A New Finite-Horizon Dynamic Programming Analysis of Nonanticipative Rate-Distortion Function for Markov Sources

Zixuan He[1], Charalambos D. Charalambous[2], and Photios A. Stavrou[1]

***Abstract*— This paper addresses the computation of a non-asymptotic lower bound, given by the nonanticipative rate-distortion function (NRDF), for the discrete-time zero-delay variable-rate lossy compression of discrete Markov sources under per-stage single-letter distortion constraints. We first derive a new information structure for the NRDF and new convexity results that allow reformulating the problem as an unconstrained partially observable finite-horizon stochastic dynamic program (DP) using Lagrange duality theorem subject to a belief state that summarizes past information and evolves in a continuous space. Rather than directly approximating the DP, we derive implicit optimal conditions via the Karush-Kuhn-Tucker (KKT) conditions and propose a novel alternating minimization (AM) scheme to approximate both the control policy and cost-to-go function through backward recursions with provable convergence guarantees. We evaluate the control policies and cost-to-go functions per-stage using an online forward algorithm that executes for any finite horizon. Our methodology yields a near-optimal approximation of the NRDF as the belief state space becomes sufficiently large. Simulation results using time-varying binary Markov sources validate the effectiveness of our approach.**

I. Introduction

The classical lossy source coding problem involves encoding long blocks of source symbols to asymptotically achieve Shannon's limit, minimizing the average bit rate for a given distortion level [2]. However, such block-based schemes introduce significant coding delays, making them undesirable for delay-sensitive applications spanning from networked control systems [3] to applications within the emerging field of semantic communications [1].

A more appropriate lossy compression paradigm for delay-sensitive applications is the zero-delay lossy source coding. Unlike classical schemes, it generates compressed source symbols by the encoder without delay and transmits over a discrete noiseless channel to the decoder, which immediately reconstructs the source symbols under a fidelity constraint. The discrete noiseless channel may operate assuming either fixed or variable rates.

Several key results are documented in the literature on zero-delay lossy compression schemes. Early foundational works [4], [5] explored the structural properties of optimal zero-delay codes for primarily fixed-rate systems using stochastic control and DP [6], with later generalizations in [7]–[9]. In [10], they developed structural theorems for variable-rate coding with and without side information at the decoder. Recently, [11] invoked reinforcement learning via a quantized Q-learning to near-optimally approximate the infinite-horizon fixed-rate zero-delay coding problem for finite-alphabet stationary Markov sources. Another relevant research direction, instead of designing directly zero-delay lossy codes subject to variable-rate constraints, derives information-theoretic upper and lower bounds building on nonanticipatory $\epsilon-$entropy[1] introduced by [14]. This line of research primarily studied bounding techniques on linear (perhaps controlled) Markov systems driven by additive Gaussian or non-Gaussian noise and the derivation of information-theoretic closed-form solutions [12], [15]–[17].

Contributions: In this paper, we analyze a non-asymptotic lower bound on the empirical rates of a discrete-time zero-delay variable-rate lossy source coding system assuming discrete and possibly time-varying Markov sources subject to per-stage average single-letter distortion criterion. We first derive a structural property of the NRDF specific to the particular class of sources and fidelity constraint (see Lemma 1). We then derive new convexity properties results (see Theorem 2, 3) that enable reformulating the problem as an unconstrained partially observable finite-horizon stochastic DP with an continuous information state space [6] (see equations (13), (14)). Instead of solving the DP recursions directly, we optimize with respect to (wrt) the control policy (which corresponds to the minimizing distribution of the NRDF) and thereby derive implicit closed-form backward recursions. These are computed via a new dynamic AM scheme (see Lemma 4) that approximates offline the control policy and the cost-to-go function (a function of the rate) using a discretized belief state space at each time stage (see Algorithm 1). A forward (online) algorithm (see Algorithm 2) is then proposed to evaluate the resulting functional quantities for any finite-time horizon. Our offline scheme has provable convergence guarantees per-stage (see Theorems 5, 6) and approaches a near-optimal solution as the discretized belief space (finite state-space) becomes sufficiently large. We corroborate our theoretical results with numerical simulations on time-varying binary Markov sources with the illustration of the stagewise rate.

*The work of Z. He was supported by the Huawei France-EURECOM Chair on Future Wireless Networks. The work of P. A. Stavrou was supported in part by the SNS JU project 6G-GOALS [1] under the EU's Horizon programme (Grant Agreement No. 101139232) and by the Huawei France-EURECOM Chair on Future Wireless Networks.

[1]Z. He and P. A. Stavrou are with the Foundation and Algorithm Group, Communication Systems Department, EURECOM, France. email: {zixuan.he,fotios.stavrou}@eurecom.fr

[2]C. D. Charalambous is with the Department of Electrical and Computer Engineering, University of Cyprus, Cyprus. email: chadcha@ucy.ac.cy

[1]Also found in the literature as sequential rate-distortion-function (RDF) [12] and nonanticipative RDF (NRDF) [13].

To the best of our knowledge, this is the first work to (i) reformulate the NRDF optimization for discrete Markov sources and single-letter distortion as an unconstrained partially observable stochastic DP with continuous state space, and (ii) approximate the optimal policy via a dynamic AM scheme that generalizes the Blahut-Arimoto algorithm (BAA) [18] into offline training, followed by an online computation.

Notation: $\mathbb{N} \triangleq \{1,2,\ldots\}$, $\mathbb{N}_0 \triangleq \{0,1,\ldots\}$, and $\mathbb{N}_j^n \triangleq \{j,\ldots,n\}$, $j \leq n$, $n \in \mathbb{N}$. We denote a sequence of random variables (RVs) by $X^t = \{X_0, X_1, \ldots, X_t\}$, $t \in \mathbb{N}_0^n$ and their values by $x^t \in \mathcal{X}^t = \{\mathcal{X}_0, \ldots, \mathcal{X}_t\}$, where $\mathcal{X}_t$ denotes the alphabet and hence $\mathcal{X}^t$ the alphabet sequence. A truncated sequence of RVs is denoted by $X_j^t = \{X_j, \ldots, X_t\}$, $j \in \mathbb{N}_0^t$, $t \geq j$, and its realizations by $x_j^t = \{x_j, \ldots, x_t\} \in \mathcal{X}_j^t = \{\mathcal{X}_j, \ldots, \mathcal{X}_t\}$, $t \geq j$. The distribution of a RV X on $\mathcal{X}$ is denoted by $P(x)$, and the conditional distribution of Y given $X = x$ denoted by $P(y|x)$. Functional dependencies are indicated using square brackets; e.g., $P[Q](x)$ and $P[y](x)$ express the functional dependence of a distribution $P(x)$ on another distribution Q and on another realization y, respectively. Expectation operator is denoted by $\mathbb{E}\{\cdot\}$, and by $\mathbb{E}^{P^o}\{\cdot\}$ when taken wrt a specific distribution $P(\cdot) = P^o(\cdot)$.

