An Achievability Proof for the Lossy Coding of Markov Sources with Feed-Forward

Shahab Asoodeh, Fady Alajaji and Tamás Linder
Department of Mathematics and Statistics
Queen's University, Kingston, ON, K7L 3N6, Canada
Email: {asoodehshahab,fady,linder}@mast.queensu.ca

***Abstract*—The lossy source coding problem with feed-forward link was recently introduced in [1] and the corresponding rate-distortion function was derived in [2] and [3] for stationary and ergodic sources and for arbitrary sources with memory, respectively. The achievability schemes of [2] and [3] are mainly based on codetrees. In this work, we give an alternative proof of achievability for binary asymmetric Markov sources via a simple coding scheme that utilizes optimal lossy coding for Bernoulli sources. We then generalize this coding scheme for *m*-ary Markov sources and show its optimality for the distortion region where the Shannon lower bound is tight.**

I. Introduction

The emerging applications of sensor networks have given special significance to the problem of source coding with side information (SI) at the decoder. In this scenario, the main goal of each sensor is to convey what it measures to the receiver. Each sensor encodes the observed source $X^n = (X_1, X_2, \dots, X_n)$ into a message M of nR bits and transmits it to the receiver where R is the communication rate. The receiver has access to a processed version Y^n of X^n. Hence, when decoding X^n the decoder has available not only the message M received directly from the sensor, but also SI Y^n. The goal is then to minimize the reconstruction distortion at a fixed transmission rate or equivalently, to minimize the rate for a given distortion. The rate-distortion function is known and given by Wyner and Ziv in [4] when $\{(X^n, Y^n)\}_{n=1}^{\infty}$ is an independent and identically distributed (i.i.d.) process. In the Wyner-Ziv model the decoder is assumed to have a non-causal SI which is crucial for its binning encoding scheme. The Wyner-Ziv problem with causal SI was considered in [5]. Another structural restriction that has been studied in the source coding literature is the delay in the SI. In this problem, there is a delay of d time instants between the time when the ith source symbol is fed into the encoder and the time when the corresponding SI, Y_i, is observed at the decoder. Clearly for memoryless sources the Wyner-Ziv setting with strictly delayed SI, i.e., with $d > 0$, reduces to the original source coding problem without any SI. However, this conclusion does not generally hold if the source has memory. In this context, a growing number of works have focused on the scenario in which $X_i = Y_i$ for $i = 1, 2, \dots, n$. This setting is typically referred to as *source coding with noiseless feed-forward* as introduced in [1]. In this setting, the encoder maps X^n to $M \in \{1, 2, \dots, 2^{nR}\}$ and sends it to the decoder. The decoder then receives M together with SI with delay $d > 0$, so that $Y_i = X_{i-d}$ for $i = 1, 2, \dots, n$. It is clear that this model is valid only if the delay is at least $n + 1$. For simplicity, we write $d = 1$ when delay is actually $d = n + 1$. Although this setup may look too idealized for applications, it can be used to model a number of scenarios in sensor networks and economics; see e.g., [3] and the references therein. The model for this setting is depicted in Figure 1.

This work was supported in part by NSERC of Canada.

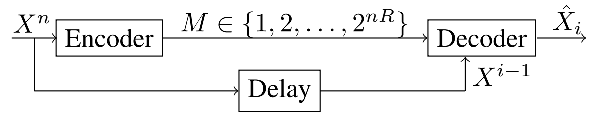

Fig. 1: Feed-forward source coding with rate R and delay 1.

The rate-distortion function for this problem was derived in [3] in terms of multi-letter spectral mutual information rates for arbitrary (not necessarily stationary and ergodic) sources with memory. This function was shown to be easily analytically evaluated for some special classes of sources in [6] and its numerical calculation for stationary and ergodic sources was addressed in [2]. A simpler formula for stationary and ergodic sources was obtained in [2] using the notion of *n*th order feed-forward rate-distortion function. In [7], Weissman and El Gamal gave a simple, yet inspiring, scheme to achieve the rate-distortion function when both decoder and encoder know the SI (causally or non-causally) based on an appropriate partitioning of the source sequence X^n before encoding. The achievability schemes proposed in [2], [3] are based on codetrees.

