Optimal Zero Delay Coding of Markov Sources: Stationary and Finite Memory Codes

Richard G. Wood, Tamás Linder, *Fellow, IEEE*, and Serdar Yüksel, *Member, IEEE*

Abstract—The optimal zero delay coding of a finite-state Markov source is considered. The existence and structure of optimal codes are studied using a stochastic control formulation. Prior results in the literature established the optimality of deterministic Markov (Walrand–Varaiya-type) coding policies for the finite time horizon problem, and the optimality of both deterministic nonstationary and randomized stationary policies for the infinite time horizon problem. Our main result here shows that for any irreducible and aperiodic Markov source with a finite alphabet, *deterministic and stationary* Markov coding policies are optimal for the infinite horizon problem. In addition, the finite block length (time horizon) performance of an optimal (stationary and Markov) coding policy is shown to approach the infinite time horizon optimum at a rate $O(1/T)$. The results are extended to systems, where zero delay communication takes place across a noisy channel with noiseless feedback.

Index Terms—Zero delay source coding, real time coding, causal coding, quantization, stochastic control, Markov sources, Markov decision processes.

I. INTRODUCTION

THIS paper is concerned with optimal zero delay coding of Markov sources for infinite time horizons. Zero delay coding is a variant of the original lossy source coding problem introduced by Shannon [1].

A. Block Coding and Zero Delay Coding

Recall Shannon's lossy source coding problem [2]: Given is an X-valued information source $\{X_t\}_{t \geq 0}$, where we assume that X is a finite set. An encoder compresses the source at a rate R bits per source symbol. A decoder reproduces the information source via the sequence $\{\hat{X}_t\}_{t \geq 0}$ of $\hat{\mathsf{X}}$-valued random variables, where $\hat{\mathsf{X}}$ is also a finite set. One is typically concerned with the transmission *rate* and the *distortion* of the system.

In particular, a $(2^{RT}, T)$-*rate distortion block code* [2] encodes T source symbols $X_{[0,T-1]} := (X_0, \ldots, X_{T-1})$ at a time, and comprises an encoding function $\eta^T : \mathsf{X}^T \to \{1, \ldots, 2^{RT}\}$ and a decoding function $\gamma^T : \{1, \ldots, 2^{RT}\} \to$

$\hat{\mathsf{X}}^T$. This code has rate R bits per source symbol, and (expected) distortion given by

$$D_T := \frac{1}{T} E \left[\sum_{t=0}^{T-1} d\left(X_t, \hat{X}_t\right) \right],$$

where $(\hat{X}_0, \ldots, \hat{X}_{T-1}) = \gamma^T(\eta^T(X_{[0,T-1]}))$ and $d : \mathsf{X} \times \hat{\mathsf{X}} \to [0, \infty)$ is a so called single letter distortion measure.

A rate distortion pair (R, D) is said to be *achievable* if there exists a sequence of $(2^{RT}, T)$-rate distortion codes (η^T, γ^T) such that

$$\limsup_{T \to \infty} D_T \leq D. \tag{1}$$

By a classical result, if the source is stationary and ergodic, the minimum achievable distortion for rate R is given by the distortion rate function of the source

$$D(R) = \lim_{T \to \infty} D_T(R), \tag{2}$$

where $D_T(R)$ is the Tth order distortion rate function which can be calculated from the statistics of the block $X_{[0,T-1]}$ (see, e.g., [3]).

As is evident from the definition of block codes, such a coding scheme relies on encoding blocks of data $(X_0, \ldots, X_{T-1})$ together, which may not be practical for many applications as the encoder has to wait until it has all T source symbols before it can start encoding and transmitting the data. In *zero delay source coding*, the encoder can produce the code of $\hat{X}_t$ as soon as the source symbol X_t is available. Such coding schemes have many practical applications in emerging fields such as networked control systems (see [4] and references therein for an extensive review and discussion of applications), real-time mobile audio-video systems (as in streaming systems [5], [6]), and real-time sensor networks [7], among other areas.

In this paper, we consider a zero delay (sequential) encoding problem where the goal is to encode an observed information source without delay. It is assumed that the information source $\{X_t\}_{t \geq 0}$ is an X-valued discrete time Markov process, where X is a finite set. The transition probability matrix P and initial distribution π_0 for X_0 completely determine the process distribution, so we will use the shorthand $\{X_t\} \sim (\pi_0, P)$. The encoder encodes (quantizes) the source samples and transmits the encoded versions to a receiver over a discrete noiseless channel with common input and output alphabet $\mathsf{M} := \{1, 2, \ldots, M\}$, where M is a positive integer.

In the following, we build on the notation in [8]. Formally, the encoder is specified by a *quantization policy* Π, which is

a sequence of functions $\{\eta_t\}_{t\geq0}$ with $\eta_t : \mathsf{M}^t \times \mathsf{X}^{t+1} \to \mathsf{M}$. At time t, the encoder transmits the M-valued message

$$q_t = \eta_t(I_t)$$

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 $X_{[0,t]} = (X_0, \ldots, X_t)$. The collection of all such zero delay encoding policies is called the set of *admissible* quantization policies and is denoted by Π_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 mapping of X into the finite set M. Thus a quantization policy at each time index t selects a quantizer $Q_t : \mathsf{X} \to \mathsf{M}$ 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 decoder generates the reconstruction $\hat{X}_t$, also without delay. A zero delay decoder policy is a sequence of functions $\gamma = \{\gamma_t\}_{t\geq0}$ of type $\gamma_t : \mathsf{M}^{t+1} \to \hat{\mathsf{X}}$, where $\hat{\mathsf{X}}$ denotes the finite reconstruction alphabet. Thus for all $t \geq 0$,

$$\hat{X}_t = \gamma_t(q_{[0,t]}).$$

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

$$E_{\pi_0}^{\Pi,\gamma}\left[\frac{1}{T}\sum_{t=0}^{T-1} d(X_t, \hat{X}_t)\right] \tag{3}$$

for some $T \geq 1$, where $d : \mathsf{X} \times \hat{\mathsf{X}} \to [0, \infty)$ is a cost (distortion) function and $E_{\pi_0}^{\Pi,\gamma}$ denotes expectation with initial distribution π_0 for X_0 and under the quantization policy Π and receiver policy γ. We assume that the encoder and decoder know the initial distribution π_0.

Since the source alphabet is finite, for any encoder policy $\Pi \in \Pi_A$ and any $t \geq 0$, there always exists an optimal receiver policy $\gamma^* = \gamma^*(\Pi)$ such that for all $t \geq 0$,

$$E_{\pi_0}^{\Pi,\gamma^*}\left[d(X_t, \hat{X}_t)\right] = \inf_\gamma E_{\pi_0}^{\Pi,\gamma}\left[d(X_t, \hat{X}_t)\right].$$

From now on, we always assume that an optimal receiver policy is used for a given encoder policy and, with an abuse of notation, $\Pi \in \Pi_A$ will mean the combined encoder and decoder policies $(\Pi, \gamma^*(\Pi))$. Using this new notation, we have for all $t \geq 0$,

$$E_{\pi_0}^{\Pi}\left[d(X_t, \hat{X}_t)\right] = \inf_\gamma E_{\pi_0}^{\Pi,\gamma}\left[d(X_t, \hat{X}_t)\right].$$

In this paper, we concentrate on the following performance criteria.

1) *Infinite Horizon Discounted Cost Problem*: In the infinite horizon discounted cost problem, the goal is to minimize the cumulative "discounted" cost

$$J_{\pi_0}^{\beta}(\Pi) := \lim_{T\to\infty} E_{\pi_0}^{\Pi}\left[\sum_{t=0}^{T-1} \beta^t d(X_t, \hat{X}_t)\right] \tag{4}$$

for some $\beta \in (0, 1)$.

2) *Infinite Horizon Average Cost Problem*: The more challenging infinite horizon average cost problem has the objective of minimizing the long term average distortion

$$J_{\pi_0}(\Pi) := \limsup_{T\to\infty} E_{\pi_0}^{\Pi}\left[\frac{1}{T}\sum_{t=0}^{T-1} d(X_t, \hat{X}_t)\right]. \tag{5}$$

We note that in source coding only the average cost problem is of interest, but we also consider the discounted cost problem since it will serve as a useful tool in studying the more difficult average cost problem.

Observe that $R = \log_2 M$ is the rate of the described zero delay codes. Then, in analogy to (1), the rate distortion pair (R, D) is said to be achievable if there exists a policy Π such that $J_{\pi_0}(\Pi) \leq D$. As opposed to the block coding case, finding the minimum achievable distortion (cost) $\min_{\Pi\in\Pi_A} J_{\pi_0}(\Pi)$ at rate R for zero delay codes is an open problem. In particular, if the source is stationary and memoryless, then this minimum is equal to $\min_f E[d(X_0, f(X_0))]$, where the minimum is taken over all "memoryless quantizers" $f : \mathsf{X} \to \hat{\mathsf{X}}$ with $|f(\mathsf{X})| \leq 2^R$ [9]–[11]. However, this optimum performance is not known for any other (more general) source classes, and in particular it is unknown when $\{X_t\}$ is a stationary and ergodic Markov source. (Some partial results on this problem are given in, e.g., [11], [12].)

Our main goal in this paper is to characterize some important properties of optimal coding policies that achieve this minimum, even though we cannot characterize the value of the minimum.

We review two results fundamental to the structure of optimal zero delay codes (see also [13]).

Theorem 1 (Witsenhausen [14]): For the problem of coding a Markov source over a finite time horizon T, any zero delay quantization policy $\Pi = \{\eta_t\}$ can be replaced, without loss in distortion 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 $t = 1, \ldots, T - 1$.

Let $\mathcal{P}(\mathsf{X})$ denote the space of probability measures on X. Given a quantization policy Π, for all $t \geq 1$ let $\pi_t \in \mathcal{P}(\mathsf{X})$ be the conditional probability defined by

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

for any set $A \subset \mathsf{X}$.

Theorem 2 (Walrand and Varaiya [15]): For the problem of coding a Markov source over a finite time horizon T, any zero delay quantization policy can be replaced, without 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 $\hat{\eta}_t$ uses π_t to select a quantizer $Q_t = \hat{\eta}(\pi_t)$ (where $Q_t : \mathsf{X} \to \mathsf{M}$), and then q_t is generated as $q_t = Q_t(x_t)$.

As discussed in [13], 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 $\mathcal{P}(\mathsf{X})$ of an optimal encoder is fixed (but of course is not finite). 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.

Recall that a Markov chain $\{X_t\}$ with finite state space X is *irreducible* if for any $a, b \in \mathsf{X}$ there exists a positive n such that $\Pr(X_n = b | X_0 = a) > 0$ (e.g., [44, Chapter 1.2]), and it is *aperiodic* if for each state $a \in \mathsf{X}$ there is a positive n such that $\Pr(X_{n'} = a | X_0 = a) > 0$ for all $n' \geq n$ (e.g., [44, Chapter 1.8]). Our assumption on the source $\{X_t\}$ is that it is an irreducible and aperiodic finite state Markov chain.

