The Degree of Suboptimality of Sending a Lossy Version of the Innovations Process in Gauss-Markov Rate-Distortion

Kwang Taik Kim
Department of Electrical and Computer Engineering
Cornell University
Ithaca, NY 14853, USA
Email: kk274@cornell.edu

Toby Berger
Department of Electrical and Computer Engineering
Cornell University
Ithaca, NY 14853, USA
Email: berger@ece.cornell.edu

***Abstract*— In critical distortion range, the MSE rate-distortion function of time-discrete stationary Gaussian first-order autoregressive source is equal to that of related time-discrete i.i.d. Gaussian source. For $0 < D \leq D_c$ it is necessary to provide additional information of non-negligible positive rate in order to obtain a D-admissible code for the original source via the R-D coding of the innovations process and additional post-processing at the decoder. In this scenario, we provide an explicit expression of additional description rate about the original source to find the degree of suboptimality of sending a lossy version of the innovations process in Gauss-Markov rate-distortion. It is shown that additional description rate is monotone increasing on $\alpha \in [0, 1)$ and is constant on all critical distortion range.**

I. Introduction

The rate-distortion (R-D) function of time-discrete stationary Gaussian first-order autoregressive source with a mean-square-error (MSE) fidelity criterion is equal to that of independent-letter identically distributed generating sequence with the same distortion measure in critical distortion range [1], [3], [4]. Even though there is no obvious intuitive reason why two R-D functions are the same in non-zero region because the relation between the average distortion of the memory source sequence and that of the related memoryless source sequence is unclear, this suggests that perhaps an optimum encoder should compute the related time-discrete independently and identically distributed (i.i.d.) Gaussian sequence from the given time-discrete stationary Gaussian first-order autoregressive source and then use a code of rate $R(D)$ to convey the memoryless sequence to the decoder with an MSE of D [1], [2], [4]. In this scenario, we [7] showed that for $0 < D \leq D_c$ a D-admissible code for the original source can't 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. Our previous result, however has had limitations in computing the additional description rate and in giving explanations of how memory property of the original source affects it.

The main result of this paper is an explicit characterization of the additional description rate as a function of the autoregression constant and the average distortion in terms of the Clausen's integral [8] by applying Grenander and Szegö's theorems on the asymptotic behavior of Toeplitz forms [5]. It is further shown that the additional description rate is monotone increasing on the autoregression constant $\alpha \in [0, 1)$, and doesn't depend on average distortion despite appearances.

Two methods will be used to evaluate the additional description rate. The first method is to apply the well-known fact that the optimum minimum mean-squared error (MMSE) estimator for jointly Gaussian process is the linear minimum mean-squared error (LMMSE) estimator, to the MMSE estimate of the original source sequence given a lossy version of its innovations process. The second method is to apply theorems on the asymptotic eigenvalue behavior of Toeplitz and approximately Toeplitz forms to the autocorrelation matrix of the error process under a lossy version of the innovations process obtained by the LMMSE estimator.

We provide the necessary results on the MMSE estimator and some other preliminaries in Section III. We prove our main result in Section IV. The next section contains a precise formulation of the problem and the statement of our main result, Theorem 2.2.

II. Problem Formulation and Main Result

A Gauss-Markov source $\{X_k, k = 0, 1, 2...\}$ is described by the equation

$$X_k = \alpha X_{k-1} + Z_k,$$

where $|\alpha| < 1$ is the autoregression constant, $\{Z_k\}$ is a sequence of i.i.d. r.v.'s, and X_r and Z_s are statistically independent if $s > r$. Let σ_Z^2 denote the variance of the Z_k, and assume for simplicity of presentation that the Z_k's have mean zero. Then it is easy to show that the variance of X_k, call it σ_k^2, satisfies the difference equation

$$\begin{aligned} E\left[X_k^2\right] &= \alpha^2 E\left[X_{k-1}^2\right] + \sigma_Z^2 \\ \sigma_k^2 &= \alpha^2 \sigma_{k-1}^2 + \sigma_Z^2. \end{aligned}$$