II. Problem Statement and Preliminaries

We consider the discrete-time zero-delay lossy source coding system illustrated in Fig. 1, operating at any finite-time horizon $n \in \mathbb{N}_0$.

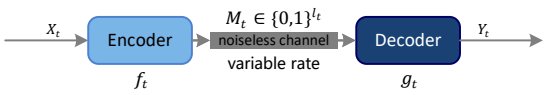


Fig. 1: A generic zero-delay lossy source coding system

Operation: At each time instant $t \in \mathbb{N}_1^n$, the *source* is modeled as a Markov process (not necessarily time-homogeneous) with transition probability distribution $P_t(x_t|x_{t-1})$ and initial distribution $P_0(x_0)$, which induce the joint distribution of the sequence of RVs X^t, $t \in \mathbb{N}_0^n$. For any $X_t \in \mathcal{X}_t$, we assume that the cardinality of $\mathcal{X}_t$ is finite. The *encoder (E)* f_t encodes the information X_t generated from the source based on the past information X^{t-1} and produces a variable-rate codeword $M_t \in \mathcal{M}_t \subset \{0,1\}^{l_t}$ of length l_t and expected rate $R_t = \mathbb{E}|l_t|$. The *decoder (D)* g_t receives the past and current codewords M^t to reproduce $Y_t \in \mathcal{Y}_t$, provided that Y^{t-1} is already reproduced. Again, we assume that for any $Y_t \in \mathcal{Y}_t$, the cardinality of $\mathcal{Y}_t$ is finite. Formally, the *(E), (D)* pair is specified by the sequence of possibly stochastic mappings $f_t : \mathcal{M}^{t-1} \times \mathcal{X}^t \to \mathcal{M}_t$, and $g_t : \mathcal{M}^t \to \mathcal{Y}_t$, respectively. We note that at $t = 0$, the encoder's output is $m_0 = f_0(x_0)$ and the decoder's output $y_0 = g_0(m_0)$. This means that no prior information is assumed at both the encoder and the decoder.

Fidelity constraint: The system in Fig. 1, is penalized at each time instant by an additive fidelity $\{D_t \geq 0 : t \in \mathbb{N}_0^n\}$. The constraint between the source process $\{X_t : t \in \mathbb{N}_0^n\}$ and the reproduction process $\{Y_t : t \in \mathbb{N}_0^n\}$ is described by the per-stage average single-letter distortion, i.e., $\mathbb{E}\{\rho_t(X_t, Y_t)\} \leq D_t$, where $\rho_t(x_t, y_t)$ is the single-letter distortion function between the source information x_t and its reproduction y_t at each t.

Performance: The system model in Fig. 1, subject to the fidelity criterion described above, can be computed by the following empirical rate optimization:

$$R_{[0,n]}^o(D_0,\ldots,D_n) \triangleq \inf_{\substack{\{(f_t,\ g_t):\\ \mathbb{E}\{\rho_t(X_t,Y_t)\}\leq D_t,\\ \forall t\in\mathbb{N}_0^n\}}} \frac{1}{n+1}\sum_{t=0}^{n} R_t. \quad (1)$$

Solving (1) requires considering all possible coding schemes, making it extremely difficult to obtain a globally optimal solution with tractable computational complexity. As a result, it makes sense to pursue approximate solutions via bounds.

A well-known lower bound on (1), assuming a Markov source and a per-stage average single-letter distortion, is the NRDF [14] given as follows:

$$R_{[0,n]}^{na}(D_0,\ldots,D_n) \triangleq \inf_{\substack{\mathcal{Q}_{[0,n]}\\ (D_0,\ldots,D_n)}} \frac{1}{n+1} I(X^n \to Y^n) \quad (2)$$

where the constraint set is expressed as

$$\mathcal{Q}_{[0,n]}(D_0,\ldots,D_n) \triangleq \{P_t(y_t|y^{t-1},x_t) : \mathbb{E}\{\rho_t(X_t,Y_t)\} \leq D_t,\ \forall t \in \mathbb{N}_0^n\} \quad (3)$$

and $I(X^n \to Y^n)$ is a variant of directed information (DI) [19] defined as follows

$$\begin{aligned} &I(X^n \to Y^n) \\ &\triangleq \sum_{t=0}^{n} \mathbb{E}\left\{\log\left(\frac{P_t(Y_t|Y^{t-1},X_t)}{P_t(Y_t|Y^{t-1})}\right)\right\} = \sum_{t=0}^{n} I(X_t;Y_t|Y^{t-1}) \\ &= \sum_{t=0}^{n}\int_{\mathcal{X}_{t-1}^t\times\mathcal{Y}^t} \log\left(\frac{P_t(y_t|y^{t-1},x_t)}{P_t(y_t|y^{t-1})}\right) P_t(x_t|x_{t-1}) \\ &\quad P_t(y_t|y^{t-1},x_t)P_t(x_{t-1}|y^{t-1})P_t(y^{t-1}) \\ &\overset{(a)}{\equiv} \mathbb{I}_{[0,n]}(\overleftarrow{P}_{[0,n]}(x^n), \overrightarrow{P}_{[0,n]}(y^n|x^n)) \end{aligned} \quad (4)$$

where (a) denotes the functional dependence of $I(X^n \to Y^n)$ wrt $\overleftarrow{P}_{[0,n]}(x^n) \triangleq \otimes_{t=0}^n P_t(x_t|x_{t-1})$ and $\overrightarrow{P}_{[0,n]}(y^n|x^n) \triangleq \otimes_{t=0}^n P_t(y_t|y^{t-1},x_t)$, respectively.