The main results in this paper are the following.

- For the problem of zero delay source coding of an irreducible and aperiodic Markov source over an infinite time horizon we show the optimality (among all admissible policies) of *deterministic* and *stationary* (i.e., time invariant) Markov (Walrand-Varaiya type) policies for both stationary and nonstationary Markov sources.
- For the same class of Markov sources, we show that the optimum performance for time horizon T converges to the optimum infinite horizon performance at least as fast as $O\left(\frac{1}{T}\right)$.
- Using the above convergence rate result, for stationary Markov sources we also show the existence of ϵ-optimal periodic zero delay codes with an explicit bound on the relationship between ϵ and the period length. This result is relevant since the complexity of the code is directly related to the length of the period (memory size).

The rest of the paper is organized as follows. In the next subsection we review some existing results on zero delay coding and related problems. In Section II we derive auxiliary results and show that stationary Walrand-Varaiya type policies are optimal in the set of all policies for the infinite horizon discounted cost problem. In Section III we consider the infinite horizon average cost problem and prove the optimality of stationary and deterministic Walrand-Varaiya type policies. The convergence rate result and the ϵ-optimality of finite memory policies are also presented here. Section IV describes the extension of these results for zero delay coding over a noisy channel with feedback. Concluding remarks are given in Section V. In the Appendix we provide a brief summary of some definitions and results we need from the theory of Markov decision processes.

B. Literature Review

Structural results for the finite horizon coding problem have been developed in a number of important papers. As mentioned before, the classic works by Witsenhausen [14] and Walrand and Varaiya [15], which use two different approaches, are of particular relevance. An extension to the more general setting of non feedback communication was given by Teneketzis [16], and [13] also extended these results to more general state spaces; see also [8] and [4] for a more detailed review.

A related lossy coding procedure was introduced by Neuhoff and Gilbert [17], which they called *causal source coding*. The main result in [17] established that for stationary memoryless sources, an optimal causal coder can either be replaced by one that time shares two memoryless coders, without loss in

performance. As noted in [17], zero delay codes form a special subclass of causal codes. We also note that *scalar quantization* is a practical (but in general suboptimal) method for zero delay coding of continuous sources. A detailed review of classical results on scalar and vector quantization is given in [18].

Causal coding under a high rate assumption for stationary sources and individual sequences was studied in [25]. Borkar et al. [26] studied the related problem of coding a partially observed Markov source and obtained existence results for dynamic vector quantizers in the infinite horizon setting. It should be noted that in [26] the set of admissible quantizers was restricted to the set of nearest neighbor quantizers, and other conditions were placed on the dynamics of the system; furthermore the proof technique used in [26] relies on the fact that the source is partially observed unlike the setup we consider here.

In [8], zero delay coding of $\mathbb{R}^d$-valued Markov sources was considered. In particular, [8] established the existence of optimal quantizers (having convex codecells) for finite horizons and the existence of optimal *deterministic nonstationary* or *randomized stationary* policies for stationary Markov sources over infinite horizons, but the optimality of *stationary and deterministic* codes was left as an open problem. Related work include [27] which considered the coding of discrete independent and identically distributed (i.i.d.) sources with limited lookahead using the average cost optimality equation. Also, [28] studied real time joint source-channel coding of a discrete Markov source over a discrete memoryless channel with feedback under a similar average cost formulation.

Some partial, but interesting results on the optimum performance of zero-delay coding over a noisy channel are available in the literature. It is shown in [15, Theorem 3] that when the source and the channel alphabets have the same cardinality and the channel satisfies certain symmetry conditions (e.g., the channel is the binary symmetric channel or a binary erasure channel), then memoryless encoding is optimal for any Markov source. Also, an information theoretic source-channel *matching* type argument can be made for special scenarios where the sequential rate-distortion [19], [20] achieving channel kernels are realized with the physical channel itself, a crucial case being the scalar Gaussian source transmitted over a scalar Gaussian channel under power constraints at the encoder [21]. Along this direction, a more modern treatment and further results are given in [22] and [23]. Optimal zero delay coding of Markov sources over noisy channels without feedback was considered in [16] and [24].

In this paper we also investigate how fast the optimum finite blocklength (time horizon) distortion converges to the optimum (infinite horizon) distortion. An analog of this problem in block coding is the speed of convergence of the finite block length encoding performance to Shannon's distortion rate function. For stationary and memoryless sources, this speed of convergence was shown to be of the type $O\left(\frac{\log T}{T}\right)$ [29], [30]. See also [31] for a detailed literature review and further finite blocklength performance bounds.

Finally, we note that control theoretic tools are playing an increasingly important role in solving certain types of problems in information theory. Several of the papers cited

above use dynamic programming as a crucial tool to analyze an average cost optimal control problem that the given information theoretic problem is reduced to. To facilitate this analysis, the *convex analytic method* [32] was used, e.g., in [8] and [33], while in [26], [34], [35], [36], [27], [37], and [38] the *average cost optimality equation* was used (typically through the *vanishing discount method*). In particular, [34], [37], and [38] applied this latter approach to solve dynamic programs that provide *explicit* channel capacity expressions. In this paper (unlike in our earlier work [8]), we also use the average cost optimality equation approach, but here certain technical subtleties complicate the analysis: (i) the structural result (on the optimality of Walrand-Varaiya type policies) only holds for finite horizon problems; and (ii) we have a controlled Markov chain (where the beliefs are the states and the quantizer maps are the actions) only when the quantizers belong to the Walrand-Varaiya class (see Definition 1). Much of our technical analysis concerns extending this line of argument to the infinite horizon case through the study of recurrence, coupling, convergence, and continuity properties of the underlying controlled Markov chain.

II. THE FINITE HORIZON AVERAGE DISTORTION AND THE INFINITE HORIZON DISCOUNTED DISTORTION PROBLEMS

A. The Finite Horizon Average Cost Problem

In view of Theorem 2, for a finite horizon problem any admissible (i.e., zero delay) quantization policy can be replaced by a Walrand-Varaiya type policy. Using the terminology of Markov decision processes, we will also refer to such policies as Markov policies. The class of all such policies is denoted by Π_W, and is formally defined below.

Definition 1: Let $\mathcal{Q}$ denote the set of all quantizers $Q : \mathsf{X} \to \mathsf{M}$. An (admissible) quantization policy $\Pi = \{\eta_t\}$ belongs to Π_W if there exists a sequence of mappings $\{\hat{\eta}_t\}$ of the type $\hat{\eta}_t : \mathcal{P}(\mathsf{X}) \to \mathcal{Q}$ such that for $Q_t = \hat{\eta}_t(\pi_t)$ we have $q_t = Q_t(X_t) = \eta_t(I_t)$. A policy in Π_W is called *stationary* if $\hat{\eta}_t$ does not depend on t. The set of such stationary policies is denoted by Π_{WS}.

Remark: It is worth pointing out that the classical definition of a stationary (time invariant or sliding block [39]) encoder involves a "two sided" infinite source sequence $\{X_t\}_{t=-\infty}^{\infty}$ and has the form $q_t = g(X_{[-\infty,t]})$ for all t, where g maps the infinite past $X_{[-\infty,t]} = \ldots, X_{t-2}, X_{t-1}, X_t$, up to time t into the symbol q_t. Clearly, for a "one sided" source $\{X_t\}_{t \geq 0}$ such a definition of stationary codes is problematic. Thus, in a sense, stationary Walrand-Varaiya type encoding policies give a useful generalization of classical stationary encoders for the case of one sided sources.

Building on [13] and [8], suppose a given quantizer policy $\Pi = \{\hat{\eta}_t\}$ in Π_W is used to encode the Markov source $\{X_t\}$. Let $P = P(x_{t+1}|x_t)$ denote the transition kernel of the source. Observe that the conditional probability of q_t given π_t and x_t is given by $P(q_t|\pi_t, x_t) = 1_{\{Q_t(x_t) = q_t\}}$ with $Q_t = \hat{\eta}_t(\pi_t)$, and is therefore determined by the quantizer policy. Then standard properties of conditional probability can be used to obtain the

following "filtering equation" for the evolution of π_t:

$$
\begin{aligned}
\pi_{t+1}&(x_{t+1}) \\
&= \frac{\sum_{x_t} \pi_t(x_t) P(q_t|\pi_t, x_t) P(x_{t+1}|x_t)}{\sum_{x_t} \sum_{x_{t+1}} \pi_t(x_t) P(q_t|\pi_t, x_t) P(x_{t+1}|x_t)} \\
&= \frac{1}{\pi_t(Q_t^{-1}(q_t))} \sum_{x_t \in Q_t^{-1}(q_t)} P(x_{t+1}|x_t) \pi_t(x_t). \quad (6)
\end{aligned}
$$

Therefore, given π_t and Q_t, π_{t+1} is conditionally independent of $(\pi_{[0,t-1]}, Q_{[0,t-1]})$. Thus $\{\pi_t\}$ can be viewed as a $\mathcal{P}(\mathsf{X})$-valued controlled Markov process [40] with $\mathcal{Q}$-valued control $\{Q_t\}$ and average cost up to time $T - 1$ given by

$$
E_{\pi_0}^{\Pi} \left[\frac{1}{T} \sum_{t=0}^{T-1} d(X_t, \hat{X}_t) \right] = E_{\pi_0}^{\Pi} \left[\frac{1}{T} \sum_{t=0}^{T-1} c(\pi_t, Q_t) \right],
$$

where

$$
c(\pi_t, Q_t) := \sum_{i=1}^{M} \min_{\hat{x} \in \hat{\mathsf{X}}} \sum_{x \in Q_t^{-1}(i)} \pi_t(x) d(x, \hat{x}). \quad (7)
$$

In this context, Π_W corresponds to the class of deterministic Markov control policies [40]. The Appendix provides a brief overview of controlled Markov processes.

The following statements follow from results in [8], but they can also be straightforwardly derived since, in contrast to [8], here we have only finitely many M-cell quantizers on X. For any $\Pi \in \Pi_A$, define

$$
J_{\pi_0}(\Pi, T) := E_{\pi_0}^{\Pi} \left[\frac{1}{T} \sum_{t=0}^{T-1} d(X_t, \hat{X}_t) \right].
$$

Proposition 1: For any $T \geq 1$, there exists a policy Π in Π_W such that

$$
J_{\pi_0}(\Pi, T) = \inf_{\Pi' \in \Pi_A} J_{\pi_0}(\Pi', T). \quad (8)
$$

Letting $J_T^T(\cdot) := 0$, $J_0^T(\pi_0) := \min_{\Pi \in \Pi_W} J_{\pi_0}(\Pi, T)$, the dynamic programming recursion

$$
\begin{aligned}
T J_t^T&(\pi) \\
&= \min_{Q \in \mathcal{Q}} \left(c(\pi, Q) + T E\big[J_{t+1}^T(\pi_{t+1}) | \pi_t = \pi, Q_t = Q \big] \right)
\end{aligned}
$$

holds for all $t = T - 1, T - 2, \ldots, 0$ and $\pi \in \mathcal{P}(\mathsf{X})$.