Since $|\alpha| < 1$, σ_k^2 converges to

$$\sigma_\infty^2 = \frac{\sigma_Z^2}{1 - \alpha^2}, \tag{1}$$

1-4244-0504-1/06/$20.00

and

$$\sigma_k^2 = \frac{1-\alpha^{2k}}{1-\alpha^2} \cdot \sigma_Z^2 = \left(1-\alpha^{2k}\right)\sigma_\infty^2. \tag{2}$$

This choice of σ_Z^2 makes a stationary Gauss-Markov random sequence. For $\alpha > 0$, all correlations are positive; for $\alpha < 0$, they oscillate in sign. Since the probability distribution of $\{X_k\}$ is Gaussian, we have

$$E\left(X_k|X_{-\infty}^{k-1}\right) = cX_{k-1}$$

and

$$\begin{aligned} E\left[X_k - E\left(X_k|X_{-\infty}^{k-1}\right)\right]^2 &= E\left(X_k - cX_{k-1}\right)^2 \\ &= \sigma_\infty^2\left(1-\alpha^{2k-2}\right) \\ &\quad \cdot \left\{(c-\alpha)^2 + \frac{1-\alpha^2}{1-\alpha^{2k-2}}\right\}, \end{aligned}$$

which is minimized for $c = \alpha$. So,

$$E\left(X_k|X_{-\infty}^{k-1}\right) = \alpha X_{k-1}$$

is the optimal predictor and its error is

$$\sigma_\infty^2(1-\alpha^2) = \sigma_Z^2 \triangleq Q_0.$$

Consider next the correlation structure of $\{X_k\}$, namely

$$E\left[X_k X_{k+n}\right] = E\left[X_k\left(\alpha X_{k+n-1} + Z_{k+n}\right)\right] \tag{3}$$

For $n > 0$, we have

$$E\left[X_k Z_{k+n}\right] = 0,$$

so solving (3) yields

$$E\left[X_k X_{k+n}\right] = \alpha^n E\left[X_k^2\right] = \alpha^n \sigma_k^2. \tag{4}$$

Since we chose the autoregression constant $|\alpha| < 1$ so that $\sigma_k^2 \to \sigma_\infty^2$ as given by (1), an asymptotically stationary correlation structure emerges given by

$$\phi_n \triangleq \lim_{k\to\infty} E\left[X_k X_{k+n}\right] = \sigma_\infty^2 \alpha^{|n|},$$

and a Gauss-Markov source is characterized by an exponentially decaying memory of the form

$$\phi_k = \sigma_\infty^2 |\alpha|^k, \qquad -1 < \alpha < 1.$$

We have

$$\Phi_X(\omega) = \sigma_\infty^2 \sum_{k=-\infty}^{\infty} |\alpha|^{|k|} e^{-i\omega k} = \sigma_\infty^2 \cdot \frac{1-|\alpha|^2}{1-2|\alpha|\cos\omega + |\alpha|^2}.$$

For $-\pi \le \omega \le \pi$, the minimum of $\Phi_X(\omega)$ is at $\omega = \pm\pi$ because $\cos\omega = -1$ there, and this minimum is

$$ess \inf \Phi_X(\omega) = \sigma_\infty^2 \cdot \frac{1-|\alpha|}{1+|\alpha|} = \frac{\sigma_Z^2}{(1+|\alpha|)^2}.$$

This implies

$$D_c = \frac{\sigma_Z^2}{(1+|\alpha|)^2},$$

which is strictly less than $Q_0 = \sigma_Z^2$.

Definition 2.1: The Clausen's integral *is defined to be*

$$\mathrm{Cl}_2(\theta) = -\int_0^\theta \log\left[2\sin\frac{1}{2}\theta\right] d\theta.$$

Let $\{\hat{Z}_k\}$ be a sequence of optimal D-predictor of $\{Z_k\}$ for $D \le D_c$. Then

$$Z_k = \hat{Z}_k + V_k,$$

where $V_k \sim N(0, D)$ and $V_k \perp \hat{Z}_k$; i.e. $\{V_k\}$ is a white Gaussian random sequence. Given $X_0, \hat{Z}_1, \hat{Z}_2, \ldots$, we find the MMSE estimate of X_n, call it $\hat{X}_n$.

