{R}^d)} \pi'(A)P(d\pi'|\pi, Q)$$

$$= \sum_{m=1}^M \hat{\pi}(m, \pi, Q)(A)P(\hat{\pi}(m, \pi, Q)|\pi, Q)$$

$$= \sum_{m=1}^M \left( \frac{1}{\pi(Q^{-1}(m))} \int_{Q^{-1}(m)} P(x_{t+1} \in A|x_t)\pi(dx_t) \right) \pi(Q^{-1}(m))$$

$$= \int_{\mathbb{R}^d} P(x_{t+1} \in A|x_t)\pi(dx_t).$$  

Now let $v \in G$ and consider the “average” $\pi_v$ under $v$ determined by

$$\pi_v(A) = \int_{P(\mathbb{R}^d) \times Q_c} \pi(A)v(d\pi dQ) = \int_{P(\mathbb{R}^d)} \pi(A)\hat{v}(d\pi),$$
Since the claim of the lemma.

The proof of the same lemma shows that $A$ has boundary of zero Lebesgue measure, the mapping $\pi \mapsto \pi(A)$ is continuous on $S$ and the definition of $G$ implies

$$
\pi_v(A) = \int_{\mathcal{P}(\mathbb{R}^d) \times \mathcal{Q}_c} \pi(A) v(d\pi dQ) = \int_{\mathcal{P}(\mathbb{R}^d) \times \mathcal{Q}_c} \int_{\mathcal{P}(\mathbb{R}^d)} \pi'(A) P(d\pi'|\pi, Q) v(d\pi dQ)
$$

$$
= \int_{\mathcal{P}(\mathbb{R}^d)} \int_{\mathcal{Q}_c} \int_{\mathcal{P}(\mathbb{R}^d)} \pi'(A) P(d\pi'|\pi, Q) \tilde{\eta}(dQ|\pi) \hat{v}(d\pi)
$$

Substituting (54) into the last integral, we obtain

$$
\pi_v(A) = \int_{\mathcal{P}(\mathbb{R}^d)} \int_{\mathcal{Q}_c} \int_{\mathbb{R}^d} P(x_{t+1} \in A|x_t) \pi(dx_t) \tilde{\eta}(dQ|\pi) \hat{v}(d\pi)
$$

$$
= \int_{\mathbb{R}^d} P(x_{t+1} \in A|x_t) \pi_v(dx_t).
$$

Since the Borel sets in $\mathbb{R}^d$ having boundaries of zero Lebesgue measure form a separating class for $\mathcal{P}(\mathbb{R}^d)$, the above holds for all Borel sets $A$, implying that $\pi_v = \pi^*$, the unique invariant measure for $\{x_t\}$. Thus

$$
\int_{\mathcal{P}(\mathbb{R}^d) \times \mathcal{Q}_c} \left( \int_{\mathbb{R}^d} \|x\|^2 \pi(d\pi) \right) v(d\pi dQ) = \int_{\mathbb{R}^d} \|x\|^2 \pi_v(dx) = \int_{\mathbb{R}^d} \|x\|^2 \pi^*(dx)
$$

for all $v \in G$. Since the last integral is finite by Assumption 4, Lemma 12 implies that $G$ is relatively compact.

**Proof of Lemma 11.** Consider a sequence $\{(\pi_n, Q_n)\}$ converging to some $(\pi, Q)$ in $S \times \mathcal{Q}_c$. Then for any $f \in C_b(\mathcal{P}(\mathbb{R}^d))$,

$$
\int_{\mathcal{P}(\mathbb{R}^d) \times \mathcal{Q}_c} f(\pi') P(d\pi'|\pi_n, Q_n) - \int_{\mathcal{P}(\mathbb{R}^d) \times \mathcal{Q}_c} f(\pi') P(d\pi'|\pi, Q)
$$

$$
= \sum_{m=1}^{M} \left( f(\hat{\pi}(m, \pi_n, Q_n)) P(\hat{\pi}(m, \pi_n, Q_n)|\pi_n, Q_n) - f(\hat{\pi}(m, \pi, Q)) P(\hat{\pi}(m, \pi, Q)|\pi, Q) \right)
$$

$$
= \sum_{m=1}^{M} \left( f(\hat{\pi}(m, \pi_n, Q_n)) \pi_n(Q_n^{-1}(m)) - f(\hat{\pi}(m, \pi, Q)) \pi(Q^{-1}(m)) \right).
$$