Proof: By Theorem 2, there exists a policy Π in Π_W such that (8) holds. Also, by Theorem 8 in the Appendix, we can use the dynamic programming recursion to solve for an optimal quantization policy $\Pi \in \Pi_W$. □

B. The Infinite Horizon Discounted Cost Problem

As discussed in Section I-A, the goal of the infinite horizon discounted cost problem is to find policies that achieve

$$
J_{\pi_0}^{\beta} := \inf_{\Pi \in \Pi_A} J_{\pi_0}^{\beta}(\Pi) \quad (9)
$$

for given $\beta \in (0, 1)$, where

$$
J_{\pi_0}^{\beta}(\Pi) = \lim_{T \to \infty} E_{\pi_0}^{\Pi} \left[\sum_{t=0}^{T-1} \beta^t d(X_t, \hat{X}_t) \right].
$$

From the viewpoint of source coding, the discounted cost problem has much less significance than the average cost problem. However the discounted cost approach will play an important role in deriving results for the average cost problem.

Proposition 2: There exists an optimal (deterministic) quantization policy in Π_{WS} among all policies in Π_A that achieves the infimum in (9).

Proof: Observe that

$$
\inf_{\Pi \in \Pi_A} \lim_{T \to \infty} E_{\pi_0}^{\Pi} \left[\sum_{t=0}^{T-1} \beta^t d(X_t, \hat{X}_t) \right]
$$

$$
\geq \limsup_{T \to \infty} \inf_{\Pi \in \Pi_A} E_{\pi_0}^{\Pi} \left[\sum_{t=0}^{T-1} \beta^t d(X_t, \hat{X}_t) \right]
$$

$$
= \limsup_{T \to \infty} \min_{\Pi \in \Pi_W} E_{\pi_0}^{\Pi} \left[\sum_{t=0}^{T-1} \beta^t d(X_t, \hat{X}_t) \right]
$$

$$
= \limsup_{T \to \infty} \min_{\Pi \in \Pi_W} E_{\pi_0}^{\Pi} \left[\sum_{t=0}^{T-1} \beta^t c(\pi_t, Q_t) \right], \qquad (10)
$$

where the first equality follows from Theorem 2 and the second from the definition of $c(\pi_t, Q_t)$ in (7). For each T, let Π_T denote the optimal policy in Π_W achieving the minimum in (10).

One can easily check that conditions (i)–(iii) of Theorem 9 in the Appendix hold in our case (with $\mathsf{Z} = \mathcal{P}(\mathsf{X})$, $\mathsf{A} = \mathcal{Q}$, $c(z, a) = c(\pi, Q)$, and $K(dz'|z, a) = P(d\pi'|\pi, Q)$). Specifically, the definition of $c(\pi, Q)$ in (7) shows that c is continuous, so (i) holds. Condition (ii) clearly holds. since $\mathcal{Q}$ is a finite set. Finally, it is easily verified that the stochastic kernel $P(d\pi_{t+1}|\pi_t, Q_t)$ is weakly continuous, i.e., that $\int_{\mathcal{P}(\mathsf{X})} f(\pi') P(d\pi'|\pi, Q)$ is continuous on $\mathcal{P}(\mathsf{X}) \times \mathcal{Q}$ for any continuous and bounded $f : \mathcal{P}(\mathsf{X}) \to \mathbb{R}$ (see [8, Lemma 11]). Thus by Theorem 9 in the Appendix, this sequence of policies, $\{\Pi_T\}$, can be obtained by using the iteration algorithm

$$
J_t(\pi) = \min_{Q \in \mathcal{Q}} \left[c(\pi, Q) + \beta \int_{\mathcal{P}(\mathsf{X})} J_{t-1}(\pi') P(d\pi'|\pi, Q) \right]
$$

with $J_0(\pi) \equiv 0$. By the same theorem, the sequence of value functions for the policies $\{\Pi_T\}$, i.e. $\{J_{\pi_0}(\Pi_W, T)\}$, converges to the value function of some deterministic policy $\Pi \in \Pi_{WS}$ (i.e., a deterministic stationary Markov policy) which is optimal in the set of policies Π_W for the infinite horizon discounted cost problem. Thus by the chain of inequalities leading to (10), Π is optimal among all policies in Π_A. □

III. MAIN RESULTS: THE INFINITE HORIZON AVERAGE DISTORTION PROBLEM

The more challenging average cost case deals with a performance measure (the long time average distortion) that is usually studied in source coding problems. Formally, the infinite horizon average cost of a coding policy Π is

$$
J_{\pi_0}(\Pi) = \limsup_{T \to \infty} E_{\pi_0}^{\Pi} \left[\frac{1}{T} \sum_{t=0}^{T-1} d(X_t, \hat{X}_t) \right] \qquad (11)
$$

and the goal is to find an optimal policy attaining

$$
J_{\pi_0} := \inf_{\Pi \in \Pi_A} J_{\pi_0}(\Pi). \qquad (12)
$$

A. Optimality of Policies in Π_W for Stationary Sources

For the infinite horizon setting structural results such as Theorems 1 and 2 are not available in the literature as the proofs are based on dynamic programming, which starts at a finite terminal time stage and optimal policies are computed by working backwards from the end. However, as in [8], we can prove an infinite horizon analog of Theorem 2 assuming that $\{X_t\}$ starts from its invariant measure π^* (which exists e.g. if $\{X_t\}$ is irreducible and aperiodic).

Proposition 3 ([8, Theorem 6]): Assume $\{X_t\}$ is a stationary Markov chain with invariant probability π^*. Then there exists an optimal policy in Π_W that solves the minimization problem (12), i.e., there exists $\Pi \in \Pi_W$ such that

$$
J_{\pi^*}(\Pi) = J_{\pi^*}.
$$

The proof of the proposition is straightforward; it relies on a construction that pieces together policies from Π_W that on time segments of appropriately large lengths increasingly well approximate the infimum of the infinite horizon cost achievable by policies in Π_A; see [8] for the details. This construction results in a policy that is nonstationary in general. However, for the finite alphabet case considered here, we will also establish the optimality of deterministic *stationary* policies even for possibly nonstationary Markov sources. The remainder of the section focuses on this problem.

B. Optimality of Stationary Coding Policies

The following theorem is the main result of the paper. It states that for any irreducible and aperiodic Markov source there exists a stationary Markov (Walrand-Varaiya type) coding policy that is optimal among all zero delay coding policies. Note that the theorem does not require the source to be stationary.

Theorem 3: Assume that $\{X_t\}$ is an irreducible and aperiodic Markov chain. Then for any initial distribution π_0,

$$
\inf_{\Pi \in \Pi_A} J_{\pi_0}(\Pi) = \min_{\Pi \in \Pi_{WS}} J_{\pi_0}(\Pi).
$$

Furthermore, there exist $\Pi^* \in \Pi_{WS}$ that achieves the minimum above simultaneously for all π_0 and which satisfies for all $T \geq 1$

$$
\frac{1}{T} E_{\pi_0}^{\Pi^*} \left[\sum_{t=0}^{T-1} d(X_t, \hat{X}_t) \right] \leq J_{\pi_0} + \frac{K}{T} \qquad (13)
$$

for some positive constant K.

The theorem is proved in the next subsection where the constant K is more explicitly identified. Here we give a brief description of the main steps. The key step in the proof is Lemma 1 where we build on the approach of Borkar [42] (but use a different construction) to show that for any two initial distributions μ_0 and ζ_0, the absolute difference of the optimal infinite horizon discounted costs $J_{\mu_0}^{\beta}$ and $J_{\zeta_0}^{\beta}$ is uniformly upper bounded by a constant times the L_1 Wasserstein distance between μ_0 and ζ_0. With the aid of this result and an Abelian lemma that relates the infinite horizon discounted cost to the average cost, Lemma 2 shows through the vanishing discount approach that for the infinite horizon average cost problem,

randomized stationary Markov policies are at least as good as deterministic policies in Π_A. Lemma 4 in turn shows that deterministic stationary Markov policies are no worse than randomized ones, which, together with Lemma 2, gives $\inf_{\Pi \in \Pi_A} J_{\pi_0}(\Pi) = \inf_{\Pi \in \Pi_{WS}} J_{\pi_0}(\Pi)$ (Lemma 5). Finally, we show that Lemma 1 implies that the *average cost optimality equation* (ACOE) (see Theorem 10 in the Appendix) holds for our controlled Markov chain, which in turn implies that the infimum $\inf_{\Pi \in \Pi_{WS}} J_{\pi_0}(\Pi)$ is achieved by some policy in Π_{WS}, proving the first statement of the theorem. The $O(1/T)$ convergence rate result is shown to be a direct consequence of the ACOE.

Definition 2 (ε-Optimality): Given an initial distribution π_0 and $\epsilon > 0$, a policy $\Pi \in \Pi_A$ is ϵ-optimal if $J_{\pi_0}(\Pi) \leq J_{\pi_0} + \epsilon$, where J_{π_0} is the optimal performance for the infinite horizon average cost problem.

Now suppose that $\{X_t\}$ is irreducible and aperiodic and it starts from the unique invariant probability π^* so that it is a stationary process. Consider the (nonstationary) coding policy that is obtained by periodically extending an initial segment of the optimal stationary policy Π^* in Theorem 3. In particular, assume $\Pi^* = \{\eta^*\}$ and for $T \geq 1$ consider the periodic policy $\Pi^{(T)} = \{\eta_t^{(T)}\}$, where $\eta_t^{(T)} = \eta^*$ for $t = kT+1, \ldots$, $(k+1)T$, $k = 0, 1, 2, \ldots$, and $\eta_t^{(T)} \equiv \eta^*(\pi^*)$ for $t = kT$, $k = 0, 1, 2, \ldots$. Since $\{X_t\}$ is stationary, the infinite horizon cost of $\Pi^{(T)}$ is

$$J_{\pi^*}(\Pi^{(T)}) = \frac{1}{T} E_{\pi_0}^{\Pi^*} \left[\sum_{t=0}^{T-1} d(X_t, \hat{X}_t) \right].$$

Since the encoder of $\Pi^{(T)}$ is reset to $\eta^*(\pi^*)$ each time after processing T source samples, we can say that it has memory length T. The following result, which is implied by the above construction and the bound (13), may have implications in the construction of practical codes since, loosely speaking, the complexity of a code is determined by its memory length.

Theorem 4: Assume $\{X_t\}$ is an irreducible and aperiodic Markov chain. If $X_0 \sim \pi^*$, where π^* is the invariant probability measure, then for every $\epsilon > 0$, there exists a finite memory, nonstationary, but periodic coding policy with period at most $\frac{K}{\epsilon}$ that is ϵ-optimal, where K is the constant from Theorem 3.