We state some noteworthy remarks related to (2).

Remark 1: (On the bound of (2)) **(i)** In accordance with the system model in Fig. 1, (2) does not assume any prior information at $t = 0$ available in the system model but only $P_0(x_0)$ and $P_0(y_0)$ are fixed. This means that at $t = 0$, $I(X_0;Y_0|Y^{-1}) \equiv I(X_0;Y_0)$. **(ii)** The bound of (2) has been thoroughly studied for continuous sources, see e.g., [20] but not adequately for discrete sources, with a notable exception perhaps the work of [21] which alas does not provide general methodologies for computation or tangible proofs to certain analytical expressions. **(iii)** Although (2) under the constraint set (3) makes sense to be a convex program, there is no available proof in the literature. On the other hand, (2) forms a convex program when $I(X^n \to Y^n)$ is a convex function

wrt the product $\overrightarrow{P}_{[0,n]}$ for a fixed $\overleftarrow{P}_{[0,n]}(x^n)$ and for a constraint set defined as

$$\overrightarrow{\mathcal{Q}}_{[0,n]}(D_0,\ldots,D_n) \triangleq \{\overrightarrow{P}(y^n|x^n) \triangleq \otimes_{t=0}^{n} P_t(y_t|y^{t-1},x_t) : \mathbb{E}\{\rho_t(X_t,Y_t)\} \leq D_t,\ \forall t \in \mathbb{N}_0^n\} \quad (5)$$

(see e.g., [22]). Hence, *no proof of the convexity of* (2) *exists wrt the constraint set* (3). **(iv)** There is no generic implicit or explicit solution of (2) under the constraint set of (3). However, implicit closed-form recursions of the optimal minimizer are reported (without a complete proof) in [20, Theorem 4.1] assuming the constraint set in (5).

In this work, we provide a generic methodology to approximate (2) assuming discrete Markov sources aiming primarily to close the gaps mentioned in Remark 1, **(iii)**, **(iv)**.

III. Main Results

In this section, we give our main results. To do it, we restrict ourselves to finite alphabet spaces, e.g., with cardinality $|\mathcal{X}_t| < \infty$, $|\mathcal{Y}_t| < \infty$, $\forall t$, throughout the paper.

First, we give a new information structure that simplifies the multi-letter variant of DI in (2).

Lemma 1: (Structural Property) For a given Markov source $\{P_t(x_t|x_{t-1}) :\ t \in \mathbb{N}_0^n\}$ and a single letter distortion function $\{\rho_t(x_t,y_t) :\ t \in \mathbb{N}_0^n\}$, the characterization in (2) can be simplified as follows[2]

$$R_{[0,n]}^{na}(D_0,\ldots,D_n) \triangleq \min_{\tilde{\mathcal{Q}}_{[0,n]}(D_0,\ldots,D_n)} \frac{1}{n+1} I(X^n \to Y^n), \quad (6)$$

where

$$\tilde{\mathcal{Q}}_{[0,n]}(D_0,\ldots,D_n) \triangleq \Big\{P_t(y_t|y_{t-1},x_t) : \mathbb{E}\{\rho_t(X_t,Y_t)\} \leq D_t,\ \forall t \in \mathbb{N}_0^n\Big\} \quad (7)$$

$$I(X^n \to Y^n) \triangleq \sum_{t=0}^{n} I(X_t;Y_t|Y_{t-1}). \quad (8)$$

The result in Lemma 1 is generic and holds for both discrete and continuous alphabets. Indeed, such property is already verified for joint Gaussian-Markov processes, e.g., [20], [23].

The next two results provide conditions to ensure new convexity properties for (6). The first result demonstrates a new convexity property of (8) for a given posterior distribution $\{P_t(x_{t-1}|y_{t-1}) \equiv P_t^o(x_{t-1}|y_{t-1}) : t \in \mathbb{N}_0^n\}$, wrt the minimizing distributions $\{P_t(y_t|y_{t-1},x_t) : t \in \mathbb{N}_0^n\}$.

Theorem 2: (Convexity of (8)) For a fixed source distribution $\{P_t(x_t|x_{t-1}) : t \in \mathbb{N}_0^n\}$, and a given posterior distribution $\{P_t^o(x_{t-1}|y_{t-1}) : t \in \mathbb{N}_0^n\}$ obtained for a fixed $Y_{t-1} = y_{t-1}$, define the conditional mutual information $I(X_t;Y_t|Y_{t-1}=y_{t-1})$ as follows

$$I(X_t;Y_t|Y_{t-1}=y_{t-1}) \triangleq \sum_{\substack{x_{t-1}\\ \in\mathcal{X}_{t-1}}} \sum_{\substack{x_t\in\mathcal{X}_t\\ y_t\in\mathcal{Y}_t}} \log\left(\frac{P_t(y_t|y_{t-1},x_t)}{P_t(y_t|y_{t-1})}\right) P_t(x_t|x_{t-1})P_t(y_t|y_{t-1},x_t)P_t^o(x_{t-1}|y_{t-1}),\ \forall t \in \mathbb{N}_0^n. \quad (9)$$

[2]For non-empty finite sets, we can replace infimum with minimum due to the compactness of the constraint sets.

TABLE I: Elements of stochastic optimal control problem

Variables of DP	Connection to (12)
belief	$P_t(x_{t-1}\|y_{t-1})$
disturbance	$P_t(x_t\|x_{t-1})$
control policy	$P_t(y_t\|y_{t-1},x_t)$
cost function	$\log\Big(\frac{P_t(y_t\|y_{t-1},x_t)}{P_t(y_t\|y_{t-1})}\Big) - s_t(\rho_t(x_t,y_t) - D_t[y_{t-1},P(x_{t-1}\|y_{t-1})])$

Then, (9) is a convex functional wrt $\{P_t(y_t|y_{t-1},x_t) :\ t \in \mathbb{N}_0^n\}$. Moreover, the additive term

$$I(X^n \to Y^n) = \sum_{t=0}^{n} \sum_{\substack{y_{t-1}\\ \in\mathcal{Y}_{t-1}}} P_t(y_{t-1}) I(X_t;Y_t|Y_{t-1}=y_{t-1}) \quad (10)$$

is also convex wrt $\{P_t(y_t|y_{t-1},x_t) :\ t \in \mathbb{N}_0^n\}$.