We are now in a position to state our main result.

Theorem 2.2: For Gauss-Markov rate-distortion,

$$\begin{aligned} f(\alpha, D) = \frac{x_\theta}{2\pi}\log\frac{D}{\theta} + \frac{1}{\pi}\Big[& y_\theta \log|\alpha| + \frac{1}{2}\mathrm{Cl}_2(2x_\theta) \\ & + \frac{1}{2}\mathrm{Cl}_2(2y_\theta) - \frac{1}{2}\mathrm{Cl}_2(2x_\theta + 2y_\theta)\Big] \end{aligned} \tag{5}$$

is the additional description rate about the original source to get a MSE of D between $\{\hat{X}_k\}$ and $\{X_k\}$, where

$$x_\theta = \cos^{-1}\left\{\frac{\theta\left(1+\alpha^2\right) - D}{2|\alpha|\theta}\right\}, \qquad 0 \le x_\theta \le \pi,$$

$$y_\theta = \tan^{-1}\left(\frac{\sqrt{2\theta\frac{1+\alpha^2}{1-\alpha^2} - 1 - \theta^2}}{1+\theta}\right), \qquad 0 \le y_\theta \le \pi,$$

and θ can be obtained by solving

$$D = \frac{\theta x_\theta}{\pi} + \frac{D}{1-\alpha^2} - \frac{2D}{\pi(1-\alpha^2)}\tan^{-1}\left\{\left(1-\alpha^2\right)\tan\frac{x_\theta}{2}\right\}.$$

The proof is deferred to Section IV. In the meantime, we establish some ancillary results.

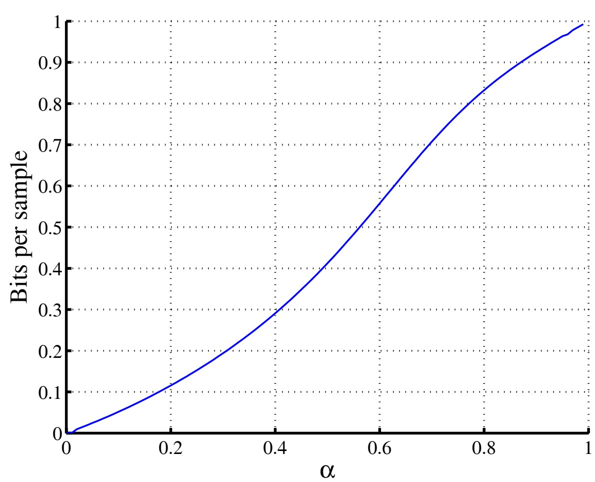

Fig. 1. The additional description rate $f(\alpha, D)$ is monotone increasing on $\alpha \in [0, 1)$. f doesn't depend on D despite appearances.

Corollary 2.3: If f is the additional description rate defined in (5), then f is monotone increasing on $\alpha \in [0,1)$, and is constant on $D \in (0, D_c]$.

The proof follows the chain rule and the second derivative test, and is omitted for brevity.

III. Background and Preliminaries

We can set $X_0 = 0$ w.l.o.g. and then the Gauss-Markov source is fully stationary without any transient period. First, we find the MMSE estimate of X_n and the MMSE of X_n, given $\{\hat{Z}_k, k \geq 1\}$. Since the joint probability distribution of two processes is jointly Gaussian, the optimum estimator for MMSE is LMMSE estimator.