C. Proof of Theorem 3

Let $\mathsf{X} = \{1, \cdots, |\mathsf{X}|\}$ be viewed as a subset of $\mathbb{R}$. The L_1 *Wasserstein distance* [41] between two distributions μ_0 and ζ_0 is defined as

$$\rho_1(\mu_0, \zeta_0) := \inf_{X \sim \mu_0, Y \sim \zeta_0} E[|X - Y|], \quad (14)$$

where the infimum is taken over all joint distributions of pairs of X-valued random variables (X, Y) such that $X \sim \mu_0$ and $Y \sim \zeta_0$. It can be shown that the infimum in the definition is in fact a minimum and that the L_1 Wasserstein distance is a metric on $\mathcal{P}(\mathsf{X})$.

Recall the definition

$$J_{\pi_0}^{\beta} := \inf_{\Pi \in \Pi_A} \lim_{T \to \infty} E_{\pi_0}^{\Pi} \left[\sum_{t=0}^{T-1} \beta^t d(X_t, \hat{X}_t) \right].$$

The following lemma is a key step in the proof.

Lemma 1: Suppose the source is an irreducible and aperiodic Markov chain. Then for any pair of initial distributions μ_0 and ζ_0, and any $\beta \in (0, 1)$, we have

$$\left| J_{\mu_0}^{\beta} - J_{\zeta_0}^{\beta} \right| \leq K_1 \|d\|_{\infty} \rho_1(\mu_0, \zeta_0),$$

where K_1 is a finite constant and $\|d\|_{\infty} = \max_{x,y} d(x, y)$.

Proof: Note that by monotone convergence for any Π and $\beta \in (0, 1)$,

$$\lim_{T \to \infty} E \left[\sum_{t=0}^{T-1} \beta^t d(X_t, \hat{X}_t) \right] = E \left[\sum_{t=0}^{\infty} \beta^t d(X_t, \hat{X}_t) \right].$$

Thus the lemma statement is equivalent to

$$\left| \inf_{\Pi \in \Pi_A} E_{\mu_0}^{\Pi} \left[\sum_{t=0}^{\infty} \beta^t d(X_t, \hat{X}_t) \right] \right.$$
$$\left. - \inf_{\Pi \in \Pi_A} E_{\zeta_0}^{\Pi} \left[\sum_{t=0}^{\infty} \beta^t d(X_t, \hat{X}_t) \right] \right|$$
$$\leq K_1 \|d\|_{\infty} \rho_1(\mu_0, \zeta_0).$$

The proof builds on the approach of Borkar [42] (see also [43] and [26]), but our argument is different (and also more direct) since the absolute continuity conditions in [42] are not applicable here due to quantization. As in [26], in the proof we will enlarge the space of admissible coding policies to allow for randomization at the encoder. Since for a discounted infinite horizon optimal encoding problem optimal policies are deterministic even among possibly randomized policies (see Proposition 2), allowing common randomness does not change the optimal performance.

In our construction, we will use the well known coupling method for Markov chains. In particular, we will use a minor modification of the coupling argument in the proof of [44, Theorem 1.8.3].

Given the initial distributions μ_0 and ζ_0, let λ be a joint distribution on $\mathsf{X} \times \mathsf{X}$, having marginals μ_0 and ζ_0, such that $E[|X - Y|] = \rho_1(\mu_0, \zeta_0)$ if $(X, Y) \sim \lambda$.

Now consider the $\mathsf{X} \times \mathsf{X}$-valued process $\{(X_t, Y_t)\}_{t \geq 0}$ such that $\{X_t\}_{t \geq 0} \sim (\mu_0, P)$, $\{Y_t\}_{t \geq 0} \sim (\zeta_0, P)$, $(X_0, Y_0) \sim \lambda$, and $\{X_t\}_{t \geq 1}$ and $\{Y_t\}_{t \geq 1}$ are conditionally independent given (X_0, Y_0). We note that given $\{X_t\} \sim (\mu_0, P)$, a process $\{(X_t, Y_t)\}$ with such a distribution can be obtained via an i.i.d. randomization process $\{W_t\}$ which is uniform on the interval $[0, 1]$ and is independent of $\{X_t\}$, and via appropriate functions $F_0, F : \mathsf{X} \times [0, 1] \to \mathsf{X}$, by letting

$$Y_0 = F_0(X_0, W_0) \quad \text{and} \quad Y_t = F(Y_{t-1}, W_t) \quad \text{for all } t \geq 1.$$

$$(15)$$

Fixing a reference state $b \in \mathsf{X}$, define

$$\tau = \inf\{t \geq 0 : X_n = Y_n = b\}.$$

Since the common transition probability P of $\{X_t\}$ and $\{Y_t\}$ is irreducible and aperiodic, it easily follows that $\{(X_t, Y_t)\}$ is an irreducible and aperiodic Markov chain [44, p. 41]. Since X is finite, this implies that the chain is positive recurrent and

thus $E[\tau] < \infty$. Define $X'_t = X_t$ for $t \geq 0$ so that $\{X'_t\} \sim (\mu_0, P)$. Also define the process $\{X''_t\}$ by

$$X''_t = \begin{cases} Y_t & \text{if } t \leq \tau \\ X_t & \text{if } t > \tau. \end{cases}$$

It is shown in [44, p. 42] that $\{X''_t\}$ is a Markov chain such that $\{X''_t\} \sim (\zeta_0, P)$.

Assume without loss of generality that $J^\beta_{\mu_0} - J^\beta_{\zeta_0} \geq 0$. Then from the above

$$\left| J^\beta_{\mu_0} - J^\beta_{\zeta_0} \right| = J^\beta_{\mu_0} - J^\beta_{\zeta_0}$$

$$= \inf_{\Pi \in \Pi_A} E^\Pi_{\mu_0} \left[\sum_{t=0}^{\infty} \beta^t d(X'_t, \hat{X}'_t) \right]$$

$$- \inf_{\Pi \in \Pi_A} E^\Pi_{\zeta_0} \left[\sum_{t=0}^{\infty} \beta^t d(X''_t, \hat{X}''_t) \right] \quad (16)$$

$$= E^{\Pi'}_{\mu_0} \left[\sum_{t=0}^{\infty} \beta^t d(X'_t, \hat{X}'_t) \right]$$

$$- E^{\Pi''}_{\zeta_0} \left[\sum_{t=0}^{\infty} \beta^t d(X''_t, \hat{X}''_t) \right], \quad (17)$$

where Π' (resp. Π'') achieves the first (resp. the second) infimum in (16); see Proposition 2.

Consider the following suboptimal coding and decoding policy for $\{X'_t\}$: In addition to observing the source $X'_t = X_t$, $t \geq 0$, the encoder is also given the randomization process $\{W_t\}$ which is independent of $\{X'_t\}$. Then the encoder can generate $Y_0, \ldots, Y_\tau$ according to the representation (15) and thus it can produce the second source process $\{X''_t\}$. The encoder for $\{X'_t\}$ feeds sequentially the $\{X''_t\}$ values to the quantizer policy Π'' and produces the same channel symbols q''_t and reproduction sequence $\hat{X}''_t = \gamma''_t(q''_{[0,t]})$ as the policy Π'' does in response to $\{X''_t\}$. Note that this procedure comprises a suboptimal *randomized encoder* and a *deterministic decoder* for coding $\{X'_t\}$. Let us denote this randomized policy by $\hat{\Pi}$. Thus we obtain the upper bound

$$E^{\Pi'}_{\mu_0} \left[\sum_{t=0}^{\infty} \beta^t d(X'_t, \hat{X}'_t) \right] \leq E^{\hat{\Pi}}_{\mu_0} \left[\sum_{t=0}^{\infty} \beta^t d(X'_t, \hat{X}''_t) \right].$$

In view of this and (17), we can write

$$\left| J^\beta_{\mu_0} - J^\beta_{\zeta_0} \right| \leq E^{\hat{\Pi}}_{\mu_0} \left[\sum_{t=0}^{\infty} \beta^t d(X'_t, \hat{X}''_t) \right]$$

$$- E^{\Pi''}_{\zeta_0} \left[\sum_{t=0}^{\infty} \beta^t d(X''_t, \hat{X}''_t) \right] \quad (18)$$

$$\leq \left| E \left[\sum_{t=0}^{\infty} \beta^t \left(d(X'_t, \hat{X}''_t) - d(X''_t, \hat{X}''_t) \right) \right] \right|$$

$$\leq E[\tau] \|d\|_\infty, \quad (19)$$

where the last inequality follows since $X'_t = X''_t$ if $t \geq \tau$.

On the other hand,

$$E[\tau] = \sum_{x,y} \lambda(x, y) E[\tau | X_0 = x, Y_0 = y]$$

and since $E[\tau] < \infty$, we have that

$$K_1 := \max_{x,y} E[\tau | X_0 = x, Y_0 = y] < \infty$$

and

$$E[\tau] \leq \sum_{x \neq y} \lambda(x, y) K_1$$

$$= \Pr(X_0 \neq Y_0) K_1 \leq K_1 \rho_1(\mu_0, \zeta_0),$$

where the the second inequality follows from the fact that $\Pr(X_0 \neq Y_0) \leq \rho_1(\mu_0, \zeta_0)$ by (14). This and (19) complete the proof of Lemma 1. □

Under any given stationary Markov policy $\Pi \in \Pi_{WS}$ the sequence $\{(\pi_t, Q_t)\}_{t \geq 0}$ is a $\mathcal{P}(\mathsf{X}) \times \mathcal{Q}$-valued Markov chain whose transition kernel is determined by Π and the transition kernel $P(d\pi_{t+1} | \pi_t, Q_t)$, which is given by the filtering equation (6) and does not depend on Π. As pointed out in the proof of Proposition 2, the transition kernel $P(d\pi_{t+1} | \pi_t, Q_t)$ is weakly continuous. This implies that the Markov process $\{(\pi_t, Q_t)\}$ is weak Feller, that is, the transition kernel $P(d(\pi_{t+1}, Q_{t+1}) | \pi_t, Q_t)$ is weakly continuous [40, C.3 Definition]. Since every weak Feller Markov process with a compact state space has an invariant probability measure [45], it follows that there exists a probability measure $\pi^*(\Pi)$ on $\mathcal{P}(\mathsf{X})$ such that if π_0 is picked randomly according to $\pi^*(\Pi)$, then $\{(\pi_t, Q_t)\}$ is a stationary process. We call $\pi^*(\Pi)$ an invariant probability on $\mathcal{P}(\mathsf{X})$ induced by $\Pi \in \Pi_{WS}$.

Note that if the initial probability π_0 is random with distribution $\pi^*(\Pi)$, the quantization policy Π becomes a randomized policy since the encoder and decoder must have access to the same random π_0. Expectations under such policies will be denoted by $E^\Pi_{\pi_0 \sim \pi^*(\Pi)}$.