The following result establishes the convexity of the constraint set in (7) for a given $\{P_t^o(x_{t-1}|y_{t-1}) :\ t \in \mathbb{N}_0^n\}$.

Theorem 3: (Convexity of (7)) For a fixed source distribution $\{P_t(x_t|x_{t-1}) :\ t \in \mathbb{N}_0^n\}$, and a given posterior distribution $\{P_t^o(x_{t-1}|y_{t-1}) :\ t \in \mathbb{N}_0^n\}$ obtained for a fixed $Y_{t-1} = y_{t-1}$, define the constraint set

$$\tilde{Q}_{[0,n]}(D_0, D_1[y_0,P_1^o],\ldots,D_n[y_{n-1},P_n^o]) \triangleq \Big\{P_t(y_t|y_{t-1},x_t), t \in \mathbb{N}_0^n : \mathbb{E}\Big\{\rho_0(X_0,Y_0)\Big\} \leq D_0; \mathbb{E}^{P_t^o}\Big\{\rho_t(X_t,Y_t)\Big|Y_{t-1}=y_{t-1}\Big\} \leq D_t[y_{t-1},P_t^o], t \in \mathbb{N}_1^n\Big\}, \quad (11)$$

where $D_t[y_{t-1},P_t^o]$ denotes the functional dependence of D_t wrt the given $\{P_t^o(x_{t-1}|y_{t-1}) :\ t \in \mathbb{N}_0^n\}$ obtained for a fixed $Y_{t-1} = y_{t-1}$. Then, (11) forms a convex set. If averaging the constraints in (11) wrt $y_{t-1} \in \mathcal{Y}_{t-1}$ for each $t \in \mathbb{N}_1^n$, (7) is also convex for a given $\{P_t^o(x_{t-1}|y_{t-1}),\ t \in \mathbb{N}_0^n\}$.

Using Theorems 2, 3, we can exploit Lagrange duality theorem [24] with Lagrange multipliers $\{s_t \leq 0,\ t \in \mathbb{N}_0^n\}$, to cast (6) for a given $\{P_t^o(x_{t-1}|y_{t-1}),\ t \in \mathbb{N}_0^n\}$ obtained for a fixed $Y_{t-1} = y_{t-1}$, into the following unconstrained convex optimization problem:

$$R_{[0,n]}^{un}(D_0, D_1[y_0,P_0^o],\ldots,D_n[y_{n-1},P_n^o]) = \sup_{\{s_t\leq 0:\ t\in\mathbb{N}_0^n\}} \min_{\{P_t(y_t|y_{t-1},x_t):\ t\in\mathbb{N}_0^n\}} \sum_{t=0}^{n} \mathbb{E}^{P_t^o}\Bigg\{\log\left(\frac{P_t(Y_t|Y_{t-1},X_t)}{P_t(Y_t|Y_{t-1})}\right) - s_t\big(\rho_t(X_t,Y_t) - D_t[y_{t-1},P_t^o]\big)\Bigg|Y_{t-1}=y_{t-1}\Bigg\}. \quad (12)$$

Note that if we average (12) wrt to $P_t(y_{t-1}) > 0$, then, we obtain (6) for a given $\{P_t^o(x_{t-1}|y_{t-1}), t \in \mathbb{N}_0^n\}$.

Hence we can reformulate (12) using stochastic optimal control arguments as an *unconstrained partially observable finite-time horizon stochastic DP with a continuous state space* [6]. The reformulation with its one-to-one correspondence in (12) is illustrated in Table I. Let $R_t(\cdot)$ denote the optimal expected cost or pay-off in (12) on the future time horizon $\{t, t+1,\ldots,n\}$. Then, for a given belief state $P_t^o(x_{t-1}|y_{t-1})$ obtained for a fixed $Y_{t-1} = y_{t-1}$, it is

described as follows

$$R_t(D_t[y_{t-1},P_t^o]) = \min_{\substack{\{P_i(y_i|y_{i-1},x_i):\\ i\in\{t,t+1,\ldots,n\}\}}} \mathbb{E}^{P_i^o}\left\{\sum_{i=t}^{n}\log\left(\frac{P_i(Y_i|Y_{i-1},X_i)}{P_i(Y_i|Y_{i-1})}\right) - \sum_{i=t}^{n} s_i\Big(\rho_i(X_i,Y_i) - D_i[y_{i-1},P_i^o]\Big)\,\Big|\,Y_{i-1}=y_{i-1}\right\}.$$

Applying the *principle of optimality* [6] yields the following exact finite-time horizon stochastic DP recursions obtained backward in time

$$R_n(D_n[y_{n-1},P_n^o]) = \min_{P_n(y_n|y_{n-1},x_n)} \sum_{x_{n-1}^n\in\mathcal{X}_{n-1}^n, y_n\in\mathcal{Y}_n} \left\{\left(\log\left(\frac{P_n(y_n|y_{n-1},x_n)}{P_n(y_n|y_{n-1})}\right) - s_n\rho_n(x_n,y_n)\right)P_n(x_n|x_{n-1}) P_n(y_n|y_{n-1},x_n)P_n^o(x_{n-1}|y_{n-1})\right\} + s_nD_n[y_{n-1},P_n^o], \tag{13}$$

$$R_t(D_t[y_{t-1},P_t^o]) = \min_{P_t(y_t|y_{t-1},x_t)} \sum_{x_{t-1}^t\in\mathcal{X}_{t-1}^t, y_t\in\mathcal{Y}_t} \left\{\left(\log\left(\frac{P_t(y_t|y_{t-1},x_t)}{P_t(y_t|y_{t-1})}\right) - s_t\rho_t(x_t,y_t) + R_{t+1}(D_{t+1}[y_t,P_{t+1}^o])\right)P_t(y_t|y_{t-1},x_t)P_t^o(x_{t-1}|y_{t-1}) P_t(x_t|x_{t-1})\right\} + s_tD_t[y_{t-1},P_t^o],\ t=n-1,\ldots,0. \tag{14}$$