In this work, we adopt the idea given in [7] and developed in [10] and propose a constructive coding scheme for m-ary Markov sources which achieves the feed-forward rate distortion function and is conceptually simpler than the one given in [2], [3].

This paper is organized as follows. In Section II, we review some basic known results about the rate-distortion function with feed-forward. We then present a constructive feed-forward achievability scheme for binary asymmetric Markov sources (BAMS) in Section III. In Section IV, we generalize this achievability scheme for m-ary Markov sources and show that the scheme is still optimal for a particular distortion region. In Section V, we conclude the paper.

II. Review of the Feed-Forward Rate-Distortion Function

Consider a stationary and ergodic source $\{X_i\}$ with finite alphabet $\mathcal{X}$, finite reconstruction alphabet $\hat{\mathcal{X}}$ and a distortion function on pairs of sequences $d_n : \mathcal{X}^n \times \hat{\mathcal{X}}^n \rightarrow \mathbb{R}^+$. We assume that d_n is the average per-letter distortion, i.e, $d_n(x^n, \hat{x}^n) = \frac{1}{n}\sum_{i=1}^n d(x_i, \hat{x}_i)$ where $d : \mathcal{X} \times \hat{\mathcal{X}} \rightarrow \mathbb{R}^+$. Note that if the joint random process $\{(X_n, \hat{X}_n)\}$ is stationary, then $E[d_n(X^n, \hat{X}^n)] = E[d(X, \hat{X})]$.

The concept of feed-forward was introduced in [1] in the context of competitive prediction and studied further in [3], [6]. In the following, we study the feed-forward rate-distortion function for binary asymmetric Markov sources in which case competitive prediction can no longer be used. We instead use the general formula for the feed-forward rate-distortion function.

Definition 1. *An $(n, 2^{nR})$ source code with feed-forward of rate R and blocklength n, consists of encoder function ψ and a set of decoder functions ξ_i at time instants $i = 1, 2 \ldots, n$, where*

$$\begin{aligned} \psi &: \mathcal{X}^n \rightarrow \{1, 2, \ldots, 2^{nR}\}, \\ \xi_i &: \{1, 2, \ldots, 2^{nR}\} \times \mathcal{X}^{i-1} \rightarrow \hat{\mathcal{X}}. \end{aligned}$$

In this scenario the decoder has causal access to the side information which consists of the previous source symbols; that is, while estimating x_i, the decoder knows all previous source symbols x^{i-1}.

Definition 2. *R is an achievable rate at expected distortion D if for any $\epsilon > 0$, for all sufficiently large n, there exists an $(n, 2^{nR})$ code such that*

$$E\big[d_n(X^n, \hat{X}^n)\big] \leq D + \epsilon.$$

Definition 3. *The feed-forward rate-distortion function, $R_{ff}(D)$, is the infimum of all achievable rates for a given distortion D.*

The feed-forward rate-distortion function for a stationary and ergodic source is derived in [3] in terms of the directed information, $I(\hat{X}^n \rightarrow X^n)$ defined as

$$\begin{aligned} I(\hat{X}^n \rightarrow X^n) &:= H(X^n) - H(X^n || \hat{X}^n) && (1) \\ &= \sum_{i=1}^n I(\hat{X}^i; X_i | X^{i-1}) && (2) \end{aligned}$$

and is given by the following theorem which combines Theorems 1 and 2 in [3].

Theorem 1 (Venkataramanan et al. [3])**.** *For a stationary and ergodic source $\{X_i\}$ with finite alphabet $\mathcal{X}$, the feed-forward rate distortion function, $R_{ff}(D)$, at expected distortion D is given by*

$$R_{ff}(D) = \inf_{P_{\hat{X}|X} : E[d(X, \hat{X})] \leq D} \lim_{n \rightarrow \infty} \frac{1}{n} I(\hat{X}^n \rightarrow X^n),$$

where the infimum is taken over all conditional distributions $P_{\hat{X}|X}$ for which the joint process $\{(\hat{X}_n, X_n)\}$ is stationary and ergodic process.