Lemma 2: If the source is irreducible and aperiodic, then for any initial distribution π,

$$\inf_{\Pi \in \Pi_A} \limsup_{T \to \infty} \frac{1}{T} E^\Pi_\pi \left[\sum_{t=0}^{T-1} d(X_t, \hat{X}_t) \right]$$

$$\geq \inf_{\Pi \in \Pi_{WS}} \limsup_{T \to \infty} \frac{1}{T} E^\Pi_{\pi_0 \sim \pi^*(\Pi)} \left[\sum_{t=0}^{T-1} d(X_t, \hat{X}_t). \right]$$

Proof: We will need the following well known Abelian result.

Lemma 3 ([40, Lemma 5.3.1]): Let $\{c_t\}_{t \geq 0}$ be a sequence of nonnegative numbers. Then

$$\liminf_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} c_t \leq \liminf_{\beta \uparrow 1} (1 - \beta) \sum_{t=0}^{\infty} \beta^t c_t$$

$$\leq \limsup_{\beta \uparrow 1} (1 - \beta) \sum_{t=0}^{\infty} \beta^t c_t$$

$$\leq \limsup_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} c_t.$$

Let $\{\Pi_k\}$ be a sequence of policies in Π_A such that $\lim_{k \to \infty} J_\pi(\Pi_k) = J_\pi$ and fix $n > 0$ such that

$$J_\pi \geq J_\pi(\Pi_n) - \epsilon. \quad (20)$$

Applying Lemma 3 with $c_t = E_\pi^{\Pi_n}[d(X_t, \hat{X}_t)]$, there exists $\beta_\epsilon \in (0, 1)$ such that for all $\beta \in (\beta_\epsilon, 1)$

$$J_\pi \geq J_\pi(\Pi_n) - \epsilon$$

$$\geq (1 - \beta) E_\pi^{\Pi_n} \left[\sum_{t=0}^\infty \beta^t d(X_t, \hat{X}_t) \right] - 2\epsilon$$

$$\geq \min_{\Pi \in \Pi_{WS}} (1 - \beta) E_\pi^\Pi \left[\sum_{t=0}^\infty \beta^t d(X_t, \hat{X}_t) \right] - 2\epsilon, \quad (21)$$

where the minimum exists by Proposition 2.

Now consider the case where for some $\Pi \in \Pi_{WS}$ the initial measure π_0 is distributed according to $\pi_0 \sim \pi^*(\Pi)$. For ease of interpretation, let X_t' denote the source process with $X_0' \sim \pi$, let X_t'' be a process with $X_0'' \sim \pi_0$ for some fixed π_0, and in addition let X_t' and X_t'' be coupled as in Lemma 1. Then for any $\beta \in (0, 1)$,

$$\left| \min_{\Pi \in \Pi_{WS}} (1 - \beta) E_\pi^\Pi \left[\sum_{t=0}^\infty \beta^t d(X_t', \hat{X}_t') \right] \right.$$

$$\left. - \min_{\Pi \in \Pi_{WS}} (1 - \beta) E_{\pi_0}^\Pi \left[\sum_{t=0}^\infty \beta^t d(X_t'', \hat{X}_t'') \right] \right|$$

$$\leq (1 - \beta) E[\tau] \|d\|_\infty$$

$$\leq (1 - \beta) K_1 \|d\|_\infty \rho_1(\pi, \pi_0), \quad (22)$$

where $\tau = \min\{t \geq 0 : X_t' = X_t''\}$, and where the first inequality follows from the coupling of the Markov chains as in Lemma 1 (see (19)) and the second also follows from the proof of Lemma 1. Since $\rho_1(\pi, \pi_0)$ is upper bounded by $|\mathsf{X}|$ (see (14)) for any π and π_0, we obtain

$$\left| \min_{\Pi \in \Pi_{WS}} (1 - \beta) E_\pi^\Pi \left[\sum_{t=0}^\infty \beta^t d(X_t', \hat{X}_t') \right] \right.$$

$$\left. - \inf_{\Pi \in \Pi_{WS}} (1 - \beta) E_{\pi_0 \sim \pi^*(\Pi)}^\Pi \left[\sum_{t=0}^\infty \beta^t d(X_t'', \hat{X}_t'') \right] \right|$$

$$\leq (1 - \beta) K_2 \|d\|_\infty, \quad (23)$$

where $K_2 = K_1 |\mathsf{X}|$.

Choosing $\bar{\beta}$ such that (23) with $\beta = \bar{\beta}$ is less than ϵ, and combining the preceding bound with (21) yields for any $\beta \in (\max\{\beta_\epsilon, \bar{\beta}\}, 1)$,

$$J_\pi \geq \min_{\Pi \in \Pi_{WS}} (1 - \beta) E_\pi^\Pi \left[\sum_{t=0}^\infty \beta^t d(X_t, \hat{X}_t) \right] - 2\epsilon$$

$$\geq \inf_{\Pi \in \Pi_{WS}} (1 - \beta) E_{\pi_0 \sim \pi^*(\Pi)}^\Pi \left[\sum_{t=0}^\infty \beta^t d(X_t, \hat{X}_t) \right] - 3\epsilon \quad (24)$$

$$\geq (1 - \beta) E_{\pi_0 \sim \pi^*(\Pi_\beta)}^{\Pi_\beta} \left[\sum_{t=0}^\infty \beta^t d(X_t, \hat{X}_t) \right] - 4\epsilon$$

$$\geq \liminf_{T \to \infty} \frac{1}{T} E_{\pi_0 \sim \pi^*(\Pi_\beta)}^{\Pi_\beta} \left[\sum_{t=0}^{T-1} d(X_t, \hat{X}_t) \right] - 5\epsilon$$

$$= \limsup_{T \to \infty} \frac{1}{T} E_{\pi_0 \sim \pi^*(\Pi_\beta)}^{\Pi_\beta} \left[\sum_{t=0}^{T-1} c(\pi_t, Q_t) \right] - 5\epsilon,$$

where the $\Pi_\beta \in \Pi_{WS}$ is chosen so that it achieves the infimum in (24) within ϵ, and where the fourth inequality holds by

Lemma 3 if $\beta \in (\max\{\beta_\epsilon, \bar{\beta}\}, 1)$ is large enough. Finally, the last equality follows since $\pi^*(\Pi_\beta)$ is invariant and hence $\{(\pi_t, Q_t)\}$ is a stationary process. Thus we obtain

$$J_\pi \geq \limsup_{T \to \infty} \frac{1}{T} E_{\pi_0 \sim \pi^*(\Pi_\beta)}^{\Pi'} \left[\sum_{t=0}^{T-1} d(X_t, \hat{X}_t) \right] - 5\epsilon$$

$$\geq \inf_{\Pi \in \Pi_{WS}} \limsup_{T \to \infty} \frac{1}{T} E_{\pi_0 \sim \pi^*(\Pi)}^\Pi \left[\sum_{t=0}^{T-1} d(X_t, \hat{X}_t) \right] - 5\epsilon,$$

where $\epsilon > 0$ is arbitrary, which completes the proof. $\square$

Lemma 4: If the source is irreducible and aperiodic, then for any initial distribution π,

$$\inf_{\Pi \in \Pi_{WS}} \limsup_{T \to \infty} \frac{1}{T} E_\pi^\Pi \left[\sum_{t=0}^{T-1} d(X_t, \hat{X}_t) \right]$$

$$= \inf_{\Pi \in \Pi_{WS}} \limsup_{T \to \infty} \frac{1}{T} E_{\pi_0 \sim \pi^*(\Pi)}^\Pi \left[\sum_{t=0}^{T-1} d(X_t, \hat{X}_t) \right].$$

Proof: First note that by Lemma 2,

$$\inf_{\Pi \in \Pi_{WS}} \limsup_{T \to \infty} \frac{1}{T} E_\pi^\Pi \left[\sum_{t=0}^{T-1} d(X_t, \hat{X}_t) \right]$$

$$\geq \inf_{\Pi \in \Pi_{WS}} \limsup_{T \to \infty} \frac{1}{T} E_{\pi_0 \sim \pi^*(\Pi)}^\Pi \left[\sum_{t=0}^{T-1} d(X_t, \hat{X}_t) \right].$$

Now apply the argument that led to the bounds (22) and (23) to obtain

$$\left| \inf_{\Pi \in \Pi_{WS}} \limsup_{T \to \infty} \frac{1}{T} E_\pi^\Pi \left[\sum_{t=0}^{T-1} d(X_t, \hat{X}_t) \right] \right.$$

$$\left. - \inf_{\Pi \in \Pi_{WS}} \limsup_{T \to \infty} \frac{1}{T} E_{\pi_0 \sim \pi^*(\Pi)}^\Pi \left[\sum_{t=0}^{T-1} d(X_t, \hat{X}_t) \right] \right|$$

$$\leq \limsup_{T \to \infty} \frac{1}{T} E_{\pi_0 \sim \pi^*(\Pi)}[\tau] \|d\|_\infty$$

$$\leq \limsup_{T \to \infty} \frac{1}{T} K_2 \|d\|_\infty = 0.$$

$\square$

The following important result immediately follows from Lemmas 2 and 4.

Lemma 5: If the source is irreducible and aperiodic, then for any initial distribution π_0,

$$\inf_{\Pi \in \Pi_A} \limsup_{T \to \infty} \frac{1}{T} E_{\pi_0}^\Pi \left[\sum_{t=0}^{T-1} d(X_t, \hat{X}_t) \right]$$

$$= \inf_{\Pi \in \Pi_{WS}} \limsup_{T \to \infty} \frac{1}{T} E_{\pi_0}^\Pi \left[\sum_{t=0}^{T-1} d(X_t, \hat{X}_t) \right]. \quad (25)$$

Remark: This lemma is crucial because it shows that without any loss we can restrict the search for optimal quantization policies to the set Π_W. Since the filtering equation (6) leads to a controlled Markov chain only for policies in Π_W, this lemma allows us to apply controlled Markov chain techniques in the study of the the average distortion problems. The rigorous

justification of this fact is one of the main contributions of this paper.

