Sending a Lossy Version of the Innovations Process is Suboptimal in QG Rate-Distortion

Kwang Taik Kim and Toby Berger
School of Electrical and Computer Engineering
Cornell University, NY 14853, U.S.A.
Email: kk274@cornell.edu, berger@ece.cornell.edu

***Abstract*— In the critical range $0 < D \leq D_c$, the MSE rate-distortion function of a time-discrete stationary autoregressive Gaussian source is equal to that of a related time-discrete i.i.d. Gaussian source. This suggests that perhaps an optimum encoder should compute the related memoryless sequence from the given source sequence with memory and then use a code of rate $R(D)$ to convey the memoryless sequence to the decoder with an MSE of D. In this scenario, the question is, "For $D \leq D_c$ can a D-admissible code for the original source be obtained via the R-D coding of the innovations process and additional post-processing at the decoder without having to provide any additional information of positive rate?" We show that the answer of this question often is "No."**

I. Introduction

For many sources with memory, the rate-distortion function equals that of a related memoryless source over an interval between zero and a certain critical distortion. For instance, in the critical range $0 < D \leq D_c$, the MSE rate-distortion function of a time-discrete stationary autoregressive Gaussian source is equal to that of a time-discrete i.i.d. Gaussian source. The sequence of successive one-step prediction errors, also called the *innovations process*, is stationary, zero-mean, uncorrelated, and Gaussian. Let us call it $\{Z_k\}$. Rate-distortion theory tells us that the sequence of successive one-step prediction errors $\{Z_k\}$ can be encoded with an MSE of distortion D using any data rate $R > \frac{1}{2}\log\left(\frac{Q_0}{D}\right)$ where Q_0 is the entropy rate power or the variance of the minimum MSE estimate of the time-discrete stationary autoregressive Gaussian source sequence $\{X_k\}$ based on $\{X_j,\ j < k\}$. Hence, in the range $0 < D \leq D_c$ where D_c is an essential infimum of $\Phi_{\mathbf{X}}(\omega)$, the MSE rate-distortion function of the memory sequence $\{X_k\}$ is equal to that of the memoryless sequence $\{Z_k\}$. This suggests that perhaps an optimum encoder should compute the memoryless sequence $\{Z_k\}$ from the memory sequence $\{X_k\}$ and then use a code of rate $\frac{1}{2}\log\left(\frac{Q_0}{D}\right)$ to convey the memoryless sequence $\{Z_k\}$ to the decoder with an MSE of distortion D. However, it is unclear whether or not the receiver could use these D-admissible estimates of lossy one-step prediction errors with a perfectly known initial state to generate D-admissible estimates of the memory sequence [3]. The situation is summarized in Figure 1. A moment thought about this question reveals that it is not possible to generate D-admssible estimates $\{\hat{X}_t(D)\}$ simply by expending all the allotted information, $\{\hat{Z}_t(D)\}$ and then summing said estimates to reproduce the sequence $\{\hat{X}_t(D)\}$ because the errors are cumulative. These make the problem nontrivial.

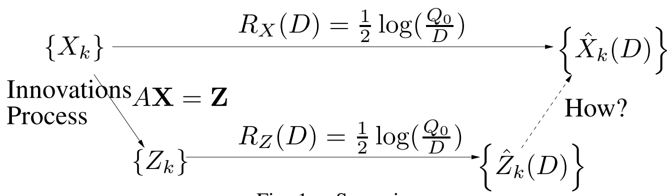

Fig. 1. Scenario

The goal of this paper is to answer this question. Through this answer, we hope to gain more insights into issues such as the intriguing and confounding relationship between the lossy compression scheme for a memory source and the lossy compression of its innovations, as well as whether or not R-D code of the latter with a perfectly known initial state can be used for lossy compression of the original source.

II. Problem Formulation

Let $\{X_k\}_{k=0}^{\infty}$ be a sequence of *time-discrete* stationary, ergodic Gaussian source. Let $\{Z_k\}_{k=0}^{\infty}$ be a sequence of successive one-step prediction errors from the original source $\{X_k\}_{k=0}^{\infty}$, also called the *innovations process* which is stationary, zero-mean, uncorrelated, and Gaussian. Let $\{\hat{X}_k(D)\}_{k=0}^{\infty}$ and $\{\hat{Z}_k(D)\}_{k=0}^{\infty}$ be a sequence of a D-admissible estimate of $\{X_k\}$ and that of $\{Z_k\}$ respectively where $0 < D \leq D_c$; i.e., D is in critical distortion region.

As the first step in our paper, we consider the time-discrete stationary, ergodic Gaussian source $\{X_k\}$ with side information of D-admissible estimates of innovations of the original source $\{\hat{Z}_k(D)\}$ at decoder. A perfectly known initial state $\mathbf{x}_0$ also plays the role of side information at decoder. The framework is summarized in Figure 2. Since encoder 2 observes the original source, it doesn't need its initial state and the lossy compression of its innovations necessarily. The focus here is on studying the characterization of rate R_2 for which this system can deliver D-admissible estimates of the original source in the critical distortion region, and its lower bound.