For stationary and ergodic sources, [2] showed that $R_{ff}(D)$ can be obtained using a simpler formula. Let $R_{n,ff}(D)$ be the nth order feed-forward rate distortion function defined by

$$R_{n,ff}(D) := \min_{P_{\hat{X}^n|X^n} : E[d_n(X^n, \hat{X}^n)] \leq D} \frac{1}{n} I(\hat{X}^n \rightarrow X^n). \quad (3)$$

The following theorem then gives the feed-forward rate-distortion function.

Theorem 2 (Naiss et al. [2])**.** *For the stationary and ergodic source described in Theorem 1, the feed-forward rate-distortion function is given by*

$$R_{ff}(D) = \lim_{n \rightarrow \infty} R_{n,ff}(D).$$

Theorem 2 states that the infimum and limit can be interchanged in Theorem 1. This brings a great deal of simplification in terms of calculation; for example, a modification of the Arimoto-Blahut algorithm is used in [2] to numerically estimate $R_{ff}(D)$.

As an immediate consequence of the results in [1], one can conclude that if $\{X_i\}$ is an i.i.d. process, then the presence of the feed-forward link does not improve the rate-distortion function. The result of [1] can also be used to show that for binary symmetric Markov sources with transition probability q, BSMS(q), $R_{ff}(D) = H_b(q) - H_b(D)$ where H_b denotes the binary entropy function. This expression is equal to the lower bound on the rate-distortion function without feed-forward obtained by Berger [9] and Gray [8] for a particular distortion region; thus the feed-forward link helps improve the rate-distortion function for BSMS.

III. Binary Asymmetric Markov Sources

Let $\mathcal{B}(p)$ denote the Bernoulli distribution with transition probability p, that is, $W \sim \mathcal{B}(p)$ if and only if $P(W = 1) = p$ and $P(W = 0) = 1 - p$. It is easy to show that any binary asymmetric Markov source $\{X_i\}$ with transition probabilities p and q, $(0 < p, q < 1)$, BAMS(p, q), can be represented by two Bernoulli sources as follows:

$$X_i = X_{i-1} W_i^1 + (1 - X_{i-1}) W_i^2, \quad (4)$$

where $\{W_i^1\}$ and $\{W_i^2\}$ are two independent processes and $W_i^1 \sim \mathcal{B}(1-q)$, $W_i^2 \sim \mathcal{B}(p)$ and X_{i-1}, W_i^1 and W_i^2 are independent for every i. Let $\pi = (\pi_1, \pi_2)$ denote the invariant distribution for BAMS(p, q) and consider the Hamming distortion measure.

As shown in [2], the feed-forward rate-distortion function for the BAMS(p, q) represented by (4) is given by

$$R_{ff}(D) = \pi_1 H_b(p) + \pi_2 H_b(q) - H_b(D). \quad (5)$$

Setting $p = q$, BAMS(p, q) reduces to BSMS(q) and (5) gives $R_{ff}(D) = H_b(q) - H_b(D)$ as proved in [1]. In the following we present an achievability scheme based on the scheme proposed in [7] and later developed in [10]. We will see later that the argument given in [10] needs refinement in our case.

We first partition the given source sequence $\{X_n\}_{n=1}^{\infty}$ into two sub-sequences, the X_i's following a 0 and the X_i's

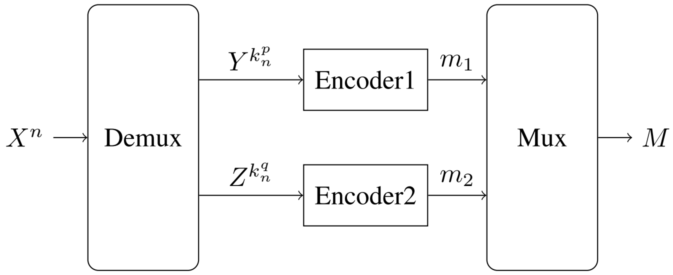

Fig. 2: The block diagram of the encoder.