Note that the given belief state for fixed $Y_{t-1}=y_{t-1}$ in (13), (14) can be identified at each time stage t by the following recursion

$$P_{t+1}^o(x_t|y_t) = \frac{P_t(x_t,y_t)}{P_t(y_t)} = \frac{\sum_{\substack{y_{t-1}\in\mathcal{Y}_{t-1}\\ x_{t-1}\in\mathcal{X}_{t-1}}} P_t(y_t|y_{t-1},x_t)P_t(x_t|x_{t-1})P_t^o(x_{t-1}|y_{t-1})}{\sum_{\substack{y_{t-1}\in\mathcal{Y}_{t-1}\\ x_{t-1}^t\in\mathcal{X}_{t-1}^t}} P_t(y_t|y_{t-1},x_t)P_t(x_t|x_{t-1})P_t^o(x_{t-1}|y_{t-1})}$$

which is Markov, conditional on $Y_t, P_t^o(x_{t-1}|y_{t-1})$. Moreover, at $t=0$, we adopt the assumptions of our system model in Fig. 1 and assume that $P_0(y_0|y_{-1},x_0)=P_0(y_0|x_0)$, and $P_0(y_0|y_{-1})=P_0(y_0)$, hence the initial time stage in (14) can be modified accordingly. To compute (6) for a given $\{P_t^o(x_{t-1}|y_{t-1}):\ t\in\mathbb{N}_1^n\}$, we need to move forward in time from R_0 and obtain the clean value of the cost-to-go functions at each time stage t averaging over $y_{t-1}\in\mathcal{Y}_{t-1}$ and exploring over all possible values of the belief state space generated by $\{P_t^o(x_{t-1}|y_{t-1}):\ t\in\mathbb{N}_1^n\}$.

The partially observable stochastic DP defined in (13), (14) can be solved offline using approximation methods [25], assuming that the belief state per-stage takes values in $(0,1)$. Instead of following that direction, we exploit the convexity of our problem and optimize wrt the control policy (test-channel) via a more efficient policy approximation method. We first discretize the given continuous belief state space into a finite state space denoted by $\mathcal{B}_t,\ \forall t\in\mathbb{N}_0^n$ and then rely on a dynamic *AM scheme* (it was used to compute the classical RDF for discrete sources [18]) to generate an offline training algorithm. The following lemma builds the dynamic AM.

Lemma 4: (Double minimization) For any $t\in\mathbb{N}_0^n$, let $s_t\le 0$ and $D_t>0$, then for a fixed Markov source $P_t(x_t|x_{t-1})$, and a given belief state $P_t^o(x_{t-1}|y_{t-1})\in\mathcal{B}_t$ obtained for a fixed $Y_{t-1}=y_{t-1}$, (13), (14) can be expressed as a double minimum as follows

$$R_t(D_t[y_{t-1},P_t^o]) = \min_{P_t(y_t|y_{t-1},x_t)}\ \min_{P_t(y_t|y_{t-1})} \sum_{x_{t-1}^t\in\mathcal{X}_{t-1}^t, y_t\in\mathcal{Y}_t} \left\{\left(\log\left(\frac{P_t(y_t|y_{t-1},x_t)}{P_t(y_t|y_{t-1})}\right) - s_t\rho_t(x_t,y_t) + R_{t+1}(D_{t+1}[y_t,P_{t+1}^o])\right)P_t(y_t|y_{t-1},x_t)P_t^o(x_{t-1}|y_{t-1}) P_t(x_t|x_{t-1})\right\} + s_tD_t[y_{t-1},P_t^o],\ t=n,n-1,\ldots,0 \tag{15}$$

where $R_{t+1}(D_{t+1}[y_t,P_{t+1}^o])$ is the cost-to-go that is equal to 0 when $t=n$, and $D_t[y_{t-1},P_t^o]$ is a functional of the distortion level expressed as

$$D_t[y_{t-1},P_t^o] = \sum_{\substack{x_t\in\mathcal{X}_t\\ y_t\in\mathcal{Y}_t}} P_t^{\mathrm{Mo}}(x_t|y_{t-1})P_t^*(y_t|y_{t-1},x_t)\rho_t(x_t,y_t) \tag{16}$$

with $P_t^*(y_t|y_{t-1},x_t)$ achieving the minimum and $P_t^{\mathrm{Mo}}(x_t|y_{t-1}) = \sum_{x_{t-1}\in\mathcal{X}_{t-1}} P_t(x_t|x_{t-1})P_t^o(x_{t-1}|y_{t-1})$. Moreover, for a fixed $P_t(y_t|y_{t-1},x_t)$, the right hand side (RHS) of (15) is minimized by

$$P_t(y_t|y_{t-1}) = \sum_{x_t\in\mathcal{X}_t} P_t^{\mathrm{Mo}}(x_t|y_{t-1})P_t(y_t|y_{t-1},x_t), \tag{17}$$

whereas for fixed $P_t(y_t|y_{t-1})$, the RHS of (15) is minimized by

$$P_t(y_t|y_{t-1},x_t) = \frac{P_t(y_t|y_{t-1})A_t[P_{t+1}^o]}{\sum_{y_t\in\mathcal{Y}_t} P_t(y_t|y_{t-1})A_t[P_{t+1}^o]} \tag{18}$$

where $A_t[P_{t+1}^o] = e^{s_t\rho_t(x_t,y_t) - R_{t+1}(D_{t+1}[y_t,P_{t+1}^o])}$.

From Lemma 4, substituting (18) into (16) yields:

$$D_{s_t}[y_{t-1},P_t^o] = \sum_{x_t\in\mathcal{X}_t, y_t\in\mathcal{Y}_t} P_t^{\mathrm{Mo}}(x_t|y_{t-1}) \frac{P_t^*(y_t|y_{t-1})A_t[P_{t+1}^o]}{\sum_{y_t\in\mathcal{Y}_t} P_t^*(y_t|y_{t-1})A_t[P_{t+1}^o]}\rho_t(x_t,y_t). \tag{19}$$

Lemma 4 provides a parametric model that establishes the use of a new dynamic version of BAA [18], enabling an AM scheme between $P_t(y_t|y_{t-1},x_t)$ and $P_t(y_t|y_{t-1})$ at any $t\in\mathbb{N}_0^n$. We now describe the convergence of the offline training algorithm that approximates the optimal control policy.