The reason why we use the framework in Figure 2 is the following; if 0 is a tight lower bound of rate R_2 in the critical distortion region, then a code for the original source can be obtained via R-D coding of the innovations process,

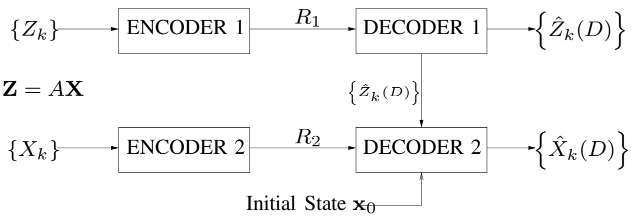

Fig. 2. Framework

and then an additional post-processing at the decoder to get a reconstruction of the original signal, based on the approximate version of the innovations it knows. Otherwise, it cannot be obtained. Here we show that 0 is not a tight lower bound of rate R_2 in the critical distortion region thereby showing that if the decoder is provided with a distorted version of the innovations process and a known initial state, it needs an additional non-negligible description rate about the source if it wants to reconstruct the source to within the same distortion.

III. Main Results

A. Suboptimality of Sending a Lossy Version of the Innovations Process

First, state the main theorem of this paper.

Theorem 3.1: Let the sequences $\{X_k\}$, $\{\hat{X}_k(D)\}$, $\{Z_k\}$ and $\{\hat{Z}_k(D)\}$ be defined as discussed previously. Assume distortion $0 < D \leq D_c$, $\min\limits_k \lambda_k^{(n)}$, where $\lambda_k^{(n)}$ is an eigenvalue of the correlation matrix of $\mathbf{X}$, Φ_n, or an inverse of essential supremum of $g(\omega)$, where $g(\omega) = \frac{1}{\sigma^2}\left|\prod_{k=0}^{m} a_k e^{-ik\omega}\right|^2$ and $a_1, ..., a_m$ are the autoregression constants; i.e., D is in critical distortion region. Given a time-discrete stationary, ergodic Gaussian source, an additional non-negligible description rate about the source besides D-admissible estimates of lossy one-step errors $\{\hat{Z}_k(D)\}$ and a known initial state, is needed to reconstruct the source to within the same distortion D.

This theorem basically says that given an initial state and the D-admissible estimates of lossy one-step errors, there is no way for decoder to generate the D-admissible estimates of the related memory source. We prove this theorem at the end of this chapter.

B. Independence of Initial State

One of the key tools used to prove theorem 3.1 is the independence of our conditional rate-distortion function from a known initial state. We assume hereafter that the initial state of $\{X_t\}$, namely $\mathbf{x}_0 = (x_0, ..., x_{-m+1})$, is known perfectly by all interested parties.

We define the rate-distortion function of $\{X_t\}$ with respect to F_ρ by the prescription

$$R_{\mathbf{X}|\hat{\mathbf{Z}},\ \mathbf{x}_0}(D_{\mathbf{X}}) = \lim_{n\to\infty} R^{(n)}_{\mathbf{X}|\hat{\mathbf{Z}},\ \mathbf{x}_0}(D_{\mathbf{X}}), \tag{1}$$

where[1]

$$R^{(n)}_{\mathbf{X}|\hat{\mathbf{Z}},\ \mathbf{x}_0}(D_{\mathbf{X}}) = \inf_{q\in Q_D} \frac{1}{n} I(\mathbf{X}^n; \hat{\mathbf{X}}^n | \hat{\mathbf{Z}}^n, \mathbf{x}_0). \tag{2}$$

In the present instances the symbols Q_D and $I(\mathbf{X}^n; \hat{\mathbf{X}}^n | \hat{\mathbf{Z}}^n, \mathbf{x}_0)$ in (2) are defined by

$$\begin{aligned} Q_D \;=\; & \{q(\hat{\mathbf{x}}^n|\mathbf{x}^n, \hat{\mathbf{z}}^n, \mathbf{x}_0) : \iiint d\mathbf{x}^n\, d\hat{\mathbf{x}}^n\, d\hat{\mathbf{z}}^n\, p(\mathbf{x}^n, \hat{\mathbf{z}}^n|\mathbf{x}_0) \\ & \cdot q(\hat{\mathbf{x}}^n|\mathbf{x}^n, \hat{\mathbf{z}}^n, \mathbf{x}_0) \cdot d(\mathbf{x}^n, \hat{\mathbf{x}}^n) \leq D_{\mathbf{X}}\}, \end{aligned} \tag{3}$$