Lemma 3.1: $\hat{X}_n = \sum_{i,j \geq 1} C_{in} R_{ij}^{-1} \hat{Z}_j$,

where

$$R_{ij} = \delta_{ij} \left\{ \sigma_\infty^2 \left(1-\alpha^2\right) - D \right\},$$

and

$$C_{ij} = \begin{cases} 0 & \text{if } i > j \text{ for all } i,j \geq 1, \\ \alpha^{j-i} \left\{ \sigma_\infty^2 \left(1-\alpha^2\right) - D \right\} & \text{otherwise}. \end{cases}$$

Proof: For $\hat{X}_n = \sum_{k \geq 1} A_{kn} \hat{Z}_k$, the MMSE estimate of X_n satisfies

$$E\left[\left(X_n - \hat{X}_n\right) \hat{Z}_k\right] = E\left[\left(X_n - \sum_{j \geq 1} A_{jn} \hat{Z}_j\right) \hat{Z}_k\right] = 0$$

for all $k \geq 1$. Let $C_{kn} = E\left[X_n \hat{Z}_k\right]$ and $R_{jk} = E\left[\hat{Z}_j \hat{Z}_k\right]$ for convenience and we have

$$\begin{aligned} E\left[X_n \hat{Z}_k\right] &= \sum_{j \geq 1} A_{jn} E\left[\hat{Z}_j \hat{Z}_k\right] \\ C_{kn} &= \sum_{j \geq 1} A_{jn} R_{jk} = \left(\mathbf{a}_n^T \mathbf{R}\right)_k \end{aligned}$$

It follows that

$$\mathbf{a}_n = \mathbf{R}^{-T} \mathbf{c}_n$$

is the best predictor and

$$\hat{X}_n = \sum_{j,k} C_{jn} R_{jk}^{-1} \hat{Z}_k.$$

We have

$$R_{ij} = E\left(Z_i - V_i\right)\left(Z_j - V_j\right)$$

But both $\{Z_i\}$ and $\{V_i\}$ are i.i.d. Specifically,

$$E\left[Z_i Z_j\right] = \sigma_Z^2 \delta_{ij} = \sigma_\infty^2 (1-\alpha^2) \delta_{ij}$$

and

$$E\left[V_i V_j\right] = D \delta_{ij}$$

and

$$\begin{aligned} E\left[V_i Z_j\right] &= \delta_{ij} E\left[V_i Z_i\right] \\ &= \delta_{ij} E\left[V_i \left(V_i + \hat{Z}_i\right)\right] \\ &= \delta_{ij} E\left[V_i^2\right] + \delta_{ij} \cdot 0 \qquad (6) \\ &= \delta_{ij} D, \qquad (7) \end{aligned}$$

where (6) follows from the fact that $V_i \perp \hat{Z}_i$ and δ_{ij} is the Kronecker delta. Thus

$$R_{ij} = \delta_{ij} \left[\sigma_\infty^2 \left(1-\alpha^2\right) - D\right].$$

If $i > j$, then

$$C_{ij} = E\left[X_j \left(Z_i - V_i\right)\right] = 0,$$

because X_j is statistically independent of both Z_i and V_i. If $i \leq j$, then it is equivalent to prove for $k = 0, 1, 2, ..., j-1$,

$$\begin{aligned} C_{j-k,j} &= E\left[X_j \left(Z_{j-k} - V_{j-k}\right)\right] \\ &= E\left[\left(\alpha X_{j-1} + Z_j\right)\left(Z_{j-k} - V_{j-k}\right)\right] \\ &= E\left(\alpha^{k+1} X_{j-k-1} + \alpha^k Z_{j-k} + \sum_{l=0}^{k-1} \alpha^l Z_{j-l}\right) \\ &\quad \cdot \left(Z_{j-k} - V_{j-k}\right) \\ &= \alpha^k E\left[Z_{j-k}\left(Z_{j-k} - V_{j-k}\right)\right] \qquad (8) \\ &= \alpha^k \left\{\sigma_\infty^2 \left(1-\alpha^2\right) - D\right\}, \qquad (9) \end{aligned}$$

where

(8) follows from the fact that $X_r \perp Z_s$ and $X_r \perp V_s$ when $r < s$, and that $\{Z_k\}$ is a white Gaussian random sequence,

(9) from (1) and (7). Letting $i = j-k$ for $C_{j-k,j}$ completes the proof. ■

Lemma 3.2: The MMSE of X_n is $D \cdot \frac{1-\alpha^{2n}}{1-\alpha^2}$.