Note that Lemma 5 immediately implies the first statement of Theorem 3 once we can show that the infimum in (25) is actually a minimum. This will be done by invoking the ACOE for controlled Markov chains. To show that the infimum is achieved by a stationary and deterministic Markov policy $\Pi \in \Pi_{WS}$ we make use of Theorem 10 in the Appendix. To do this we have to verify that the conditions of the theorem are satisfied with $Z = \mathcal{P}(X)$, $A = \mathcal{Q}$, $c(z, a) = c(\pi, Q)$, and $K(dz'|z, a) = P(\pi'|\pi, Q)$. We have already shown in the proof of Proposition 2 that conditions (i)—(iii) hold. Since $\mathcal{P}(X)$ is the standard probability simplex in $\mathbb{R}^{|X|}$ and $\mathcal{Q}$ is a finite set, condition (iv) clearly holds. Finally, condition (v) holds since the family of functions

$$\left\{ h_\beta(\zeta) := J_\zeta^\beta - J_{\zeta_0}^\beta : \beta \in (0, 1) \right\}$$

for some arbitrary but fixed $\zeta_0 \in \mathcal{P}(X)$ is equicontinuous on $\mathcal{P}(X)$ by Lemma 1 which states that

$$|h_\beta(\zeta) - h_\beta(\zeta')| = \left| J_\zeta^\beta - J_{\zeta'}^\beta \right| \leq K_1 \|d\|_\infty \rho_1(\zeta, \zeta'). \quad (26)$$

Thus we can apply Theorem 10 to deduce the existence of a policy in Π_{WS} achieving the minimum in (25). This completes the proof of the first statement in Theorem 3.

To prove the second statement (13) we use the result in (38) in the Appendix. Note that by Lemma 1 we have for all $\beta \in (0, 1)$ and $\zeta \in \mathcal{P}(X)$

$$|h_\beta(\zeta)| = \left| J_\zeta^\beta - J_{\zeta'}^\beta \right| \leq \frac{K}{2},$$

where

$$K := 2K_1 \|d\|_\infty \rho_1(\zeta, \zeta_0) \leq 2K_1 \|d\|_\infty |X|.$$

Thus equation (38) implies, with $z_0 = \pi_0$, $g^* = J_{\pi_0}$, and Π^* being the optimal policy in Π_{WS} achieving the minimum in (25), that

$$J_{\pi_0}(\Pi^*, T) - J_{\pi_0} \leq \frac{K}{T}$$

as claimed.   □

IV. Zero-Delay Coding Over a Noisy Channel With Feedback

In this section, we briefly describe the extension of our main results to zero delay lossy coding over a noisy channel. As in Section I-A, the encoder processes the observed information source without delay. It is assumed that the source $\{X_t\}_{t \geq 0}$ is a discrete time Markov process with finite alphabet X. The encoder encodes the source samples without delay and transmits the encoded versions to a receiver over a discrete channel with input alphabet $M = \{1, \ldots, M\}$ and output alphabet $M' := \{1, \ldots, M'\}$, where M and M' are positive integers.

In contrast with the setup described in Section I-A, here the channel between the encoder and decoder is a discrete and memoryless *noisy* channel characterized by the transition probability $T(b|a) = \Pr(q' = b|q = a)$, $a \in M$, $b \in M'$.

We assume that the encoder has access to the previous channel outputs in the form of *feedback*. In particular, the encoder is specified by a *quantization policy* Π, which is a sequence of functions $\{\eta_t\}_{t \geq 0}$ with $\eta_t : M^t \times (M')^t \times X^{t+1} \to M$. At time t, the encoder transmits the M-valued message

$$q_t = \eta_t(I_t),$$

where $I_0 = X_0$, $I_t = (q_{[0,t-1]}, q'_{[0,t-1]}, X_{[0,t]})$ for $t \geq 1$, and q'_t is the received (noisy) version of q_t. The collection of all such zero delay policies is called the set of admissible quantization policies and is denoted by Π_A.

Upon receiving q'_t, the receiver generates the reconstruction, $\hat{X}_t$, also without delay. A zero delay receiver policy is a sequence of functions $\gamma = \{\gamma_t\}_{t \geq 0}$ of type $\gamma_t : (M')^{t+1} \to \hat{X}$, where $\hat{X}$ is the finite reproduction alphabet. Thus

$$\hat{X}_t = \gamma_t(q'_{[0,t]}) \quad \text{for all } t \geq 0.$$

Note that, due to the presence of feedback, the encoder also has access to $q'_{[0,t]}$ at time $t+1$. The finite and infinite horizon coding problems are defined analogously to the noiseless case.

The following result is a known extension of Witsenhausen's structure theorem [14].

Theorem 5 ([4, Theorem 10.7.1]): For the problem of transmitting T samples of a Markov source over a noisy channel with feedback, 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 $t = 1, \ldots, T - 1$.

Given a quantization policy Π, for all $t \geq 1$ let $\pi_t \in \mathcal{P}(X)$ be the conditional probability defined by

$$\pi_t(A) := \Pr(X_t \in A|q'_{[0,t-1]})$$

for any set $A \subset X$.

The following result is due to Walrand and Varaiya.

Theorem 6 ([15]): For the problem of transmitting T samples of a Markov source over a noisy channel with feedback, 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 π_t to select a quantizer $Q_t : X \to M$ and then q_t is generated as $q_t = Q_t(X_t)$.

Under a Walrand-Varaiya type policy the filtering equation (6) is modified as

$$\pi_{t+1}(x_{t+1})$$
$$= \frac{\sum_{x_t, q_t} \pi_t(x_t) T(q'_t|q_t) P(q_t|\pi_t, x_t) P(x_{t+1}|x_t)}{\sum_{x_t, q_t} \sum_{x_{t+1}} \pi_t(x_t) T(q'_t|q_t) P(q_t|\pi_t, x_t) P(x_{t+1}|x_t)}.$$

Thus, as before, given π_t and Q_t, π_{t+1} is conditionally independent of $(\pi_{[0,t-1]}, Q_{[0,t-1]})$ and it follows that $\{\pi_t\}$ can be viewed as $\mathcal{P}(X)$-valued controlled Markov process [40] with $\mathcal{Q}$-valued control $\{Q_t\}$ and average cost up to time $T - 1$ given by

$$E_{\pi_0}^{\Pi, \gamma} \left[\frac{1}{T} \sum_{t=0}^{T-1} d(X_t, \hat{X}_t) \right] = E_{\pi_0}^{\Pi} \left[\frac{1}{T} \sum_{t=0}^{T-1} c(\pi_t, Q_t) \right].$$

The set of deterministic Markov coding policies Π_W and deterministic stationary Markov policies Π_{WS} is defined analogously to Definition 1.

It can be checked that the properties concerning the continuity of the kernel and the existence of invariant measures apply identically to the new controlled Markov state pair (π_t, Q_t). Under the assumption that $\{X_t\}$ is irreducible and aperiodic, the simulation argument also applies identically by considering the same channel noise realizations for both processes X'_t and X''_t; i.e., in the simulation argument we can compare the performance of the coding schemes by taking the expectations over the channel noise realizations. Thus, the finite coupling time argument in Lemma 1 applies to this case as well. The following theorem compactly summarizes the noisy channel analogues of our results in the previos sections.

Theorem 7:

(i) *For the minimization of the finite horizon average distortion (3), an optimal solution in Π_W exists and a noisy channel analog of Proposition 1 holds.*

(ii) *For the minimization of the infinite horizon discounted distortion (4), an optimal solution exists and such a solution is in Π_{WS}, i.e., a noisy channel analog of Proposition 2 holds.*

(iii) *The noisy channel version of Theorem 3 holds: If $\{X_t\}$ is irreducible and aperiodic, there exists a policy in Π_{WS} that minimizes the infinite horizon average distortion (5). Furthermore, the convergence rate result (13) holds for this optimal policy.*

(iv) *Under the assumption that $\{X_t\}$ is irreducible and aperiodic, if $X_0 \sim \pi^*$, where π^* is the invariant probability measure, for any $\epsilon > 0$, there exists $K > 0$ and a finite memory, nonstationary, but periodic quantization policy with period less than $\frac{K}{\epsilon}$ that achieves ϵ-optimal performance Thus the noisy channel version of Theorem 4 holds.*

V. CONCLUSION

Zero delay lossy coding of finite alphabet Markov sources was considered. The main result showed that for any irreducible and aperiodic (not necessarily stationary) Markov chain there exists a stationary and deterministic Markov (Walrand-Varaiya type) policy that is optimal in the set of zero delay coding policies. This result significantly generalizes existing results in [14], [15], and [8].

In addition, it was shown that the distortion of an optimal stationary policy for time horizon (block length) T converges to the optimal infinite horizon distortion at a rate $O(1/T)$. As a corollary, the ϵ-optimality of periodic zero delay codes is established with an explicit bound on the relationship between ϵ and the period length. This result is of potential practical importance since the code's complexity is directly related to the length of the period (memory size). Extensions of these results to zero delay lossy coding over noisy channels with feedback were also given.

An interesting open problem is the generalization of the results to continuous sources such as real or $\mathbb{R}^d$-valued Markov sources. Such a generalization would be facilitated by an appropriate extension of Lemma 1 to continous alphabets.

Some related results in this direction are available in [46]. Another, more challenging open problem of information theoretic flavor is to find a (preferably) single-letter characterization of the optimum infinite horizon average distortion of zero delay coding of Markov sources. As mentioned before, such a characterization is only known for stationary and memoryless (i.i.d.) sources, while for the block coding problem the distortion rate function gives a (non single-letter) characterization, and even closed form expressions exist for binary symmetric Markov sources in a certain range of distortion values [47] as well as explicit lower and upper bounds [48].

VI. APPENDIX
MARKOV DECISION PROCESSES

Let Z be a Borel space (i.e., a Borel subset of a complete and separable metric space) and let $\mathcal{P}(\mathsf{Z})$ denote the set of all probability measures on Z.

Definition 3 (Markov Control Model [40]): A discrete time *Markov control model* (*Markov decision process*) is a system characterized by the 4-tuple

$$(\mathsf{Z}, \mathsf{A}, K, c),$$

where

1) Z is the *state space*, the set of all possible states of the system;

2) A (a Borel space) is the *control space* (or action space), the set of all controls (actions) $a \in \mathsf{A}$ that can act on the system;

3) $K = K(\cdot | z, a)$ is the *transition probability* of the system, a stochastic kernel on Z given $\mathsf{Z} \times \mathsf{A}$, i.e., $K(\cdot | z, a)$ is a probability measure on Z for all state-action pairs (z, a), and $K(B | \cdot, \cdot)$ is a measurable function from $\mathsf{Z} \times \mathsf{A}$ to $[0, 1]$ for each Borel set $B \subset \mathsf{Z}$;

4) $c : \mathsf{Z} \times \mathsf{A} \to [0, \infty)$ is the *cost per time stage function* of the system, a function $c(x, a)$ of the state and the control.

Define the *history* spaces H_t at time $t \geq 0$ of the Markov control model by $\mathsf{H}_0 := \mathsf{Z}$ and $\mathsf{H}_t := (\mathsf{Z} \times \mathsf{A})^t \times \mathsf{Z}$. Thus a specific history $h_t \in \mathsf{H}_t$ has the form $h_t = (z_0, a_0, \ldots, z_{t-1}, a_{t-1}, z_t)$.