where

$$d(\mathbf{x}^n, \hat{\mathbf{x}}^n) = \frac{1}{n}\sum_{t=1}^{n} d(x_t - \hat{x}_t), \tag{4}$$

where

$$d(x_t - \hat{x}_t) = (x_t - \hat{x}_t)^2, \tag{5}$$

and

$$\begin{aligned} I(\mathbf{X}^n; \hat{\mathbf{X}}^n | \hat{\mathbf{Z}}^n, \mathbf{x}_0) \;=\; & \iiint d\mathbf{x}^n\, d\hat{\mathbf{x}}^n\, d\hat{\mathbf{z}}^n\, p(\mathbf{x}^n, \hat{\mathbf{z}}^n|\mathbf{x}_0) \\ & \cdot q(\hat{\mathbf{x}}^n|\mathbf{x}^n, \hat{\mathbf{z}}^n, \mathbf{x}_0) \\ & \cdot \log \frac{q(\hat{\mathbf{x}}^n|\mathbf{x}^n, \hat{\mathbf{z}}^n, \mathbf{x}_0)}{q(\hat{\mathbf{x}}^n|\hat{\mathbf{z}}^n, \mathbf{x}_0)}, \end{aligned} \tag{6}$$

where

$$q(\hat{\mathbf{x}}^n|\hat{\mathbf{z}}^n, \mathbf{x}_0) = \int d\mathbf{x}^n\, p(\mathbf{x}^n|\hat{\mathbf{z}}^n, \mathbf{x}_0) \cdot q(\hat{\mathbf{x}}^n|\mathbf{x}^n, \hat{\mathbf{z}}^n, \mathbf{x}_0). \tag{7}$$

In words, $R^{(n)}_{\mathbf{X}|\hat{\mathbf{Z}},\ \mathbf{x}_0}(D_{\mathbf{X}})$ is the minimum average mutual information per letter given $\hat{\mathbf{Z}} = (\hat{Z}_1, ..., \hat{Z}_n)$ between $\mathbf{X} = (X_1, ..., X_n)$, and $\hat{\mathbf{X}} = (\hat{X}_1, ..., \hat{X}_n)$ sufficient to guarantee fidelity $D_{\mathbf{X}}$ when $\mathbf{x}_0$ is the initial state. The justification for defining $R_{\mathbf{X}|\hat{\mathbf{Z}},\ \mathbf{x}_0}(D_{\mathbf{X}})$ in the above manner resides in coding theorems and techniques discussed in [1], [2], [7] and [13]. Here, however, we are concerned solely with computation of the lower bound of rate $R_{\mathbf{X}|\hat{\mathbf{Z}},\ \mathbf{x}_0}(D_{\mathbf{X}})$.

We show that $R^{(n)}_{\mathbf{X}|\hat{\mathbf{Z}},\ \mathbf{x}_0}(D_{\mathbf{X}})$ does not depend on $\mathbf{x}_0$ despite the fact the the set Q_D of admissible conditional probability densities does.

Theorem 3.2: Let the sequences $\{X_k\}$, $\{\hat{X}_k(D)\}$, $\{Z_k\}$ and $\{\hat{Z}_k(D)\}$ be defined as discussed previously. Given a two-dimensional stationary, ergodic source $X\hat{Z}$ and scalar distortion measures

$$R_{\mathbf{X}\ |\ \hat{\mathbf{Z}},\ \mathbf{x}_0}(D_{\mathbf{X}}) = R_{\mathbf{X}\ |\ \hat{\mathbf{Z}}}(D_{\mathbf{X}}), \tag{8}$$

where $\mathbf{x}_0 = (x_0, ..., x_{-m+1})$, known perfectly by all interested parties. In other words, the rate $R_{\mathbf{X}\ |\ \hat{\mathbf{Z}}}(D_{\mathbf{X}})$ does not depend

[1]Note that $R^{(n)}_{\mathbf{X}|\hat{\mathbf{Z}},\ \mathbf{x}_0}(D_{\mathbf{X}})$ as defined here for autoregressive sources usually *increases* with n for fixed $D_{\mathbf{X}}$, reflecting the diminishing value of the knowledge of the initial state. Accordingly, $R^{(n)}_{\mathbf{X}|\hat{\mathbf{Z}},\ \mathbf{x}_0}(D_{\mathbf{X}})$ for finite n does *not* serve as an upper bound to $R_{\mathbf{X}|\hat{\mathbf{Z}},\ \mathbf{x}_0}(D_{\mathbf{X}})$. This contrasts with the stationary case in which only statistical knowledge of the initial state was assumed, and the $R^{(n)}_{\mathbf{X}|\hat{\mathbf{Z}},\ \mathbf{x}_0}(D_{\mathbf{X}})$ curves always converged to $R_{\mathbf{X}|\hat{\mathbf{Z}},\ \mathbf{x}_0}(D_{\mathbf{X}})$ from above.