Proof: The corresponding MMSE is

$$\begin{aligned} E\left(X_n - \hat{X}_n\right)^2 &= E(X_n - \hat{X}_n)(X_n - \sum_{j,k} C_{jn} R_{jk}^{-1} \hat{Z}_k) \\ &= E\left(X_n - \hat{X}_n\right) X_n - 0 \qquad (10) \\ &= \sigma_n^2 - E\left[X_n \mathbf{c}_n^T R^{-1} \hat{\mathbf{Z}}\right] \\ &= \sigma_n^2 - \mathbf{c}_n^T R^{-1} \mathbf{c}_n, \qquad (11) \\ &= \sigma_n^2 - \left\{\sigma_\infty^2 \left(1-\alpha^2\right) - D\right\} \sum_{i=1}^n \left(\alpha^2\right)^{n-i} \\ &= D \cdot \frac{1-\alpha^{2n}}{1-\alpha^2}, \qquad (12) \end{aligned}$$

where

(10) follows from the fact that $E\left[\left(X_n - \hat{X}_n\right) \hat{Z}_k\right] = 0$ for all $k \geq 1$,

(12) from (2). ■

In particular, since $|\alpha| < 1$, passing the corresponding MMSE to the limit $n \to \infty$ yields $\frac{D}{1-\alpha^2} > D$. This suggests that perhaps sending a lossy version of the innovations process is suboptimal in Gauss-Markov rate-distortion but it's false to conclude that the additional description rate is

$$\frac{1}{2} \log\left(\frac{D/(1-\alpha^2)}{D}\right) = \frac{1}{2} \log\left(\frac{1}{1-\alpha^2}\right) \text{ [bits/sample]}$$

to get a MSE of D between $\{\hat{X}_k\}$ and $\{X_k\}$ because the probability distribution of the error process $\{X_k - \hat{X}_k\}$ under a lossy version $\{\hat{Z}_k\}$ of the innovations process is colored Gaussian, not white Gaussian. Hence it isn't necessarily simple to find how suboptimal sending a lossy version of the innovations process is in Gauss-Markov rate-distortion and is required to evaluate the $n \times n$ autocorrelation matrix of $\{X_k - \hat{X}_k, 1 \le k \le n\}$ under $\{\hat{Z}_k, k \ge 1\}$, call it $\Xi_n = \{\xi_{ij}\}$.

Lemma 3.3: $\xi_{ij} = D\alpha^{|i-j|}\frac{1-\alpha^{2\cdot\min(i,j)}}{1-\alpha^2}$.

Proof: If $i \ge j$, then

$$\begin{aligned}
\xi_{ij} &= E\left(X_i - \hat{X}_i\right)\left(X_j - \hat{X}_j\right) \\
&= E\left(X_i - \hat{X}_i\right)\left(X_j - \sum_{l,k\ge 1} C_{lj}R_{lk}^{-1}\hat{Z}_k\right) \\
&= E\left(X_i - \hat{X}_i\right)X_j - 0 \qquad (13)\\
&= \alpha^{i-j}\sigma_j^2 - EX_j\sum_{l,k\ge 1} C_{li}R_{lk}^{-1}\hat{Z}_k \qquad (14)\\
&= \alpha^{i-j}\sigma_j^2 - \sum_{l,k\ge 1} C_{li}R_{lk}^{-1}C_{kj} \\
&= \alpha^{i-j}\sigma_j^2 - \sum_{k=1}^{j} C_{ki}R_{kk}^{-1}C_{kj} \\
&= \alpha^{i-j}\sigma_j^2 - \left\{\sigma_\infty^2\left(1-\alpha^2\right) - D\right\}\sum_{k=1}^{j}\alpha^{i+j-2k} \\
&= D\alpha^{i-j}\frac{1-\alpha^{2j}}{1-\alpha^2},
\end{aligned}$$

where

(13) follows from the fact that $E\left(X_i - \hat{X}_i\right)\hat{Z}_k = 0$ for all $k \ge 1$,

