INFORMATION AND CONTROL 29, 11–28 (1975)

Block Coding for Stationary Gaussian Sources with Memory Under a Square-Error Fidelity Criterion*

HARRY H. TAN

Department of Electrical Engineering, Princeton, University, Princeton, New Jersey 08540

In this paper, we present a new version of the source coding theorem for the block coding of stationary Gaussian sources with memory under a square-error distortion criterion. For both time-discrete and time-continuous Gaussian sources, the average square-error distortion of the optimum block source code of rate $R > R(D)$ is shown to decrease at least exponentially in block-length to D, where $R(D)$ is the square-error criterion rate distortion function of the stationary Gaussian source with memory. In both cases, the exponent of convergence of average distortion is explicitly derived.

I. INTRODUCTION

In 1959, the foundations of rate distortion theory was established by Shannon (1959) when he defined the rate distortion function of an information source with respect to a fidelity criterion and proved the fundamental source coding theorems which give this function its operational significance. The standard source coding theorems give the existence of block codes of rate R and average distortion D when used to represent the information source, if and only if $R > R(D)$, the rate distortion function of the source. Thus in a communication system, where the codewords of the source code are properly channel coded for reliable transmission over the channel, $R(D)$ represents an absolute lower bound on the channel capacity required to achieve an overall system average distortion D. However, it is known that as the available channel capacity of the communication system approaches $R(D)$, the source code required to achieve this level of performance must have increasing blocklength. Since the complexity of the source encoder which

* This research was supported by the National Science Foundation under Grant No. GK-42080.

implements the source code will necessarily increase with code blocklength, a reduction in channel capacity requirement entails a higher source encoder complexity. Thus, in the selection of source codes for a communication system, the trade-off between channel capacity requirement and the source encoder complexity required to implement the source code should be considered. To resolve this problem, a study of the rate at which the average distortion of the optimum source code of rate $R > R(D)$ approaches D as its blocklength approaches infinity, is necessary.

Recently source coding theorems in the above context have been proved by Blahut (1972), Omura (1973), King (1973), and Marton (1974) for finite-alphabet discrete-time memoryless sources under bounded distortion measures. These works showed the exponential rate of convergence of average distortion to D in blocklength n for optimum source codes of rate $R > R(D)$. In this paper source coding theorems for stationary Gaussian sources with memory under a square-error distortion criterion are proved which gives an upper bound to the rate of convergence of average distortion. This is a class of sources and distortion measure which play a fundamental role in many theoretical and practical problems.

The main results of this paper are stated in Theorems 1 and 3 of Section II. Theorem 1 gives the source coding theorem for time-discrete stationary Gaussian sources and Theorem 3 the corresponding result for the time-continuous case. In both cases the average distortion of an optimum source code of rate $R > R(D)$ is shown to decrease at least exponentially in blocklength to D. The proof of Theorem 1 is contained in Section III of this paper and the proof of Theorem 3 in Section IV. The techniques we use are motivated by the methods developed by Omura (1973) and have some similarity to the derivation of the random channel coding bound for additive Gaussian noise channels (Gallager, 1968). We note that a source coding theorem for time-continuous Gaussian sources under square-error distortion criterion has been given by Gallager (1968, p. 486). In the proof of Theorem 3 here, a few of our steps follow Gallager's approach except that rigorous proofs are supplied for these steps which were only formal manipulations before in Gallager (1968, p. 487).

II. The Source Coding Theorem

We first consider a time-discrete stationary zero mean Gaussian source $\{X_t: t = 0, \pm 1, \ldots\}$, where the random n-vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)^T$ has covariance matrix $\mathbf{\Phi}_n = \{\varphi_{i-j}\}$ and where $\varphi_{i-j} = E[X_i X_j]$. We want to represent the source output with a sequence of reproduced letters from the real line according to an accuracy prescribed by a single-letter fidelity criteria generated by the mean square-error distortion criterion. That is, if a block of n source letters $\mathbf{x}^T = (x_1, ..., x_n)$ in $\mathscr{R}^n$ is represented by a block of reproduced letters $\mathbf{y}^T = (y_1, ..., y_n)$ in $\mathscr{R}^n$, a distortion $\rho_n(\mathbf{x}, \mathbf{y})$ given by

$$\rho_n(\mathbf{x}, \mathbf{y}) = \frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2 \tag{1}$$

is incurred. For any integer $n \geqslant 1$ and any $R > 0$, a set $B^n = \{\mathbf{y}_1, ..., \mathbf{y}_K\}$ of reproduced vectors in $\mathscr{R}^n$, where $K = \lfloor e^{nR} \rfloor$ (i.e., K is an integer such that $e^{nR} \geqslant K > e^{nR} - 1$) is called a (n, R) source code. Such a code B^n is is said to be of rate R and blocklength n. When the code B^n is used to represent the source output, each source word $\mathbf{x} \in \mathscr{R}^n$ is mapped into the codeword $\mathbf{y} \in B^n$ which minimizes $\rho_n(\mathbf{x}, \mathbf{y})$. The average distortion incurred when the code B^n is used to represent the source is defined to be

$$\rho_n(B^n) = E[\min_{\mathbf{y} \in B^n} \rho_n(\mathbf{X}, \mathbf{y})], \tag{2}$$

where the expectation is taken over the source ensemble. If $\rho_n(B^n) \leqslant D$, we say that the code B^n is D-admissible. In a communication system, the K codewords of a (n, R) source code B^n may be channel coded for reliable transmission over a channel of capacity C if and only if $R < C$. Thus the source coding problem is to determine for a prescribed average distortion level $D > 0$,[1] the smallest rate of any D-admissible code. The solution to this problem is contained in Theorems 1 and 2 below which show that given a $D > 0$ and any $\epsilon > 0$, a $(D + \epsilon)$-admissible (n, R) code exists for sufficiently large n if $R > R(D)$ and that no D-admissible source code has rate less than $R(D)$, where $R(D)$ is the well-known (Kolmogorov (1956)) mean-square error rate distortion function of the Gaussian source $\{X_t)$ and is given parametrically by

$$D_\theta = \frac{1}{2\pi} \int_{-\pi}^{\pi} \min(\theta, \Phi(\lambda))\, d\lambda, \tag{3}$$

$$R(D_\theta) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \max \left[0, \log \frac{\Phi(\lambda)}{\theta}\right] d\lambda, \tag{4}$$