on initial points $\mathbf{x}_0$ *despite the fact that the set* Q_D *of admissible conditional probability densities does, where*

$$Q_D = \{q(\cdot|\cdot\cdot) \quad : \quad \iiint d\mathbf{x}^n \, d\hat{\mathbf{x}}^n \, d\hat{\mathbf{z}}^n \, p\left(\mathbf{x}^n \hat{\mathbf{z}}^n | \mathbf{x}_0\right) \cdot q\left(\hat{\mathbf{x}}^n | \mathbf{x}^n \hat{\mathbf{z}}^n \mathbf{x}_0\right) \cdot d(\mathbf{x}^n, \hat{\mathbf{x}}^n) \leq D_{\mathbf{X}}\} \quad (9)$$

We omit the proof for brevity. The significance of this result is that the knowledge of the initial state doesn't affect on R_2 and that when the rate R_2 is evaluated, the starting point doesn't matter.

C. The Lower Bound

To prove the theorem 3.1, we state the lower bound of conditional rate-distortion function and its necessary condition for equality.

Theorem 3.3: Let the sequences $\{X_k\}$, $\{\hat{X}_k(D)\}$, $\{Z_k\}$ *and* $\{\hat{Z}_k(D)\}$ *be defined as discussed previously. Given a two-dimensional stationary, ergodic source* $X\hat{Z}$ *and the squared-error distortion measure*

$$R_{\mathbf{X} \mid \hat{\mathbf{Z}}}(D_{\mathbf{X}}) \geq \lim_{n\to\infty} \frac{1}{n}\left[h(\mathbf{X}^n|\hat{\mathbf{Z}}^n) - h(g_s)\right] \quad (10)$$

where $h(g_s) = \frac{n}{2}\log(2\pi e D_{\mathbf{X}})$*. The necessary condition for the equality is that a probability density* $q(\cdot|\cdot)$ *can be found that satisfies*

$$p(\mathbf{x}^n|\hat{\mathbf{z}}^n) = \int d\hat{\mathbf{x}}^n q(\hat{\mathbf{x}}^n|\hat{\mathbf{z}}^n) g_s(\mathbf{x}^n - \hat{\mathbf{x}}^n) \quad (11)$$

Proof: We define $R^{(n)}_{\mathbf{X}|\hat{\mathbf{Z}}}(D_{\mathbf{X}})$ which is the MSE rate-distortion function of a hypothetical source which generates a sequence of i.i.d. n-dimensional random vectors, each having probability density $p^{(n)}_{\mathbf{X}|\hat{\mathbf{Z}}}(\mathbf{x}^n)$, which is the probability density governing the first n outputs of the source, i.e., $\mathbf{X}^n = [X_1, ..., X_n]^T$. The Shannon lower bound of $R^{(n)}_{\mathbf{X}|\hat{\mathbf{Z}}}(D_{\mathbf{X}})$ will be denoted by $R^{(n)}_{\mathbf{X}|\hat{\mathbf{Z}},L}(D_{\mathbf{X}})$.

The MSE Shannon lower bound of $R^{(n)}_{\mathbf{X}|\hat{\mathbf{Z}}}(D_{\mathbf{X}})$, $R^{(n)}_{\mathbf{X}|\hat{\mathbf{Z}},L}(D_{\mathbf{X}})$, can be expressed parametrically as

$$R^{(n)}_{\mathbf{X}|\hat{\mathbf{Z}},L}(D_s) = \frac{1}{n}\left[h(p(\mathbf{x}^n|\hat{\mathbf{z}}^n)) - h(g_s(\mathbf{x}^n))\right] \quad (12)$$

$$D_s = \int d\mathbf{x}^n \rho_n(\mathbf{x}^n) g_s(\mathbf{x}^n), \quad (13)$$

where $\rho_n(\mathbf{x}^n)$ is the squared error per component distortion measure, i.e.,

$$\rho_n(\mathbf{x}^n) \triangleq \frac{1}{n}(\mathbf{x}^n)^T\mathbf{x}^n, \quad (14)$$

and $g_s(\mathbf{x}^n)$ is the probability density given by

$$g_s(\mathbf{x}^n) \triangleq \frac{e^{s\rho_n(\mathbf{x}^n)}}{\int d\mathbf{z}^n e^{s\rho_n(\mathbf{z}^n)}} \quad (15)$$

The n-dimensional version of Theorem 4.3.1 of [2] gives a necessary and sufficient condition for equality in the Shannon lower bound for any difference distortion measure $\rho(\cdot)$, and will be useful in our analysis. It says that given any $s \leq 0$, the Shannon lower bound is tight if and only if a probability density $q(\cdot|\cdot)$ can be found that satisfies

$$p(\mathbf{x}^n|\hat{\mathbf{z}}^n) = \int d\hat{\mathbf{x}}^n q(\hat{\mathbf{x}}^n|\hat{\mathbf{z}}^n) g_s(\mathbf{x}^n - \hat{\mathbf{x}}^n) \quad (16)$$