(14) from (4). If $i < j$, then it can be done in the same way by symmetry. ■

IV. Proof of the Main Result

The additional description rate can be expressed in terms of the eigenvalues $\nu_{k,n}$ of the autocorrelation matrix Ξ_n [1], [6] as

$$\frac{1}{n}\sum_{k=1}^{n}\max\left[0, \frac{1}{2}\log\left(\frac{\nu_{k,n}}{\theta}\right)\right], \qquad (15)$$

where θ can be obtained by solving

$$D = \frac{1}{n}\sum_{k=1}^{n}\min\left(\theta, \nu_{k,n}\right). \qquad (16)$$

Note that $\nu_{k,n} > 0$ for all finite k and n since Ξ_n is positive definite. To evaluate the limit

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\max\left[0, \frac{1}{2}\log\left(\frac{\nu_{k,n}}{\theta}\right)\right]$$

we have to find the asymptotic eigenvalue behavior of Ξ_n since it is Hermitian, not Toeplitz. While these eigenvalues cannot be evaluated exactly except in special cases, the limits as $n \to \infty$ of the sums of (15) and (16) can be evaluated using two of Grenander and Szegö's theorems on the asymptotic eigenvalue behavior of Toeplitz and approximately Toeplitz matrices [5, pp. 62-65, 102-105]. Define the Toeplitz matrix Ψ_n by

$$\psi_{i,j} = \frac{D\alpha^{|i-j|}}{1-\alpha^2}, \qquad 1 \le i,j \le n. \qquad (17)$$

The matrix Ψ_n is intuitively a candidate for a simple matrix that is asymptotically equivalent to Ξ_n; we need prove only that it is indeed both asymptotically equivalent and simple.

Lemma 4.1: The matrices Ξ_n and Ψ_n defined in Lemma 3.3 and (17) are asymptotically equivalent; i.e. there exists a number $M < \infty$ such that

$$\begin{aligned}
\|\Xi_n\| &\le M, \qquad \forall n;\\
\|\Psi_n\| &\le M, \qquad \forall n;
\end{aligned}$$

and

$$\lim_{n\to\infty}|\Xi_n - \Psi_n| = 0.$$

Proof: Ξ_n and Ψ_n are uniformly bounded in strong (and hence in weak) norm; i.e. there exists a $M < \infty$ such that

$$\begin{aligned}
\|\Xi_n\| &\overset{(a)}{=} \max_{1\le k\le n}|\nu_{k,n}| \le \frac{D}{(1-|\alpha|)^2} = M < \infty,\\
\|\Psi_n\| &\overset{(b)}{=} \max_{1\le k\le n}|\tau_{k,n}| \le \frac{D}{(1-|\alpha|)^2} = M < \infty
\end{aligned}$$

where

$\tau_{k,n}$ are eigenvalues of the Toeplitz matrix Ψ_n,

(a) and (b) follow from the fact that Ξ_n and Ψ_n are Hermitian.

We have

$$\begin{aligned}
|\Xi_n - \Psi_n| &= \left|\left\{\frac{D\alpha^{i+j}}{1-\alpha^2}\right\}_{1\le i,j\le n}\right| \\
&= \sqrt{\frac{1}{n}\sum_{k=0}^{n-1}\zeta_{k,n}^2} \\
&= \sqrt{\frac{1}{n}\frac{D^2\alpha^4\left(1-\alpha^{2n}\right)^2}{\left(1-\alpha^2\right)^4}},
\end{aligned}$$

where $\zeta_{k,n}$ are eigenvalues of $\Xi_n - \Psi_n$. Hence there exists a $K < \infty$ such that

$$|\Xi_n - \Psi_n| = K(1-\alpha^{2n})n^{-1/2} \overset{n\to\infty}{\longrightarrow} 0.$$

■

Lemma 4.2: If r is a positive integer, then

$$\int_0^\theta \log\left(1 - 2r\cos\phi + r^2\right)d\phi$$

$$= -2\omega\log r - \mathrm{Cl}_2(2\omega) - \mathrm{Cl}_2(2\theta) + \mathrm{Cl}_2(2\omega + 2\theta),$$

where $\omega = \tan^{-1}\left(\frac{r\sin\theta}{1-r\cos\theta}\right)$.