From Lemma 3 we have that $\pi_n Q_n \to \pi Q$ in total variation which implies via Lemma 6 that $\hat{\pi}(m, \pi_n, Q_n) \to \hat{\pi}(m, \pi, Q)$ in total variation and thus weakly for all $m$ with $\pi(Q^{-1}(m)) > 0$. The proof of the same lemma shows that $\pi_n(Q_n^{-1}(m)) \to \pi(Q^{-1}(m))$ for all $m = 1, \ldots, M$. Since $f$ is continuous and bounded, the sum in (57) converges to zero as $n \to \infty$, proving the claim of the lemma.

**VIII. ACKNOWLEDGEMENTS**

We are grateful to Vivek S. Borkar and Naci Saldi for technical discussions related to the paper.

May 7, 2014
REFERENCES

[1] E. A. Abaya and G. L. Wise, “Convergence of vector quantizers with applications to optimal quantization,” *SIAM Journal on Applied Mathematics*, vol. 44, pp. 183–189, 1984.

[2] A. Arapostathis, V. S. Borkar, E. Fernandez-Gaucherand, M. K. Ghosh and S. I. Marcus, “Discrete-Time controlled Markov processes with average cost criterion: A survey,” *SIAM J. Control and Optimization*, vol. 31, pp. 282–344, 1993.

[3] A. Arapostathis, V. S. Borkar and M. K. Ghosh, *Ergodic Control of Diffusion Processes*, Cambridge University Press, 2012.

[4] H. Asnani and T. Weissman, “On real time coding with limited lookahead,” *IEEE Trans. Inform. Theory*, vol. 59, no. 6, pp. 3582–3606, Jun. 2013.

[5] L. Bao, M. Skoglund, and K. H. Johansson, “Iterative encoder-controller design for feedback control over noisy channels,” *IEEE Trans. on Automatic Control*, vol. 57, no. 2, pp. 265–278, Feb. 2011.

[6] P. Billingsley, *Probability and Measure*. New York: Wiley, 2nd ed., 1986.

[7] V. S. Borkar, “Convex analytic methods in Markov Decision Processes,” *Handbook of Markov Decision Processes: Methods and Applications*, Kluwer, Boston, 2002.

[8] V. S. Borkar and S. K. Mitter, “LQG control with communication constraints,” in *Kailath Festschrift*, Kluwer, Boston, 1997.

[9] V.S. Borkar, S.K. Mitter, A. Sahai and S. Tatikonda, “Sequential source coding: An optimization viewpoint,” *Proc. IEEE Conference on Decision and Control*, pp. 1035–1042, Seville, Spain, Dec. 2005.

[10] V. S. Borkar, S. K. Mitter, and S. Tatikonda, “Optimal sequential vector quantization of Markov sources,” *SIAM J. Control and Optimization*, vol. 40, no. 1, pp. 135–148, 2001.

[11] R. M. Dudley, *Real Analysis and Probability*, Cambridge University Press, Cambridge, 2nd ed., 2002.

[12] I. I. Gikhman and A. V. Skorokhod, *Controlled Stochastic Processes*, Springer, New York, 1979.

[13] R. M. Gray and D. L. Neuhoff, “Quantization,” *IEEE Trans. Inform. Theory* (Special Commemorative Issue), vol. 44, pp. 2325–2383, Oct. 1998.

[14] A. György and T. Linder, “Codecell convexity in optimal entropy-constrained vector quantization,” *IEEE Trans. Inf. Theory*, vol. 49, pp. 1821–1828, July 2003.

[15] O. Hernandez-Lerma, J.B. Lasserre, *Discrete-Time Markov Control Processes, Basic Optimality Criteria*, Springer, New York, 1996.

[16] O. Hernandez-Lerma, J.B. Lasserre, *Further topics on discrete-time Markov control processes*, Springer, New York, 1999.