By Theorem 4.6.1 of [2], we have

$$R_{\mathbf{X}|\hat{\mathbf{Z}}}(D_s) \geq R^{(n)}_{\mathbf{X}|\hat{\mathbf{Z}}}(D_s) + h - \frac{1}{n}h(p(\mathbf{x}^n|\hat{\mathbf{z}}^n)), \quad (17)$$

where $h \triangleq \lim_{n\to\infty} \frac{1}{n}h(p_{\mathbf{X}^n|\hat{\mathbf{Z}}^n}) = \lim_{n\to\infty} \frac{1}{n}h(\mathbf{X}^n|\hat{\mathbf{Z}}^n)$.

It follows from (12) that

$$R_{\mathbf{X}|\hat{\mathbf{Z}}}(D_s) \geq h - \frac{1}{n}h(g_s(\mathbf{x}^n)) \quad (18)$$

It is clear that a necessary (although by no means sufficient) condition for equality is that equality must hold in (12); that is, $R^{(n)}_{\mathbf{X}|\hat{\mathbf{Z}}}(D_{\mathbf{X}})$ must equal its ordinary Shannon lower bound at the value of $D_{\mathbf{X}}$. Otherwise, the lower bound of (18) would be strictly less than the lower bound of (17) and therefore could not possibly be tight.

Now we evaluate the distortion, D_s and the differential entropy of $g_s(\mathbf{x}^n)$, $h(g_s(\mathbf{x}^n))$. We have

$$g_{s,1}(x_k) = \frac{e^{s\rho_1(x_k)}}{\int dz e^{s\rho_1(z)}}, \quad (19)$$

where $\rho_1(x_k) = x_k^2$. Note that $g_{s,1}(x_k)$ has the form of a 1-dimensional density of i.i.d. Gaussian random (scalar) variable, with mean zero and variance $-\frac{1}{2s}$. Thus we have

$$g_{s,1}(x_k) = \frac{1}{\sqrt{\frac{\pi}{-s}}}\exp(sx_k^2). \quad (20)$$

Likewise, we have

$$g_s(\mathbf{x}^n) = \frac{1}{\left(\frac{\pi n}{-s}\right)^{\frac{n}{2}}}\exp\left(\frac{s}{n}\left(\mathbf{x}^n\right)^T\mathbf{x}^n\right). \quad (21)$$

By the definition of differential entropy, we have

$$h(g_s(\mathbf{x}^n))) = \frac{n}{2}\log\left(2\pi e \cdot \frac{n}{-2s}\right). \quad (22)$$

We can now express $h(g_s(\mathbf{x}^n)))$ in terms of D_s by the following calculation:

$$\begin{aligned} D_s &= \int d\mathbf{x}^n \rho_n\left(\mathbf{x}^n\right) g_s(\mathbf{x}^n) && (23)\\ &= \frac{1}{n}\left(\frac{\pi n}{-s}\right)^{-\frac{n}{2}} \int \left(\mathbf{x}^n\right)^T \mathbf{x}^n \exp\left(\frac{s}{n}\left(\mathbf{x}^n\right)^T\mathbf{x}^n\right) d\mathbf{x}^n \\ &= \frac{1}{n}\left(\frac{\pi n}{-s}\right)^{-\frac{n}{2}} \int \sum_{k=1}^{n} x_k^2 \prod_{j=1}^{n} e^{\frac{s}{n}x_j^2} dx_1 \cdots dx_n, && (24) \end{aligned}$$

where x_k denotes the kth scalar component of $\mathbf{x}^n$. This then yields

$$\begin{aligned} D_s &= \frac{1}{n}\left(\frac{\pi n}{-s}\right)^{-\frac{n}{2}} n \int x_k^2 \prod_{j=1}^{n} e^{\frac{s}{n}x_j^2} dx_1 \cdots dx_n && (25)\\ &= \frac{1}{n}\left(\frac{\pi n}{-s}\right)^{-\frac{n}{2}} n \cdot \frac{n}{-2s}\left(\frac{\pi n}{-s}\right)^{\frac{n}{2}} && (26)\\ &= \frac{n}{-2s}. && (27) \end{aligned}$$

Thus, from eq. (22) we have

$$h(g_s(\mathbf{x}^n))) = \frac{n}{2}\log(2\pi e D_s). \tag{28}$$

The conclusion follows. ■

In terms of characteristic functions the requirement for the necessary condition is that the Fourier transform of $q(\cdot|\cdot)$ must be a multivariate characteristic function. In general it is very difficult to determine conditions under which the function is a multivariate characteristic function. However, this can be done in the case of the squared-error distortion measure.

D. The Proof of Theorem 3.1

This is a proof via the lower bound of conditional rate-distortion function and the independence of a known initial state.