The proof is deferred to Appendix.

Proof of Theorem 2.2: Combing Lemmas 3.1-4.2 and the fundamental eigenvalue distribution theorem [5], we have

$$\Psi(\omega) = \sum_{k=-\infty}^{\infty}\psi_k e^{-jk\omega} = \frac{D}{1-2\alpha\cos\omega+\alpha^2},$$

and the following explicit expression of the additional description rate $f(\alpha, D)$ about the original source:

$$\begin{aligned} f(\alpha,D) &= \frac{1}{4\pi}\int_{-\pi}^{\pi}\max\left[0,\log\frac{\Psi(\omega)}{\theta}\right]d\omega \\ &= \frac{1}{4\pi}\int_{-x_\theta}^{x_\theta}\log\frac{\Psi(\omega)}{\theta}d\omega \\ &= \frac{x_\theta}{2\pi}\log\frac{D}{\theta}-\frac{1}{2\pi}\int_0^{x_\theta}\log\left(1-2|\alpha|\cos\omega+\alpha^2\right)d\omega \\ &= \frac{x_\theta}{2\pi}\log\frac{D}{\theta}+\frac{1}{\pi}\left[y_\theta\log|\alpha|+\frac{1}{2}\mathrm{Cl}_2(2x_\theta)\right. \\ &\qquad \left.+\frac{1}{2}\mathrm{Cl}_2(2y_\theta)-\frac{1}{2}\mathrm{Cl}_2(2x_\theta+2y_\theta)\right], \end{aligned} \tag{18}$$

where
(18) follows from Lemma 4.2,

$$x_\theta = \cos^{-1}\left\{\frac{\theta\left(1+\alpha^2\right)-D}{2|\alpha|\theta}\right\}, \qquad 0 \le x_\theta \le \pi,$$

$$y_\theta = \tan^{-1}\left(\frac{\sqrt{2\theta\frac{1+\alpha^2}{1-\alpha^2}-1-\theta^2}}{1+\theta}\right), \quad 0 \le y_\theta \le \pi,$$

and θ can be obtained by solving

$$\begin{aligned} D &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\min\left[\theta,\Psi(\omega)\right]d\omega \\ &= \frac{1}{\pi}\int_{x_\theta}^{\pi}\frac{D}{1-2|\alpha|\cos\omega+\alpha^2}d\omega+\frac{1}{\pi}\int_0^{x_\theta}\theta d\omega \\ &= \frac{\theta x_\theta}{\pi}+\frac{D}{1-\alpha^2}-\frac{2D}{\pi(1-\alpha^2)}\tan^{-1}\left\{\left(1-\alpha^2\right)\tan\frac{x_\theta}{2}\right\}. \end{aligned}$$

This implies the desired conclusion.

V. Appendix

A. Proof of Lemma 4.2

First define

$$\mathrm{Li}_2(z) = -\int_0^z \frac{\log(1-x)}{x}dx$$

and consider the imaginary part of a function $\mathrm{Li}_2\left(re^{i\theta}\right)$.

$$\begin{aligned} \Im\left[\mathrm{Li}_2\left(re^{i\theta}\right)\right] &= \Im\left[\mathrm{Li}_2(r)-\int_r^{re^{i\theta}}\frac{\log(1-z)}{z}dz\right] \\ &= \Im\left[\mathrm{Li}_2(r)-i\int_0^{\theta}\log\left(1-re^{i\phi}\right)d\phi\right] \quad (19) \\ &= -\int_0^{\theta}\frac{1}{2}\log\left(1-2r\cos\phi+r^2\right)d\phi \quad (20) \end{aligned}$$

where
(19) follows from changing integration variable from z to $re^{i\phi}$,
(20) from separating the logarithm into real and imaginary parts.