Theorem 5: (Offline training algorithm) For each $t\in\mathbb{N}_0^n$, consider a fixed source $P_t(x_t|x_{t-1})$, and a given

$P_t^o(x_{t-1}|y_{t-1})$ obtained for a fixed $Y_{t-1} = y_{t-1}$. Moreover, let $s_t \leq 0$ and $P_t^{(0)}(y_t|y_{t-1}) > 0$ be the initial output probability distribution, and let $P_t^{(k+1)}(y_t|y_{t-1}) = P_t[P_t^{(k)}(y_t|y_{t-1})](y_t|y_{t-1}, x_t)$ and $P_t^{(k+1)}(y_t|y_{t-1}, x_t) = P_t[P_t^{(k)}(y_t|y_{t-1})](y_t|y_{t-1})$ be expressed as follows

$$P_t^{(k+1)}(y_t|y_{t-1}, x_t) = \frac{P_t^{(k)}(y_t|y_{t-1})A_t[P_{t+1}^o]}{\sum_{y_t\in\mathcal{Y}_t} P_t^{(k)}(y_t|y_{t-1})A_t[P_{t+1}^o]}, \tag{20}$$

$$P_t^{(k+1)}(y_t|y_{t-1}) = P_t^{(k)}(y_t|y_{t-1}) \sum_{x_t\in\mathcal{X}_t} \frac{P_t^{\mathrm{Mo}}(x_t|y_{t-1})A_t[P_{t+1}^o]}{\sum_{y_t\in\mathcal{Y}_t} P_t^{(k)}(y_t|y_{t-1})A_t[P_{t+1}^o]}. \tag{21}$$

Then as $k \to \infty$, we obtain for any t that

$$D_t[y_{t-1}, P_t^o, P_t^{(k)}(y_t|y_{t-1}, x_t)] \to D_{s_t}[y_{t-1}, P_t^o]$$
$$\mathbb{I}_t[y_{t-1}, P_t^o, P_t^{(k)}(y_t|y_{t-1}, x_t)] \to R_t(D_{s_t}[y_{t-1}, P_t^o])$$

where $(D_{s_t}[y_{t-1}, P_t^o], R_t(D_{s_t}[y_{t-1}, P_t^o]))$ denotes a point on the cost-to-go curve parametrized by s_t given $P_t^o(x_{t-1}|y_{t-1})$ obtained for a fixed $Y_{t-1} = y_{t-1}$ and

$$\mathbb{I}_t[y_{t-1}, P_t^o, P_t^{(k)}(y_t|y_{t-1}, x_t)] = \sum_{x_t\in\mathcal{X}_t, y_t\in\mathcal{Y}_t} P_t^{\mathrm{Mo}}(x_t|y_{t-1}) P_t^{(k)}(y_t|y_{t-1}, x_t) \Bigg(\log \left(\frac{P_t^{(k)}(y_t|y_{t-1}, x_t)}{P_t^{(k)}(y_t|y_{t-1})} \right) + R_{t+1}(D_{t+1}[y_t, P_{t+1}^o]) \Bigg), \tag{22}$$

$$D_t[y_{t-1}, P_t^o, P_t^{(k)}(y_t|y_{t-1}, x_t)] = \sum_{x_t\in\mathcal{X}_t, y_t\in\mathcal{Y}_t} P_t^{\mathrm{Mo}}(x_t|y_{t-1}) P_t^{(k)}(y_t|y_{t-1}, x_t)\rho_t(x_t, y_t). \tag{23}$$

The implementation of Theorem 5 is illustrated in Algorithm 1. The following theorem supplements Algorithm 1 with a stopping criterion after a finite number of steps.

Theorem 6: (Stopping criterion of Algorithm 1) For each $t \in \mathbb{N}_0^n$, the point $D_{s_t}[y_{t-1}, P_t^o]$ given by (19) admits the following bounds

$$R_t(D_{s_t}[y_{t-1}, P_t^o]) \leq s_t D_{s_t}[y_{t-1}, P_t^o] - \sum_{x_t\in\mathcal{X}_t} \Bigg(P_t^{\mathrm{Mo}}(x_t|y_{t-1}) \log\Big(\sum_{y_t\in\mathcal{Y}_t} P_t(y_t|y_{t-1})A_t[P_{t+1}^o] \Big) \Bigg) - \sum_{y_t\in\mathcal{Y}_t} P_t(y_t|y_{t-1})c_t[y_{t-1}](y_t) \log c_t[y_{t-1}](y_t), \tag{24}$$

$$R_t(D_{s_t}[y_{t-1}, P_t^o]) \geq s_t D_{s_t}[y_{t-1}, P_t^o] - \sum_{x_t\in\mathcal{X}_t} \Bigg(P_t^{\mathrm{Mo}}(x_t|y_{t-1}) \log\Big(\sum_{y_t\in\mathcal{Y}_t} P_t(y_t|y_{t-1})A_t[P_{t+1}^o] \Big) \Bigg) - \max_{y_t\in\mathcal{Y}_t} \log c_t[y_{t-1}](y_t), \tag{25}$$

where $c_t[y_{t-1}](y_t)$ is expressed as a function of fixed y_{t-1}

$$c_t[y_{t-1}](y_t) = \sum_{x_t\in\mathcal{X}_t} \frac{P_t^{\mathrm{Mo}}(x_t|y_{t-1})A_t[P_{t+1}^o]}{\sum_{y_t\in\mathcal{Y}_t} P_t(y_t|y_{t-1})A_t[P_{t+1}^o]}.$$

Algorithm 1 Approximation of the Control Policy Backward in Time (Offline Training)

Input: $\{P_t(x_t|x_{t-1}) : t \in \mathbb{N}_0^n\}$, $\{s_t \leq 0 : t \in \mathbb{N}_0^n\}$, given belief state $P_t^o(x_{t-1}|y_{t-1}) \in \mathcal{B}_t$, $\epsilon > 0$.