following a 1 and then encode separately these two sub-sequences. We describe in detail the encoding process for one sub-sequence as the other one is similar. Given the source sequence $\{X_n\}_{n=1}^{\infty}$, let N_i be the time index of the ith zero in the sequence and $Y_i := X_{N_i+1}$. It is easy to show that $\{Y_n\}$ is an i.i.d. process generated by $\mathcal{B}(p)$. To see this, consider the following

$$\begin{aligned}
P(Y^i = y^i) &= \prod_{j=1}^{i} P(Y_j = y_j | Y^{j-1} = y^{j-1}) \\
&= \prod_{j=1}^{i} \sum_{n=1}^{\infty} P(Y_j = y_j | Y^{j-1} = y^{j-1}, N_j = n) \\
&\qquad \times P(N_j = n | Y^{j-1} = y^{j-1}) \\
&\overset{(a)}{=} \prod_{j=1}^{i} \sum_{n=1}^{\infty} P(X_{n+1} = y_j | X_n = 0) \\
&\qquad \times P(N_j = n | Y^{j-1} = y^{j-1}) \\
&\overset{(b)}{=} \prod_{j=1}^{i} p^{y_j} (1-p)^{1-y_j},
\end{aligned}$$

where (a) is due to the Markovity of the source and the fact that event $\{N_j = n\}$ implies $\{X_n = 0\}$ and (b) holds because from (4), $P(X_{n+1} = y_j | X_n = 0) = p^{y_j}(1-p)^{1-y_j}$.

The key idea of the encoding scheme is to use an optimal rate-distortion code of a Bernoulli $\mathcal{B}(p)$ source to encode the sequence Y^i. By the strong law of large numbers for Markov chains, we know that the number of zeros in a sufficiently large source sequence X^n is approximately $n\pi_1$, in other words, as $n \to \infty$, with probability one,

$$\frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i = 0\}} \to \pi_1.$$

Let $k_n^p = \lceil n(\pi_1 - \delta) \rceil$ and $\mathcal{E}_n$ be a binary random variable defined as follows

$$\mathcal{E}_n = \begin{cases} 0 & \text{if } N_{k_n^p} \le n, \\ 1 & \text{if } N_{k_n^p} > n. \end{cases} \tag{6}$$

When $\mathcal{E}_n = 0$ we encode $(Y_1, Y_2, \ldots, Y_{k_n^p})$ using an optimal rate-distortion code for the source $\mathcal{B}(p)$ at rate R and if $\mathcal{E}_n = 1$ we do not encode and simply send a particular vector. The analysis given in [10] is not applicable in our case since $Y^{k_n^p}$ is not i.i.d. when conditioned on $\mathcal{E}_n = 0$ and thus we need a more involved conditioning argument.

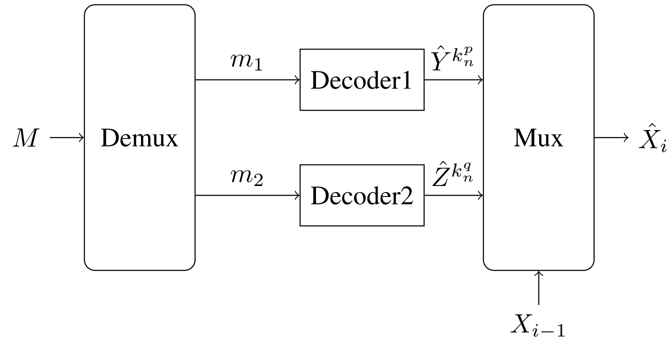

Fig. 3: The block diagram of the decoder at the ith time instant.