Alternatively, we have

$$\begin{aligned} \Im\left[\mathrm{Li}_2\left(re^{i\theta}\right)\right] &= \Im\left[-\int_0^r\frac{\log\left(1-xe^{i\theta}\right)}{x}dx\right] \quad (21) \\ &= \int_0^r\tan^{-1}\left[\frac{y\sin\theta}{1-y\cos\theta}\right]\frac{1}{y}dy, \quad (22) \\ &= \int_0^{\frac{r\sin\theta}{1-r\cos\theta}}\tan^{-1}\left(\frac{1}{t}-\frac{1}{t+\tan\theta}\right)dt \quad (23) \\ &= \int_0^{\omega}\phi\left(\frac{\sec^2\phi}{\tan\phi}-\frac{\sec^2\phi}{\tan\phi+\tan\theta}\right)d\phi \quad (24) \\ &= \int_0^{\omega}\phi\frac{\sin\theta}{\sin\phi\sin(\theta+\phi)}d\phi \\ &= \left[\phi\log\frac{\sin\phi}{\sin(\phi+\theta)}\right]_0^{\omega} \\ &\quad -\int_0^{\omega}\left[\log\sin\phi-\log\sin(\phi+\theta)\right]d\phi \quad (25) \\ &= \omega\log r+\frac{1}{2}\mathrm{Cl}_2(2\omega)+\frac{1}{2}\mathrm{Cl}_2(2\theta) \\ &\quad -\frac{1}{2}\mathrm{Cl}_2(2\omega+2\theta), \quad (26) \end{aligned}$$

where
(21) follows from changing the integration variable from z to $xe^{i\theta}$,
(22) from separating the logarithm into real and imaginary parts,
(23) follows from letting $\frac{y\sin\theta}{1-y\cos\theta}=t$,
(24) from changing integration variable from t to $\tan\phi$ and $\omega=\tan^{-1}\left(\frac{r\sin\theta}{1-r\cos\theta}\right)$,
(25) follows from integration by parts,
(26) from $\omega=\tan^{-1}\left(\frac{r\sin\theta}{1-r\cos\theta}\right)$ and hence $r=\frac{\sin\omega}{\sin(\omega+\theta)}$.
Combining (20) and (26) completes the proof.

Acknowledgment

This work is supported in part by a grant from National Science Foundation and a doctoral fellowship from Korean Government.

References

[1] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, N.J.: Prentice-Hall, 1971.
[2] T. Berger, J. Gibson, "Lossy Source Coding," *IEEE Trans. Information Theory*, vol. IT-44, No. 6, pp. 2693-2723, October 1998.
[3] R. M. Gray, "Information rates of autoregressive processes," Ph.D. dissetation, Elec. Eng., Univ. South. Calif., Los Angeles, CA, 1969.
[4] R. M. Gray, "Information rates of autoregressive processes," *IEEE Trans. Information Theory*, vol. IT-16, pp. 412-421, July 1970.
[5] U. Grenander and G. Szego, *Toeplitz Forms and Their Applications*. University of California Press, Berkeley and Los Angeles, California, 1958.
[6] A. N. Kolmogorov, "On the Shannon Theory of information transmission in the case of continuous signals," *IRE Trans. Information Theory*, vol. IT-2, pp. 102-108, December 1956.
[7] K. Kim and T. Berger, "Sending a Lossy Version of the Innovations Process is Suboptimal in QG Rate-Distortion," *Proc. IEEE Int. Symp. Inform. Theory*, pp. 209-213, Adelaide, Australia, September 2005.
[8] L. Lewin, *Polylogarithms and associated functions*. New York: North Holland, 1981.