[17] O. Hernandez-Lerma, J.B. Lasserre, *Markov Chains and Invariant Probabilities*, Birkhäuser, Basel, 2003.

[18] T. Javidi and A. Goldsmith, “Dynamic joint source-channel coding with feedback,” *Proc. IEEE International Symposium on Information Theory*, pp. 16-20, Istanbul, 2013.

[19] Y. Kaspi and N. Merhav, “Structure theorems for real-time variable rate coding with and without side information,” *IEEE Trans. Inf. Theory*, vol. 58, no.12, pp. 7135–7153, Dec. 2012.

[20] T. Linder and R. Zamir, “Causal coding of stationary sources and individual sequences with high resolution,” *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 662–680, Feb. 2006.

[21] A. Mahajan and D. Tenekezis, “On the design of globally optimal communication strategies for real-time noisy communication with noisy feedback”, *IEEE Journal on Special Areas in Communications*, vol. 28, pp. 580–595, May 2008.

[22] A. Mahajan and D. Tenekezis, “Optimal design of sequential real-time communication systems”, *IEEE Transactions on Inform. Theory*, vol. 55, pp. 5317–5338, November 2009.

[23] S. P. Meyn, *Control Techniques for Complex Networks*, Cambridge, UK: Cambridge University Press.

[24] S. P. Meyn and R. Tweedie, *Markov Chains and Stochastic Stability*, Springer, London, 1993.

[25] G. N. Nair, F. Fagnani, S. Zampieri, and J. R. Evans, “Feedback control under data constraints: an overview,” *Proceedings of the IEEE*, vol. 95, no. 1, pp. 108–137, Jan. 2007.

[26] D. L. Neuhoff and R. K. Gilbert, “Causal source codes,” *IEEE Trans. Inform. Theory*, vol. 28, pp. 701–713, Sep. 1982.

[27] D. Pollard, “Quantization and the method of k-means,” *IEEE Trans. Inf. Theory*, vol. 28, pp. 199–205, 1982.

[28] N. Saldi, T. Linder and S. Yüksel, “Randomized quantization and optimal design with a marginal constraint,” *Proc. IEEE International Symposium on Information Theory*, pp. 2349–2353, Istanbul, 2013.
[29] S. Tatikonda and S. Mitter, “The capacity of channels with feedback,” IEEE Trans. Inf. Theory, vol. 55, pp. 323–349, Jan. 2009.

[30] S. Tatikonda, A. Sahai, and S. Mitter, “Stochastic linear control over a communication channels”, IEEE Trans. Aut. Control, vol. 49, pp. 1549–1561, Sept. 2004.

[31] D. Teneketzis, “On the structure of optimal real-time encoders and decoders in noisy communication,” IEEE Trans. Inf. Theory, vol. 52, pp. 4017–4035, Sep. 2006.

[32] J. C. Walrand and P. Varaiya, “Optimal causal coding-decoding problems,” IEEE Trans. Inf. Theory, vol. 19, pp. 814–820, Nov. 1983.

[33] T. Weissman and N. Merhav, “On causal source codes with side information,” IEEE Trans. Inform. Theory, vol. 51, no. 11, pp. 4003–4013, Nov. 2005.

[34] H. S. Witsenhausen, “On the structure of real-time source coders,” Bell Syst. Tech. J., vol. 58, pp. 1437–1451, Jul./Aug. 1979.

[35] S. Yüksel, “On optimal causal coding of partially observed Markov sources in single and multi-terminal settings,” IEEE Trans. Inf. Theory, vol. 59, pp. 424–437, Jan. 2013.

[36] S. Yüksel, “Jointly optimal LQG quantization and control policies for multi-dimensional systems” IEEE Trans. on Automatic Control, vol. 59, Jun. 2014.

[37] S. Yüksel, T. Başar, Stochastic Networked Control Systems: Stabilization and Optimization under Information Constraints, Birkhäuser, Boston, MA, 2013.

[38] S. Yüksel and T. Linder, “Optimization and convergence of observation channels in stochastic control,” SIAM Journal on Control and Optimization, vol. 50, no. 2, pp. 864–887, 2012.