Let $(\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_{k_n^p})$ be the reproduction sequence and assume that the Hamming distortion between the two sequences is normalized, that is,

$$d(y^{k_n^p}, \hat{y}^{k_n^p}) = \frac{1}{k_n^p} \sum_{i=1}^{k_n^p} d(y_i, \hat{y}_i).$$

Let D_n denote the distortion in encoding $Y^{k_n^p}$ using a rate-distortion code of rate R that is optimal for $\mathcal{B}(p)$. Then

$$\begin{aligned}
D_n &:= E\big[d(Y^{k_n^p}, \hat{Y}^{k_n^p})\big] \\
&= E\big[d(Y^{k_n^p}, \hat{Y}^{k_n^p}) | \mathcal{E}_n = 0\big] P(\mathcal{E}_n = 0) \\
&\quad + E\big[d(Y^{k_n^p}, \hat{Y}^{k_n^p}) | \mathcal{E}_n = 1\big] P(\mathcal{E}_n = 1).
\end{aligned} \tag{7}$$

Since the sequence $Y_1, Y_2, \ldots$ is i.i.d. with distribution $\mathcal{B}(p)$, then obviously

$$\lim_{n\to\infty} D_n = D_p(R), \tag{8}$$

where $D_p(R)$ is the distortion-rate function of a Bernoulli source $\mathcal{B}(p)$ at rate R. Note that since all terms in (7) are nonnegative, we have

$$\begin{aligned}
D_p(R) &= \lim_{n\to\infty} D_n & (9)\\
&\ge \limsup_{n\to\infty} E\big[d(Y^{k_n^p}, \hat{Y}^{k_n^p}) | \mathcal{E}_n = 0\big] P(\mathcal{E}_n = 0).
\end{aligned}$$

On the other hand, since $d(x, y) \le 1$ for $x, y \in \{0, 1\}$, the distortion is deterministically upper bounded by 1 when $\mathcal{E}_n = 1$. Thus, if D_n^p denotes the expected distortion of our scheme, we have

$$D_n^p \le E\big[d(Y^{k_n^p}, \hat{Y}^{k_n^p}) | \mathcal{E}_n = 0\big] P(\mathcal{E}_n = 0) + P(\mathcal{E}_n = 1). \tag{10}$$

Hence together with the fact that $P(\mathcal{E}_n = 1) \to 0$ as $n \to \infty$, we obtain

$$\begin{aligned}
\limsup_{n\to\infty} D_n^p &\le \limsup_{n\to\infty} E\big[d(Y^{k_n^p}, \hat{Y}^{k_n^p}) | \mathcal{E}_n = 0\big] P(\mathcal{E}_n = 0) \\
&\le D_p(R).
\end{aligned} \tag{11}$$

The encoding scheme for the other sub-sequence is similar to the one described above. Let M_i be the time index of ith one in the source sequence X^n and $Z_i := X_{M_i+1}$. We can again show that sequence $\{Z_i\}$ is i.i.d. with distribution $\mathcal{B}(q)$. Letting k_n^q be equal to $\lceil n(\pi_2 - \delta) \rceil$, we can use the same coding scheme as before for the sequence $Z^{k_n^q}$. Similarly, let D_n^q define the distortion of the encoding scheme in this case.

The encoder structure is depicted in Figure 2, where m_1 and m_2 are the messages corresponding to $Y^{k_n^p}$ and $Z^{k_n^q}$ and are produced by two Bernoulli rate-distortion encoders.

At the receiver side, we receive the codewords for $Y^{k_n^p}$ and $Z^{k_n^q}$ and hence are able to reconstruct $\hat{Y}^{k_n^p}$ and $\hat{Z}^{k_n^q}$. We then need the causal information, i.e., X^{i-1} at time i to reconstruct the source sequence. In other words, at time i, causal information X^{i-1} helps the decoder pick the appropriate letter between $\hat{Y}_i$ and $\hat{Z}_i$ depending on whether $X_{i-1} = 0$ or $X_{i-1} = 1$. The decoder structure is depicted in Figure 3. The total distortion for encoding the source sequence X^n using our parallel encoding scheme is the sum of the distortion of each sub-sequence and therefore can be obtained in terms of k_n^p, D_n^p, k_n^q and D_n^q. Note that by the definition of k_n^p and k_n^q, there are at most $2n\delta$ many source letters which are not encoded and hence contribute to the total normalized distortion by at most 2δ. For the total normalized distortion, we can write

$$D_{tot} \leq \frac{1}{n}\big(k_n^p D_n^p + k_n^q D_n^q + 2n\delta\big), \tag{12}$$