Initialize $\{P_t^{(0)}(y_t|y_{t-1}) : t \in \mathbb{N}_0^n\}$
for $t = n : 1$ **do**
  $k \leftarrow 0$
  while $T_{L_t}[y_{t-1}, P_t^o] - T_{U_t}[y_{t-1}, P_t^o] > \epsilon$ **do**
    $P_t^{(k)}(y_t|y_{t-1}, x_t) \leftarrow$ (20)
    $P_t^{(k+1)}(y_t|y_{t-1}) \leftarrow$ (21)
    $R_t(D_t[y_{t-1}, P_t^o]) \leftarrow$ (22)
    $k \leftarrow k + 1$
  end while
end for

Output:
$\{P_t^*[P_t^o](y_t|y_{t-1}, x_t) : t \in \mathbb{N}_1^n\}$, $\{P_t^*[P_t^o](y_t|y_{t-1}) : t \in \mathbb{N}_1^n\}$, $\{R_t(D_{s_t}[y_{t-1}, P_t^o]) : t \in \mathbb{N}_1^n\}$.

Theorem 6, generates a stopping criterion for Algorithm 1 at the k-th iteration by setting the estimation error ϵ per stage, i.e., $T_{L_t}[y_{t-1}, P_t^o] - T_{U_t}[y_{t-1}, P_t^o]|$ where

$$T_{U_t}[y_{t-1}, P_t^o] = \sum_{y_t\in\mathcal{Y}_t} P_t(y_t|y_{t-1})c_t[y_{t-1}](y_t) \log c_t[y_{t-1}](y_t)$$
$$T_{L_t}[y_{t-1}, P_t^o] = \max_{y_t\in\mathcal{Y}_t} \log c_t[y_{t-1}](y_t).$$

Comments on Algorithms 1, 2: Algorithm 1 approximates the control policy $\{P_t^*[P_t^o](y_t|y_{t-1}, x_t) : t \in \mathbb{N}_1^n\}$, the output distribution $\{P_t^*[P_t^o](y_t|y_{t-1}) : t \in \mathbb{N}_1^n\}$, and the cost-to-go function $\{R_t(D_{s_t}[y_{t-1}, P_t^o]) : t \in \mathbb{N}_1^n\}$ as functions of the fixed $Y_{t-1} = y_{t-1}$, the quantized belief state $P_t^o(x_{t-1}|y_{t-1}) \in \mathcal{B}_t$, and also the one-step lookahead belief state $P_{t+1}^o(x_t|y_t) \in \mathcal{B}_{t+1}$. After computing these quantities backward in time, the online Algorithm 2 operates forward in time to evaluate the cost-to-go and identify the approximate minimizers of (6). The initial source and output distributions $P_0(x_0)$ and $P_0(y_0)$ at $t = 0$ are given, yielding the initial control policy $P_0(y_0|x_0)$ and the corresponding posterior $P_1(x_0|y_0)$, which initialize belief state $P_1^*(x_0|y_0)$. At each t, the best policy $P_t^*(y_t|y_{t-1}, x_t)$ is determined by following the best trajectory $P_{t+1}^*(x_t|y_t)$ such that

$$P_{t+1}^*(x_t|y_t) = \arg \min_{P_{t+1}^o(x_t|y_t)\in\mathcal{B}_{t+1}} \sum_{y_{t-1}\in\mathcal{Y}_{t-1}} R_t(D_{s_t}[y_{t-1}, P_t^*])P_t(y_{t-1}), \ \forall t = \mathbb{N}_2^{n-1}, \tag{26}$$

and eventually the minimum in (6) is approximated. Clearly, the larger the search space of the finite belief state, the better the approximation. Ideally, a sufficiently large belief state space can approximate near-optimally the minimum in (6).

IV. Numerical Examples

This section provides numerical simulations to support our theoretical findings that led to Algorithms 1, 2. We assume binary alphabet spaces $\{\mathcal{X}_t = \mathcal{Y}_t = \{0, 1\} : t \in \mathbb{N}_0^n\}$, with Hamming distortion metric given by

$$\rho_t(x_t, y_t) \equiv \rho(x_t, y_t) = \begin{cases} 0, & \text{if } x_t = y_t \\ 1, & \text{if } x_t \neq y_t \end{cases}, \quad \forall t \in \mathbb{N}_0^n. \tag{27}$$

Algorithm 2 Forward Computation of the Approximate Control Policy (Online Computation)

Input: $\{\mathcal{B}_t : t \in \mathbb{N}_1^n\}$ of given $\{P_t^o(x_{t-1}|y_{t-1}) : t \in \mathbb{N}_1^n\}$, outputs of Algorithm 1.

Initialize $P_0(x_0)$, $P_0(y_0)$, $P_1^*(x_0|y_0) = P(x_0|y_0)$
for $t = 1 : n-1$ **do**
  $P_{t+1}^*(x_t|y_t) \leftarrow$ (26)
  $P_t^*(y_t|y_{t-1}, x_t) \leftarrow$
    $P_t^*[P_t^*(x_{t-1}|y_{t-1}), P_{t+1}^*(x_t|y_t)](y_t|y_{t-1}, x_t)$
end for
$P_n^*(y_n|y_{n-1}, x_n) \leftarrow P_n^*[P_n^*(x_{n-1}|y_{n-1})](y_n|y_{t-1}, x_n)$

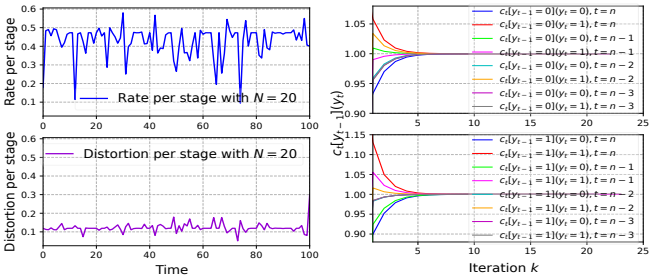
Output:
$\{P_t^*(x_{t-1}|y_{t-1}) : t \in \mathbb{N}_0^n\}$, $\{P_t^*(y_t|y_{t-1}, x_t) : t \in \mathbb{N}_0^n\}$, $R_{[0,n]}^{na}(D_0, D_1, \ldots, D_n)$.