Definition 4 (Admissible Control Policy [40]): An *admissible control policy* $\Pi = \{\alpha_t\}_{t \geq 0}$, also called a *randomized control policy* (more simply a *control policy* or a *policy*) is a sequence of stochastic kernels on the action space A given the history H_t. The set of all randomized control policies is denoted by Π_A. A *deterministic policy* Π is a sequence of functions $\{\alpha_t\}_{t \geq 0}$, $\alpha_t : \mathsf{H}_t \to \mathsf{A}$, that determine the control used at each time stage deterministically, i.e., $a_t = \alpha_t(h_t)$. The set of all deterministic policies is denoted Π_D. Note that $\Pi_D \subset \Pi_A$. A *Markov policy* is a policy Π such that for each time stage the choice of control only depends on the current state z_t, i.e., $\Pi = \{\alpha_t\}_{t \geq 0}$ with $\alpha_t : \mathsf{Z} \to \mathcal{P}(\mathsf{A})$. The set of all Markov policies is denoted by Π_M. The set of deterministic Markov policies is denoted by Π_{MD}. A *stationary policy* is a Markov policy $\Pi = \{\alpha_t\}_{t \geq 0}$ such that $\alpha_t = \alpha$ for all $t \geq 0$ for some $\alpha : \mathsf{Z} \to \mathcal{P}(\mathsf{A})$. The set of all stationary policies is denoted by Π_S and the set of deterministic stationary policies is denoted by Π_{SD}.

According to the Ionescu Tulcea theorem (see [40]), the transition kernel K, an initial probability distribution π_0 on Z, and a policy Π define a unique probability measure $P_{\pi_0}^{\Pi}$ on $\mathsf{H}_{\infty} = (\mathsf{X} \times \mathsf{A})^{\infty}$, the distribution of the state-action process $\{(Z_t, A_t)\}_{t \geq 0}$. The resulting state process $\{Z_t\}_{t \geq 0}$ is called a *controlled Markov process*. The expectation with respect to $P_{\pi_0}^{\Pi}$ is denoted by $E_{\pi_0}^{\Pi}$. If $\pi_0 = \delta_z$, the point mass at $z \in \mathsf{Z}$, we write P_z^{Π} and E_z^{Π} instead of $P_{\delta_z}^{\Pi}$ and $E_{\delta_z}^{\Pi}$.

In an *optimal control problem*, a performance objective J of the system is given and the goal is to find the controls that minimize (or maximize) that objective. Some common optimal control problems for Markov control models are the following:

1) *Finite Horizon Average Cost Problem*: Here the goal is to find policies that minimize the average cost

$$J_{\pi_0}(\Pi, T) := E_{\pi_0}^{\Pi}\left[\frac{1}{T}\sum_{t=0}^{T-1} c(Z_t, A_t)\right], \quad (27)$$

for some $T \geq 1$.

2) *Infinite Horizon Discounted Cost Problem*: Here the goal is to find policies that minimize

$$J_{\pi_0}^{\beta}(\Pi) := \lim_{T \to \infty} E_{\pi_0}^{\Pi}\left[\sum_{t=0}^{T-1} \beta^t c(Z_t, A_t)\right], \quad (28)$$

for some $\beta \in (0, 1)$.

3) *Infinite Horizon Average Cost Problem*: In the more challenging infinite horizon control problem the goal is to find policies that minimize the average cost

$$J_{\pi_0}(\Pi) := \limsup_{T \to \infty} E_{\pi_0}^{\Pi}\left[\frac{1}{T}\sum_{t=0}^{T-1} c(Z_t, A_t)\right]. \quad (29)$$

The Markov control model together with the performance objective is called a *Markov decision process*.

A common method to solving finite horizon Markov control problems is by *dynamic programming*, which involves working backwards from the final time stage to solve for the optimal sequence of controls to use. The optimality of this algorithm is guaranteed by Bellman's principle of optimality.

Theorem 8 (Bellman's Principle of Optimality [40, Chapter 3.2]): Given a finite time horizon $T \geq 1$, define a sequence of functions $J_T, \ldots, J_0$ on Z recursively such that

$$J_T(z_T) \equiv 0,$$

and for $0 \leq t < T$ and $z \in \mathsf{Z}$,

$$J_t(z) := \min_{a \in \mathsf{A}}\left[c(z, a) + \int_{\mathsf{Z}} J_{t+1}(z')K(dz'|z, a)\right]. \quad (30)$$

If the J_t are measurable and there exist measurable $f_t : \mathsf{Z} \to \mathsf{A}$ such that $a = f_t(z)$ achieves the above minimum for all $t = 0, \ldots, T-1$, then the deterministic Markov policy $\Pi := (f_0, \ldots, f_{T-1})$ is optimal with cost $J_{z_0}(\Pi, T) = J_0(z_0)$.

Quite general conditions exist under witch the two assumptions of the above theorem hold [40, Chapter 3.3].

For the infinite horizon discounted cost Markov control problem, one can also use an iteration algorithm to obtain an optimal policy. This approach is commonly called the *successive approximations* or value iteration method [40, Chapter 4.2].

A stochastic kernel K on Z given $\mathsf{Z} \times \mathsf{A}$ is called weakly continuous if the function

$$(a, z) \mapsto \int_{\mathsf{Z}} \upsilon(z')K(dz'|z, a)$$

is continuous whenever υ is a *bounded and continuous* real function on $\mathsf{Z} \times \mathsf{A}$. It is called strongly continuous if the $(a, z) \mapsto \int_{\mathsf{Z}} \upsilon(z')K(dz'|z, a)$ is continuous whenever υ is a *measurable and bounded* real function on $\mathsf{Z} \times \mathsf{A}$. The next theorem follows from [49, Chapter 8.5].

Theorem 9: Suppose the following conditions hold:

(i) *The one stage cost c is continuous, nonnegative, and bounded;*

(ii) A *is compact;*

(iii) *the transition kernel K is weakly continuous.*

Then for any $\beta \in (0, 1)$, the pointwise limit $J(z)$ as $t \to \infty$, of the sequence defined by

$$J_t(z) = \min_{a \in \mathsf{A}}\left[c(x, a) + \beta \int_{\mathsf{Z}} J_{t-1}(z')K(z'|z, a)\right], \quad z \in \mathsf{Z},$$

with $J_0(z) \equiv 0$, yields the optimum cost in the infinite horizon discounted cost problem (i.e., $\inf_{\Pi \in \Pi_A} J_z^{\beta} = J(z)$). Furthermore, there exists a measurable function $f : \mathsf{Z} \to \mathsf{A}$ such that

$$J(z) = c(x, f(z)) + \beta \int_{\mathsf{Z}} J_{t-1}(z')K(z'|z, f(z))$$

and the policy $\Pi = \{f\}$ is an optimal stationary Markov policy.

Finally, for the infinite horizon average cost Markov control problem, we give a brief overview of the *average cost optimality equation* (ACOE). When the ACOE holds for a deterministic and stationary Markov policy Π, we know Π is optimal for the infinite horizon average cost problem.

Definition 5: Let h and g be measurable real functions on Z and let $f : \mathsf{Z} \to \mathsf{A}$ be measurable. Then (g, h, f) is said to be a canonical triplet if for all $z \in \mathsf{Z}$,

$$g(z) = \inf_{a \in \mathsf{A}} \int_{\mathsf{Z}} g(z')K(dz'|z, a) \quad (31)$$

$$g(z) + h(z) = \inf_{a \in \mathsf{A}}\left(c(z, a) + \int_{\mathsf{Z}} h(z')K(dz'|z, a)\right) \quad (32)$$

and

$$g(z) = \int_{\mathsf{Z}} g(z')K(dz'|z, f(z)) \quad (33)$$

$$g(z) + h(z) = c(z, f(z)) + \int_{\mathsf{Z}} h(z')K(dz'|z, f(z)). \quad (34)$$

Equations (31)–(34) are called the *canonical equations*. In case g is a constant, $g \equiv g^* \in [0, \infty)$, these equations reduce to

$$g^* + h(z) = \inf_{a \in \mathsf{A}}\left(c(z, a) + \int_{\mathsf{Z}} h(z')K(dz'|z, a)\right) \quad (35)$$

$$g^* + h(z) = c(z, f(z)) + \int_{\mathsf{Z}} h(z')K(dz'|z, f(z)) \quad (36)$$

and (35) is called the *average cost optimality equation* (ACOE).

The ACOE is of central importance in the theory of infinite horizon average cost problems since (as can be shown [40, Chapter 5.2]), with the additional condition that $\limsup_{T\to\infty}(1/T)E_{z_0}^{\Pi}[h(Z_T)] = 0$ for all $z_0 \in \mathsf{Z}$ and $\Pi \in \Pi_A$, it implies that the deterministic and stationary Markov policy $\Pi^* = \{f\}$ is optimal in Π_A and g^* is the value function, i.e.,

$$g^* = J_{z_0}(\Pi^*) = \inf_{\Pi \in \Pi_A} J_{z_0}(\Pi).$$

Although several general sufficient conditions for the ACOE to hold exist (see, e.g., Assumptions 4.2.1 and 5.5.1 in [40]), these conditions are restrictive in our setup since they involve the strong continuity of the transition kernel. In our results we take Z to be the space probability measures, which makes strong continuity too strict a condition in general [50], [51]. More relaxed conditions that involve weak continuity are available in the literature, see [52], [53]. Since for us it is enough to consider compact state and action spaces and uniformly bounded cost, the following theorem will suffice. Recall that

$$J_z^{\beta} = \inf_{\Pi \in \Pi_A} J_z^{\beta}(\Pi).$$

Theorem 10 ([50, Theorem 3.3]): Suppose conditions (i)–(iii) of Theorem 9 hold and, in addition,

(iv) the state space Z is compact;

(v) the family of functions $\{h_\beta : \beta \in (0, 1)\}$, with

$$h_\beta(z) = J_z^{\beta} - J_{z_0}^{\beta}$$

for some fixed $z_0 \in \mathsf{Z}$, is uniformly bounded and equicontinuous.

Then there exist a constant $g^* \geq 0$, a continuous and bounded function $h : \mathsf{Z} \to \mathbb{R}$, and a measurable function $f^* : \mathsf{Z} \to \mathsf{A}$ such that (g^*, h, f^*) is a canonical triplet that satisfies the ACOE. Thus the deterministic and stationary Markov policy $\Pi^* = \{f^*\}$ is optimal in Π_A and g^* is the value function, i.e.,

$$g^* = J_{z_0}(\Pi^*) = \inf_{\Pi \in \Pi_A} \limsup_{T \to \infty} E_{z_0}^{\Pi}\left[\frac{1}{T}\sum_{t=0}^{T-1} c(Z_t, A_t)\right],$$

for all $z_0 \in \mathsf{Z}$.