Proof: To prove Theorem 3.1, first note that

$$\begin{aligned}
R_2 &= R_{\mathbf{X}\ |\ \hat{\mathbf{z}},\ \mathbf{x}_0}(D_{\mathbf{X}}) && (29)\\
&= R_{\mathbf{X}\ |\ \hat{\mathbf{z}}}(D_{\mathbf{X}}) && (30)\\
&\geq \lim_{n\to\infty}\frac{1}{n}\left[h(\mathbf{X}^n|\hat{\mathbf{Z}}^n) - h(g_s)\right] && (31)\\
&= \lim_{n\to\infty}\frac{1}{n}\left[h(\mathbf{X}^n|\hat{\mathbf{Z}}^n) - \frac{n}{2}\log(2\pi e D_{\mathbf{X}})\right] && (32)\\
&= \lim_{n\to\infty}\frac{1}{n}h(\mathbf{X}^n,\ \hat{\mathbf{Z}}^n) - \lim_{n\to\infty}\frac{1}{n}h(\hat{\mathbf{Z}}^n)\\
&\quad -\frac{1}{2}\log(2\pi e D_{\mathbf{X}}) && (33)\\
&= \lim_{n\to\infty}\frac{1}{2}\log\left[(2\pi e)^2\frac{(\sigma^2 - D_{\mathbf{X}})D_{\mathbf{X}}}{\sigma^2}|\Phi_n|^{\frac{1}{n}}\right]\\
&\quad -\frac{1}{2}\log(2\pi e)(\sigma^2 - D_{\mathbf{X}}) - \frac{1}{2}\log(2\pi e)D_{\mathbf{X}} && (34)\\
&= \lim_{n\to\infty}\frac{1}{2}\log|\Phi_n|^{\frac{1}{n}} - \frac{1}{2}\log\sigma^2 && (35)\\
&= \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\frac{1}{2}\log\lambda_k^{(n)} - \frac{1}{2}\log\sigma^2 && (36)\\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{2}\log\Phi(\omega)\,d\omega - \frac{1}{2}\log\sigma^2 && (37)\\
&= \frac{1}{2}\log\sigma^2 - \frac{1}{2}\log\sigma^2 = 0, && (38)
\end{aligned}$$

where

(30) follows from Theorem 3.2,
(31) and (32) from Theorem 3.3,
(33) follows from the definition of mutual information,
(34) from the definition of differential entropy,
(36) follows from the eigenvalue decomposition theorem,
(37) from Toeplitz Theorem [2],
(38) from the fact that here the entropy rate power is the one-step prediction error, σ^2,

To show that equality of (30) and (31) doesn't hold, we check the necessary condition by asking if a probability density $q(\cdot|\cdot)$ can be found that satisfies

$$p(\mathbf{x}^n|\hat{\mathbf{z}}^n) = \int d\hat{\mathbf{x}}^n q(\hat{\mathbf{x}}^n|\hat{\mathbf{z}}^n)g_s(\mathbf{x}^n - \hat{\mathbf{x}}^n) \tag{39}$$

Now denote by $Q_{\mathbf{X}|\hat{\mathbf{Z}}}(\omega^{\mathbf{n}})$ and $Q_{\hat{\mathbf{X}}|\hat{\mathbf{Z}}}(\mu^{\mathbf{n}})$ the characteristic functions associated with the random vectors $\mathbf{X}^n$ and $\hat{\mathbf{X}}^n$ given $\hat{\mathbf{Z}}^n = [Z_1,\cdots Z_n]^T$ respectively, and denote by $G_s(\omega^n)$ the characteristic function associated with the joint probability density $g_s(\cdot)$. Then theorem 3.3 implies that a necessary condition for equality of (30) and (31) is that the function $\frac{Q_{\hat{\mathbf{X}}|\hat{\mathbf{Z}}}(\omega^{\mathbf{n}})}{G_s(\omega^n)}$ be itself a characteristic function.

Since $g_s(\mathbf{x}^n)$ has the form of the density of n i.i.d. 1-dimensional random vectors, we have

$$G_s(\omega^n) = \prod_{k=1}^{n} G_{s,1}(\omega_k), \tag{40}$$

where ω_k is the kth variable comprising ω^n, and $G_{s,1}$ is the characteristic function associated with (19). Thus we have

$$\begin{aligned}
G_{s,1}(\omega_k) &= E_{g_{s,1}}\left[\exp\left(-iX_k\omega_k\right)\right] && (41)\\
&= \exp\left\{-\frac{1}{2}\left(\frac{1}{-2s}\right)\omega_k^2\right\}. && (42)
\end{aligned}$$

Likewise, we have

$$G_s(\omega^n) = \exp\left\{-\frac{1}{2}\left(\frac{n}{-2s}\right)(\omega^n)^T\omega^n\right\}. \tag{43}$$