We consider a belief state $P_t^o(x_{t-1}|y_{t-1}) \in \mathcal{B}_t$, that consists of a matrix comprising two "quantized" binary probability distributions drawn from the continuous space. We denote with N_t each quantization level per t, which leads to a belief state space $\mathcal{B}_t$ with size $|\mathcal{B}_t| = N_t^2$, representing the combinations of 2 out of N_t quantized binary distributions.

Example 1: (Time-varying binary symmetric Markov source) The source distribution $P_t(x_t|x_{t-1})$ at each $t \in \mathbb{N}_1^n$ is chosen such that for each t, we have

$$P_t(x_t|x_{t-1}) = \begin{pmatrix} 1-\alpha_t & \alpha_t \\ \alpha_t & 1-\alpha_t \end{pmatrix}, \ \alpha_t \in (0,1). \tag{28}$$

Moreover, we choose the quantization levels $\{N_t = N : t \in \mathbb{N}_1^n\}$ and the stagewise Lagrange multipliers $\{s_t = s : t \in \mathbb{N}_0^n\}$. We demonstrate the results applying Algorithms 1, 2 in Fig. 2 for $N = 20$, $s = -2$, and $n = 100$, whereas Fig. 2b illustrates several time stages selected during backward computation to verify the convergence of Algorithm 1.



(a) Stagewise rate & distortion (b) Stagewise convergence

Fig. 2: Illustration of the stagewise rate & distortion and convergence for the time-varying case.

V. Conclusion

We derived a non-asymptotic lower bound for a zero-delay variable-rate lossy source coding system assuming discrete Markov sources. We derived new structural and convexity properties of NRDF that helped us cast the problem as an unconstrained partially observable finite-horizon stochastic DP and solved it approximately via a novel dynamic AM scheme to compute the control policy and the cost-to-go function through an offline training algorithm followed by an online computation. Our theoretical results are supplemented with simulation studies.

References

[1] E. C. Strinati et al., "Goal-oriented and semantic communication in 6G AI-native networks: The 6G-GOALS approach," in *Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*, 2024, pp. 1–6.

[2] C. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Conv. Rec.*, pp. 142–163, 1993.

[3] S. Yüksel and T. Başar, *Stochastic networked control systems: stabilization and optimization under information constraints*. Birkhäuser New York, NY, 2014.

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

[5] P. Varaiya and J. Walrand, "Causal coding and control for markov chains," *Systems & Control Letters*, vol. 3, no. 4, pp. 189–192, 1983.

[6] D. P. Bertsekas, *Dynamic programming and optimal control*. Athena Scientific, 2005.

[7] 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.

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

[9] M. Ghomi, T. Linder, and S. Yüksel, "Zero-delay lossy coding of linear vector Markov sources: Optimality of stationary codes and near optimality of finite memory codes," *IEEE Trans. Inf. Theory*, vol. 68, no. 5, pp. 3474–3488, 2022.

[10] 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, 2012.

[11] L. Cregg, T. Linder, and S. Yüksel, "Reinforcement learning for near-optimal design of zero-delay codes for Markov sources," *IEEE Trans. Inf. Theory*, vol. 70, no. 11, pp. 8399–8413, 2024.

[12] S. Tatikonda, A. Sahai, and S. Mitter, "Stochastic linear control over a communication channel," *IEEE Trans. Autom. Control*, vol. 49, pp. 1549 – 1561, 2004.

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

[14] A. K. Gorbunov and M. S. Pinsker, "Nonanticipatory and prognostic epsilon entropies and message generation rates," *Problems Inf. Transmiss.*, vol. 9, no. 3, pp. 184–191, July-Sept. 1972.

[15] T. Tanaka, K. K. K. Kim, P. A. Parrilo, and S. K. Mitter, "Semidefinite programming approach to Gaussian sequential rate-distortion trade-offs," *IEEE Trans. Autom. Control*, vol. 62, no. 4, pp. 1896–1910, April 2017.

[16] P. A. Stavrou, J. Østergaard, and C. D. Charalambous, "Zero-delay rate distortion via filtering for vector-valued Gaussian sources," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 5, pp. 841–856, Oct 2018.

[17] P. A. Stavrou, M. Skoglund, and T. Tanaka, "Sequential source coding for stochastic systems subject to finite rate constraints," *IEEE Trans. Autom. Control*, vol. 67, no. 8, pp. 3822–3835, 2022.

[18] R. E. Blahut, *Principles and practice of information theory*. Addison-Wesley Longman Publishing Co., Inc., 1987.

[19] J. L. Massey, "Causality, feedback and directed information," in *Proc. Int. Symp. Inf. Theory Appl.*, Nov. 27-30 1990, pp. 303–305.

[20] P. A. Stavrou, T. Charalambous, C. D. Charalambous, and S. Loyka, "Optimal estimation via nonanticipative rate distortion function and applications to time-varying Gauss-Markov processes," *SIAM J. on Control Optim.*, vol. 56, no. 5, pp. 3731–3765, 2018.

[21] P. A. Stavrou, C. D. Charalambous, and C. K. Kourtellaris, "Information nonanticipative rate distortion function and its applications," *arxiv.org*, 2014.

[22] C. D. Charalambous and P. A. Stavrou, "Directed information on abstract spaces: Properties and variational equalities," *IEEE Trans. Inf. Theory*, vol. 62, no. 11, pp. 6019–6052, Nov 2016.

[23] A. K. Gorbunov and M. S. Pinsker, "Prognostic epsilon entropy of a Gaussian message and a Gaussian source," *Problems Inf. Transmiss.*, vol. 10, no. 2, pp. 93–109, Apr.-June 1972, translation from Problemy Peredachi Informatsii, vol. 10, no. 2, pp. 5–-25, April-June 1974.

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

[25] D. Bertsekas, *Reinforcement learning and optimal control*. Athena Scientific, 2019.