where $2n\delta$ is the contribution of uncoded bits. Letting $n \to \infty$, we can write:

$$\begin{aligned} D_{tot} &\leq (\pi_1 - \delta) D_P(R) + (\pi_2 - \delta) D_q(R) + 2\delta \\ &\leq \pi_1 D_p(R) + \pi_2 D_q(R) \\ &\qquad + \delta(\underbrace{1 - D_p(R)}_{\geq 0}) + \delta(\underbrace{1 - D_q(R)}_{\geq 0}) \\ &= \pi_1 D_p(R) + \pi_2 D_q(R) + \epsilon. \end{aligned} \tag{13}$$

The entire encoding function can be described via the following mapping

$$\begin{aligned} \{0,1\}^{k_n^p + k_n^q} \quad \to \quad & \{1, 2, \ldots, 2^{k_n^p R}, 2^{k_n^p R} + 1\} \times \\ & \{1, 2, \ldots, 2^{k_n^q R}, 2^{k_n^q R} + 1\}, \end{aligned}$$

which emphasizes that for the sequence $Y^{k_n^p}$ we need an index chosen from $\{1, 2, \ldots, 2^{k_n^p R}\}$ and also one extra index for the case of $\mathcal{E}_n = 1$ and similarly for $Z^{k_n^q}$. Clearly the rate of this coding scheme is

$$\begin{aligned} R_{tot} &= \frac{1}{n} \log\left[(2^{k_n^p R} + 1)(2^{k_n^q R} + 1)\right] \leq \frac{1}{n}(k_n^p R + k_n^q R + 2) \\ &\leq R + \epsilon, \end{aligned} \tag{14}$$

where we use the obvious inequality $\log(1 + x) \leq 1 + \log x$ for $x \geq 1$. Combining (13) and (14) with the fact that $R_p(D) = H_b(p) - H_b(D)$ and $R_q(D) = H_b(q) - H_b(D)$, we can conclude that $R_{ff}(D) = \pi_1 H_b(p) + \pi_2 H_b(q) - H_b(D)$ is achievable.

IV. Non-Binary Markov Sources

Consider a stationary m-ary Markov source $\{X_i\}$ with transition matrix

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}$$

with $0 < p_{ij} < 1$, $1 \leq i, j \leq m$ and invariant distribution $P(X_1 = i) = \pi_i$ for $i = 1, 2, \ldots, m$. Then clearly $H(X_n|X_{n-1}) = \sum_{i=1}^m \pi_i H(\mathbf{P}_i)$ where $\mathbf{P}_j$ denotes the jth row of $\mathbf{P}$. The following theorem gives the feed-forward rate-distortion function, $R_{ff}(D)$ for $\{X_i\}$ when D is under a certain threshold.

Theorem 3. *For an m-ary stationary Markov source with transition matrix $\boldsymbol{P}$ and invariant distribution $\pi = (\pi_1, \pi_2, \ldots, \pi_m)$,*

$$R_{ff}(D) = \sum_{i=1}^m \pi_i H(\boldsymbol{P}_i) - H(D) - D\log(m-1)$$

for $0 \leq D \leq (m-1)p_{min}$ where

$$p_{min} = \min_{0 \leq i,j \leq m} \{p_{i,j}\}.$$

Proof. The converse part is an easy application of Fano's inequality as follows

$$\begin{aligned} I(\hat{X}^n \to X^n) &= \sum_{i=1}^n I(\hat{X}^i; X_i | X^{i-1}), \\ &= \sum_{i=1}^n H(X_i|X^{i-1}) - H(X_i|X^{i-1}, \hat{X}^i) \\ &\geq H(\pi) + (n-1)\sum_{k=1}^m \pi_k H(\mathbf{P}_i) \\ &\qquad - \sum_{i=1}^n H(X_i|\hat{X}_i), \end{aligned} \tag{15}$$

where the inequality is due to the fact that conditioning reduces entropy. Applying Fano's inequality to (15), normalizing by n and then taking the limit as $n \to \infty$, we can write