By (27), we can express $G_s(\omega^n)$ in terms of D_s.

$$\begin{aligned}
G_s(\omega^n) &= \exp\left\{-\frac{1}{2}D_s(\omega^n)^T\omega^n\right\} && (44)\\
&= \prod_{k=1}^{n}\exp\left\{-\frac{1}{2}D_s\omega_k^2\right\}. && (45)
\end{aligned}$$

Since $p(\mathbf{x}^n|\hat{\mathbf{z}}^n)$ also has the form of the density of n-dimensional Gaussian vectors, with mean $A^{-1}\hat{\mathbf{z}}^n$ and variance $\left|\frac{D_s}{\sigma^2}\Phi_n\right|$, where $\Phi_n = E\left[\mathbf{X}\mathbf{X}^T\right] = \sigma^2\left(A^TA\right)^{-1}$, we have

$$p(\mathbf{x}^n|\hat{\mathbf{z}}^n) = \frac{1}{(2\pi)^{\frac{n}{2}}\left|\frac{D_s}{\sigma^2}\Phi_n\right|^{\frac{1}{2}}}\cdot$$

$$\exp\left[-\frac{1}{2}\left(\mathbf{x}^n - A^{-1}\hat{\mathbf{z}}^n\right)^T\frac{\sigma^2}{D_s}\Phi_n^{-1}\left(\mathbf{x}^n - A^{-1}\hat{\mathbf{z}}^n\right)\right]. \tag{46}$$

By the transformation via $A\mathbf{X} = \mathbf{Z}$, we have

$$p(\mathbf{z}^n|\hat{\mathbf{z}}^n) = \prod_{k=1}^{n}\frac{1}{\sqrt{2\pi D_s}}\exp\left\{-\frac{1}{2}\frac{(z_k - \hat{z}_k)^2}{D_s}\right\}. \tag{47}$$

From the fact that $A\mathbf{X} = \mathbf{Z}$, we can write

$$\begin{aligned}
Q_{\mathbf{X}|\hat{\mathbf{Z}}}(\omega^n) &= E_{\mathbf{X}|\hat{\mathbf{Z}}}\left[\exp\left\{-i(\mathbf{X}^n)^T\omega^n\right\}\right] && (48)\\
&= E_{\mathbf{Z}|\hat{\mathbf{Z}}}\left[\exp\left\{-i(\mathbf{Z}^n)^T\mu^n\right\}\right] && (49)\\
&= Q_{\mathbf{Z}|\hat{\mathbf{Z}}}(\mu^{\mathbf{n}}) && (50)\\
&= \prod_{k=1}^{n}\exp\left(-\frac{1}{2}D_s\mu_k^2 + i\hat{z}_k\mu_k\right) && (51)\\
&= \prod_{k=1}^{n}\exp\left(-\frac{1}{2}D_s\mu_k^2\right)M(\mu_k) && (52)\\
&= Q_{\mathbf{X}|\hat{\mathbf{Z}}}(\mu^n), && (53)
\end{aligned}$$

where we define

$$\omega \triangleq A^T \mu, \tag{54}$$

and

$$M(\mu_k) \triangleq \exp(i\hat{z}_k \mu_k). \tag{55}$$

Now we evaluate the following function, $Q_{\hat{\mathbf{X}}|\hat{\mathbf{Z}}}(\mu^n)$ to check whether or not it is a Gaussian multivariate characteristic function;

$$\begin{aligned} Q_{\hat{\mathbf{X}}|\hat{\mathbf{Z}}}(\mu^n) &= \frac{Q_{\mathbf{X}|\hat{\mathbf{Z}}}(\mu^n)}{G_s(\omega^n)} & (56) \\ &= \prod_{k=1}^{n} \exp\left(-\frac{1}{2}D_s\mu_k^2 + \frac{1}{2}D_s\omega_k^2\right) M(\mu_k) \\ &= \exp\left[-\mu^T\left(\frac{D_sI_n - D_sAA^T}{2}\right)\mu\right] \\ &\quad \cdot \prod_{k=1}^{n} M(\mu_k), & (57) \end{aligned}$$

where $\prod_{k=1}^{n} M(\mu_k)$ is a multivariate characteristic function.