Recall the definition

$$J_{\pi_0}(\Pi, T) := E_{\pi_0}^{\Pi}\left[\frac{1}{T}\sum_{t=0}^{T-1} c(Z_t, A_t)\right].$$

For the canonical triplet (ρ^*, h, f^*) in the preceding theorem, [40, p. 80] shows that for all $z_0 \in \mathsf{Z}$ and $T \geq 1$,

$$J_{z_0}(\Pi, T) = g^* + \frac{1}{T}\left(h(z_0) - E_{z_0}^{\Pi^*} h(Z_T)\right). \qquad (37)$$

Also, the function h in Theorem 10 is the pointwise limit of the sequence $\{h_{\beta_n}(z)\}$ along some sequence of discount factors $\{\beta_n\}$ such that $\lim_{n\to\infty} \beta_n = 1$. Thus if $\{h_{\beta_n}(z)\}$ is uniformly bounded, say $|h_\beta(z)| \leq K/2$ for all $z \in \mathsf{Z}$ and $\beta \in (0, 1)$, then $|h(z)| \leq K/2$ for all z, and so (37) implies

$$J_{z_0}(\Pi^*, T) - g^* = J_{z_0}(\Pi^*, T) - J_{z_0}(\Pi^*) \leq \frac{K}{T} \qquad (38)$$

for all $T \geq 1$.

For further details on controlled Markov processes, see [40].

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.

[3] T. Berger, *Rate Distortion Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1971.

[4] S. Yüksel and T. Başar, *Stochastic Networked Control Systems: Stabilization and Optimization Under Information Constraints*. New York, NY, USA: Birkhäuser, 2013.

[5] S. C. Draper, C. Chang, and A. Sahai, "Lossless coding for distributed streaming sources," *IEEE Trans. Inf. Theory*, vol. 60, no. 3, pp. 1447–1474, Mar. 2014.

[6] F. Etezadi, A. Khisti, and M. Trott, "Zero-delay sequential transmission of Markov sources over burst erasure channels," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4584–4613, Aug. 2014.

[7] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Comput. Netw.*, vol. 38, pp. 393–422, 2002.

[8] T. Linder and S. Yüksel, "On optimal zero-delay quantization of vector Markov sources," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 5975–5991, Oct. 2014.

[9] T. Ericson, "A result on delayless information transmission," in *Proc. IEEE Int. Symp. Inf. Theory*, Grignano, Italy, 1979.

[10] N. T. Gaarder and D. Slepian, "On optimal finite-state digital transmission systems," in *Proc. IEEE Int. Symp. Inf. Theory*, Grignano, Italy, 1979.

[11] N. Gaarder and D. Slepian, "On optimal finite-state digital transmission systems," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 167–186, Mar. 1982.

[12] G. Gábor and Z. Györfi, *Recursive Source Coding*. New York, NY, USA: Springer-Verlag, 1986.

[13] S. Yüksel, "On optimal causal coding of partially observed Markov sources in single and multiterminal settings," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 424–437, Jan. 2013.

[14] H. S. Witsenhausen, "On the structure of real-time source coders," *Bell Syst. Tech. J.*, vol. 58, no. 6, pp. 1437–1451, Jul. 1979.

[15] J. C. Walrand and P. Varaiya, "Optimal causal coding—Decoding problems," *IEEE Trans. Inf. Theory*, vol. 29, no. 6, pp. 814–820, Nov. 1983.

[16] D. Teneketzis, "On the structure of optimal real-time encoders and decoders in noisy communication," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4017–4035, Sep. 2006.

[17] D. Neuhoff and R. Gilbert, "Causal source codes," *IEEE Trans. Inf. Theory*, vol. 28, no. 5, pp. 701–713, Sep. 1982.

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

[19] S. Tatikonda, "Control under communication constraints," Ph.D. dissertation, Dept. Elect. Comput. Eng., Massachusetts Inst. Technol., Cambridge, MA, USA, 2000.

[20] S. Tatikonda, A. Sahai, and S. Mitter, "Stochastic linear control over a communication channel," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1549–1561, Sep. 2004.

[21] R. Bansal and T. Başar, "Simultaneous design of measurement and control strategies for stochastic systems with feedback," *Automatica*, vol. 25, no. 5, pp. 679–694, Sep. 1989.

[22] C. D. Charalambous, P. A. Stavrou, and N. U. Ahmed, "Nonanticipative rate distortion function and relations to filtering theory," *IEEE Trans. Autom. Control*, vol. 59, no. 4, pp. 937–952, Apr. 2014.

[23] C. K. Kourtellaris, C. D. Charalambous, and J. J. Boutros, "Nonanticipative transmission for sources and channels with memory," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2015, pp. 521–525.

[24] A. Mahajan and D. Teneketzis, "Optimal design of sequential real-time communication systems," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5317–5338, Nov. 2009.

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

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

[27] H. Asnani and T. Weissman, "Real-time coding with limited lookahead," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3582–3606, Jun. 2013.

[28] T. Javidi and A. Goldsmith, "Dynamic joint source-channel coding with feedback," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 16–20.

[29] R. Pilc, "Coding theorems for discrete source-channel pairs," Ph.D. dissertation, Dept. Elect. Eng., Massachusetts Inst. Technol., Cambridge, MA, USA, 1967.

[30] Z. Zhang, E.-H. Yang, and V. K. Wei, "The redundancy of source coding with a fidelity criterion. 1. Known statistics," *IEEE Trans. Inf. Theory*, vol. 43, no. 1, pp. 71–91, Jan. 1997.

[31] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3309–3338, Jun. 2012.

[32] V. S. Borkar, "Convex analytic methods in Markov decision processes," in *Handbook of Markov Decision Processes*. Boston, MA, USA: Kluwer, 2002.

[33] G. Como, S. Yüksel, and S. Tatikonda, "The error exponent of variable-length codes over Markov channels with feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2139–2160, May 2009.

[34] H. H. Permuter, P. Cuff, B. Van Roy, and T. Weissman, "Capacity of the trapdoor channel with feedback," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3150–3165, Jul. 2008.

[35] H. H. Permuter, T. Weissman, and A. J. Goldsmith, "Finite state channels with time-invariant deterministic feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 644–662, Feb. 2009.

[36] S. Tatikonda and S. Mitter, "The capacity of channels with feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 323–349, Jan. 2009.

[37] O. Sabag, H. H. Permuter, and N. Kashyap, "The feedback capacity of the binary erasure channel with a no-consecutive-ones input constraint," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 8–22, Jan. 2016.

[38] O. Sabag, H. H. Permuter, and N. Kashyap. (2017). "Feedback capacity and coding for the BIBO channel with a no-repeated-ones input constraint." [Online]. Available: https://arxiv.org/abs/1701.06174

[39] R. M. Gray, *Entropy and Information Theory*, 2nd ed. Springer, 2011.

[40] O. Hernandez-Lerma and J. B. Lasserre, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*. Springer, 1996.

[41] C. Villani, *Optimal Transport: Old and New*. Springer, 2008.

[42] V. S. Borkar, "Average cost dynamic programming equations for controlled Markov chains with partial observations," *SIAM J. Control Optim.*, vol. 39, no. 3, pp. 673–681, 2000.

[43] V. S. Borkar, "Dynamic programming for ergodic control of Markov chains under partial observations: A correction," *SIAM J. Control Optim.*, vol. 45, no. 6, pp. 2299–2304, 2007.

[44] J. R. Norris, *Markov Chains*. Cambridge, U.K.: Cambridge Univ. Press, 1997.

[45] S. Yüksel and S. P. Meyn, "Random-time, state-dependent stochastic drift for Markov chains and application to stochastic stabilization over erasure channels," *IEEE Trans. Autom. Control*, vol. 58, no. 1, pp. 47–59, Jan. 2013.

[46] A. Arapostathis, V. S. Borkar, and M. K. Ghosh, *Ergodic Control of Diffusion Processes*, vol. 143. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[47] R. Gray, "Information rates of autoregressive processes," *IEEE Trans. Inf. Theory*, vol. 16, no. 4, pp. 412–421, Jul. 1970.

[48] T. Berger, "Explicit bounds to R(D) for a binary symmetric Markov source," *IEEE Trans. Inf. Theory*, vol. 23, no. 1, pp. 52–59, Jan. 1977.

[49] O. Hernandez-Lerma and J. B. Lasserre, *Further Topics on Discrete-Time Markov Control Processes*. Springer, 1999.

[50] N. Saldi, S. Yüksel, and T. Linder. (Oct. 2014). "Near optimality of quantized policies in stochastic control under weak continuity conditions." [Online]. Available: https://arxiv.org/abs/1410.6985

[51] E. A. Feinberg, P. O. Kasyanov, and M. Z. Zgurovsky. (2014). "Partially observable total-cost Markov decision processes with weakly continuous transition probabilities." [Online]. Available: https://arxiv.org/abs/1401.2168

[52] E. Gordienko and O. Hernández-Lerma, "Average cost Markov control processes with weighted norms: Existence of canonical policies," *Appl. Math.*, vol. 23, no. 2, pp. 199–218, 1995.

[53] O. Vega-Amaya, "The average cost optimality equation: A fixed point approach," *Bol. Soc. Mat. Mexicana*, vol. 9, no. 3, pp. 185–195, 2003.

Richard G. Wood received his BScE and MASc degrees in Mathematics and Engineering from Queen's University, Kingston, ON, Canada, in 2012 and 2015 respectively. He is the President & Co-Founder of the Bobcaygeon Brewing Company

Tamás Linder (S'92–M'93–SM'00–F'13) received the M.S. degree in electrical engineering from the Technical University of Budapest, Hungary, in 1988, and the Ph.D degree in electrical engineering from the Hungarian Academy of Sciences in 1992. He was a post-doctoral researcher at the University of Hawaii in 1992 and a Visiting Fulbright Scholar at the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign during 1993-1994. From 1994 to 1998 he was a faculty member in the Department of Computer Science and Information Theory at the Technical University of Budapest. From 1996 to 1998 he was also a visiting research scholar in the Department of Electrical and Computer Engineering, University of California, San Diego. In 1998 he joined Queen's University where he is now a Professor of Mathematics and Engineering in the Department of Mathematics and Statistics. His research interests include communications and information theory, source coding and vector quantization, machine learning, and statistical pattern recognition. Dr. Linder received the Premier's Research Excellence Award of the Province of Ontario in 2002 and the Chancellor's Research Award of Queen's University in 2003. He was an Associate Editor for Source Coding of the IEEE TRANSACTIONS ON INFORMATION THEORY in 2003-2004.

Serdar Yüksel (M'11) received his B.Sc. degree in Electrical and Electronics Engineering from Bilkent University in 2001; M.S. and Ph.D. degrees in Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign in 2003 and 2006, respectively. He was a post-doctoral researcher at Yale University for a year before joining Queen's University as an Assistant Professor of Mathematics and Engineering in the Department of Mathematics and Statistics, where he is now an Associate Professor. He has also held a visiting position at Bilkent University, Turkey. He has been awarded the 2013 CAIMS/PIMS Early Career Award in Applied Mathematics. His research interests are on stochastic control theory, networked control, and information theory. He currently serves as an associate editor for IEEE TRANSACTIONS ON AUTOMATIC CONTROL.