$$R_{ff}(D) \geq \sum_{k=1}^m \pi_k H(\mathbf{P}_i) - H(D) - D\log(m-1).$$

The achievability scheme is similar to the one proposed for BAMS except that here we have m i.i.d subsequences $\{Y_{i,j}\}$ for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots$ with probability mass functions $\mathbf{P}_i, i = 1, \ldots, m$. We encode each of these subsequences using an optimal rate-distortion code with rate $R_i(D)$ for which we know the following [11, Page 61]

$$R_i(D) = H(\mathbf{P}_i) - H_b(D) - D\log(m-1),$$

for $0 \leq D \leq (m-1)p_{min}^i$ where $p_{min}^i = \min\{p_{i1}, \ldots, p_{im}\}$. Therefore using the argument given in Section III, the coding rate for the entire scheme is

$$\begin{aligned} R &= \sum_{i=1}^m \pi_i R_i(D) \\ &= \sum_{k=1}^m \pi_k H(\mathbf{P}_i) - H(D) - D\log(m-1), \end{aligned}$$

for $0 \leq D \leq (m-1)p_{min}$ while the total distortion is

$$D_{tot} \leq D + \epsilon,$$

which completes the proof.

V. Conclusion

In this paper we considered the rate-distortion function, $R_{ff}(D)$, for Markov sources when a noiseless feed-forward link is causally available from the encoder to the decoder. Here we propose a constructive coding scheme to achieve $R_{ff}(D)$ for binary asymmetric Markov sources which uses the idea of partitioning the source sequence prior to encoding. Using this scheme, we show that achieving $R_{ff}(D)$ for binary (symmetric and asymmetric) Markov sources reduces to the optimal rate-distortion coding of Bernoulli sources. This scheme is also generalized for m-ary Markov sources and is shown to be optimal when the distortion D belongs to the region where the Shannon lower bound is met with equality.

References

[1] T. Weissman and N. Merhav, ”On competitive prediction and its relation to rate-distortion theory”, *IEEE Trans. Inf. Theory*, vol. 49, pp. 3185-3194, Sep. 2003.

[2] I. Naiss and H. Permuter, ”Computable bounds for rate distortion with feed-forward for stationary and ergodic sources”, [Online]. Available: arXiv:1106.0895v1.

[3] R. Venkataramanan and S. S. Pradhan, ”Source coding with feed-forward: Rate-distortion theorems and error exponents for a general source”, *IEEE Trans. Inform. Theory*, vol. 53, pp. 2154-2179, Jun. 2007.

[4] A. Wyner and J. Ziv, ”The rate-distortion function for source coding with side information at the deocer”, *IEEE Trans. Inf. Theory*, vol. 22, pp. 1-10, Jan. 1976.

[5] A. El Gamal and T. Weissman, ”Source coding with causal side information at the decoder”, *In Proc. 43rd Annual Allerton Conf. on Communications, Control, and Computing*, Monticello, IL, Sept. 28th-30th, 2005.

[6] R. Venkataramanan and S. S. Pradhan, ”On computing the feedback capacity of channels and the feed-forward rate-distortion function of sources”, *IEEE Trans. Comm.*, vol. 58, pp. 1889-1896, Jul. 2010.

[7] T. Weissman and A. El Gamal, ”Source coding with limited-look-ahead side information at the decoder”, *IEEE Trans. Comm.*, vol. 52, Issue 12, pp. 5218-5239, December 2006.

[8] R. M. Gray, ”Information rates of autoregressive processes”, *IEEE Trans. Inf. Theory*, Vol. 16, No. 9, pp. 412-421, July 1970.

[9] T. Berger, ”Explicit bounds to R(D) for a binary symmetric Markov source”, *IEEE Trans. Inf. Theory*, vol. 23, pp. 52-59, Jan. 1977.

[10] O. Simeone and H. Permuter, ”Source coding when the side information may be delayed”, *submitted to IEEE Trans. Inf. Theory, July, 2012*, [Online]. Available: arxiv.org/abs/1109.1293.

[11] T. Berger, *Rate Distortion Theory; A Mathematical Basis for Data Compression*, New York, Printice Hall, 1971.