Since for all $n > 0$, $D_s\left(1 - \frac{\sigma^2}{\min_k \lambda_k{}^{(n)}}\right) < 0$ and when $n \to \infty$, $D_s\left(1 - \sigma^2 \cdot ess\sup g(\omega)\right) < 0$ by the theorem of the asymptotic equal distribution of eigenvalue of Toeplitz matrices [2], [9], where

$$g(\omega) = \sum_{k=-\infty}^{\infty} g_k e^{-i\omega} \tag{58}$$

and

$$g_k = \frac{1}{\sigma^2}\sum_{l=0}^{m} a_l a_{l+k}, \tag{59}$$

$D_sI_n - D_sAA^T$ has at least one negative eigenvalue and isn't a correlation matrix for all D_s in the critical distortion region. Hence $\exp\left[-\mu^T\left(\frac{D_sI_n - D_sAA^T}{2}\right)\mu\right]$ isn't a Gaussian multivariate characteristic function and 0 is not a tight lower bound of R_2. The conclusion follows. ∎

IV. Conclusion

We studied the problem whether or not the D-admissible estimates of lossy one step prediction errors can be used to reconstruct the original memory source to within the same distortion. We considered the following scenario. An optimum encoder computes the memoryless sequence from the memory sequence and then uses a code of rate $\frac{1}{2}\log\left(\frac{Q_0}{D}\right)$ to convey the memoryless sequence to the decoder with an MSE of distortion D. There is a receiver which tries to use these D-admissible estimates of lossy one-step prediction errors with a known initial state perfectly to generate D-admissible estimates of the memory sequence.

We focus primarily on a time-discrete stationary, ergodic Gaussian model for the source that is to be compressed with side information of D-admissible estimates of lossy one-step prediction errors. The essential characteristic of the model is that the source and the side information are jointly Gaussian stationary-ergodic pair of time-discrete random process with joint per-letter alphabet. Specifically, the sequence of successive one-step prediction errors, also called the *innovations process*, is stationary, zero-mean, uncorrelated, and Gaussian.

The central theme of our results is that if the decoder is provided with a distorted version of the innovations process and a known initial state, it needs an additional non-negligible description rate about the source if it wants to reconstruct the source to within the same distortion. Moreover, knowing the initial state does not help since we show our conditional rate-distortion function is independent of initial state. These results offer further insights into the intriguing and confounding relation between the rate-distortion function of memory source and that of its innovations process.

Acknowledgment

The authors would like to thank Jun Chen, Yegnaswamy Sermadevi and Prapun Suksompong of Cornell University for informative discussions and valuable comments.

References

[1] T. Berger, "Information rates of Wiener processes," *IEEE Trans. Information Theory*, vol. IT-16, pp. 134-138, March 1970

[2] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, N.J.: Prentice-Hall, 1971

[3] T. Berger, J. Gibson, "Lossy Source Coding," *IEEE Trans. Information Theory*, vol. IT-44, No. 6, pp. 2693-2723, October 1998

[4] T. M. Cover and J.A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991

[5] R. M. Fano, *Transmission of Information*, New York: Wiley, 1961

[6] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968

[7] R. M. Gray, "Information rates of autoregressive processes," Ph.D. dissetation, Elec. Eng., Univ. South. Calif., Los Angeles, CA, 1969

[8] R. M. Gray, "Information rates of autoregressive processes," *IEEE Trans. Information Theory*, vol. IT-16, pp. 412-421, July 1970

[9] R. M. Gray, "On the Asymptotic Eigenvalue Distribution of Toeplitz Matrices," *IEEE Trans. Information Theory*, vol. IT-18, pp. 725-730, November 1972

[10] R. M. Gray, "A New Class of Lower Bounds to Information Rates of Stationary Sources via Conditional Rate-Distortion Functions," *IEEE Trans. Information Theory*, vol. IT-19, No. 4, pp. 480-489, July 1973

[11] U. Grenander and G. Szego, *Toeplitz Forms and Their Applications*. University of California Press, Berkeley and Los Angeles, California, 1958

[12] P. R. Halmos, *Introduction to Hilbert Space*, Chelsea, New York, 1951

[13] B. M. Leiner, "Rate-distortion Theory for Sources with Side Information," Ph.D. dissertaion, Stanford University, Stanford, California, august, 1973

[14] B. M. Leiner and R. M. Gray, "Rate-Distortion Theory for Ergodic Sources with Side Information," *IEEE Trans. Information Theory*, vol. IT-20, No. 5, pp. 672-675, September 1974

[15] B. McMillan, "The Basic Theorems of Information Theory," *Ann. Math. Stat.*, Vol. 24, 196-219, 1953

[16] C. E. Shannon and W. Weaver, "The Mathematical Theory of Communication," *Bell System Technology Journal*, vol. 27, 379-423, 623-656, 1948. (Also in book from with postscript by W. Weaver, University of Illinois Press, Urbana, Ill., 1949.)

[17] C. E. Shannon, "Coding Theorems for a Discrete Source with a Fidelity Criterion," *IRE Nat'l. Conv. Rec., part 4*, pp. 142-163, 1959. (Also in *Information and Decision Processes*. R. E. Machol, ed. McGraw-Hill, Inc., New York, 1960, 93-126.)

[18] A. D. Wyner and J. Ziv, "Bounds on the Rate-distortion Function for Stationary Sources with Memory," *IEEE Trans. Information Theory*, vol. IT-17, No. 5, pp. 508 - 513, September 1971