[1] Since this is a continuous amplitude source, any finite-rate code used to represent the source will incur nonzero average distortion.

and where $\Phi(\lambda)$ is the spectral density function of the source $\{X_t\}$ and is given by

$$\Phi(\lambda) = \sum_{k=-\infty}^{\infty} \varphi_k e^{jk\lambda}, \qquad \lambda \in [-\pi, \pi]. \tag{5}$$

Moreover the nonzero portion of the $R(D)$ curve is generated as the parameter θ in (3) and (4) traverses the interval $0 < \theta < \Delta = \text{ess sup } \Phi(\lambda)$ and $R(D) = 0$ for $D \geqslant D_{\max} \triangleq D_\theta |_{\theta=\Delta} = \varphi_0 = E[X_t^2]$.

For $D \geqslant D_{\max}$, the positive side of the above assertion, that is the existence of a $(D + \epsilon)$-admissible code of rate $R > R(D)$, is trivial since the code $B^1 = \{0\}$ is a D-admissible zero rate code. For $0 < D < D_{\max}$, the following theorem holds. The proof of this theorem is given in Section III.

THEOREM 1. *Consider the discrete-time stationary zero mean Gaussian source* $\{X_t, t = 0, \pm 1, ...\}$ *with spectral density function* $\Phi(\lambda)$ *given by* (5) *and mean square error criterion rate distortion function* $R(D)$ *given by* (3) *and* (4). *Then for each* $D \in (0, D_{\max})$, *each* $R \in (R(D), \infty)$ *and each* $\epsilon_1 \in (0, D)$ *and* $\epsilon_2 > 0$, *there exist an integer* $N(D, R, \epsilon_1, \epsilon_2)$ *such that for every* $n \geqslant N(D, R, \epsilon_1, \epsilon_2)$, *there exist* (n, R) *source codes* B^n *with average distortion when used to represent the source bounded by*

$$\rho_n(B^n) \leqslant D + \exp[-n(E(R, D - \epsilon_1) - \epsilon_2)], \tag{6}$$

where for each fixed $D \in (0, D_{\max})$, $E(R, D) > 0$ *and is strictly convex in* R *for* $R > R(D)$. *Moreover for each* $\theta \in (0, \Delta)$ *and* D_θ *given by* (3), $E(R, D_\theta)$ *is given parametrically by*

$$E(R_\rho, D_\theta) = -\rho \frac{\partial E_\infty(\rho, \theta)}{\partial \rho} + E_\infty(\rho, \theta), \tag{7}$$

$$R_\rho = \frac{\partial E_\infty(\rho, \theta)}{\partial \rho}, \tag{8}$$

where $\rho \in (-1, 0)$ *and* $E_\infty(\rho, \theta)$ *is given by*

$$E_\infty(\rho, \theta) = -\frac{1}{4\pi} \int_{-\pi}^{\pi} \max \left[0, \rho \log \left(\frac{\theta(1 + \rho)}{\Phi(\lambda) + \rho\theta}\right)\right] d\lambda. \tag{9}$$

The range of the parameter ρ *in* (7) *and* (8) *corresponds to a range of rates* $R_\rho \in (R(D_\theta), \infty)$ *with* $\partial R_\rho / \partial \rho < 0$. ▮

The following converse source coding theorem follows from the variational definition of the rate distortion function (p. 105 of Berger, 1971) and can be established using an approach similar to that used by Berger (1971, p. 281).

THEOREM 2. *There are no D-admissible codes of rate $R < R(D)$.* ▮

Thus Theorem 1 has demonstrated the existence of (n, R) source codes B^n of rate $R > R(D)$ and average distortion $\rho_n(B^n)$ which decreases at least exponentially in blocklength n to D when used to represent a discrete-time Gaussian source under square-error distortion criterion. We note that this contrasts with the previous work of Bunin and Wolf (1971) where the rate of convergence of average distortion of block codes of rate $R = R(D)$ was shown to be algebraic in blocklength for a time-discrete stationary first-order Gauss–Markov source under square-error distortion criterion. It is interesting to speculate whether this is the fastest possible rate of convergence in block coding of Gaussian sources under square-error distortion criterion. We would tend to conjecture that this is the fastest rate of convergence possible when $D > D_{\text{crit}}$, where $D_{\text{crit}} = D_\theta |_{\theta=\delta}$ and $\delta = \text{ess inf}\, \Phi(\lambda)$. For $D \leqslant D_{\text{crit}}$, there is less confidence in such a conjecture in view of the double exponential rate of convergence demonstrated by Goblick (1962) and Omura and Shohara (1973) in the block coding of certain symmetric finite-alphabet memoryless sources.

We now consider a time-continuous zero-mean stationary Gaussian process $\{X(t): -\infty < t < \infty\}$, where $E[X^2(t)] < \infty$ and where

$$r(\tau) = E[X(t)\, X(t + \tau)], \qquad \tau \in \mathscr{R} \tag{10}$$

is the autocorrelation function of $\{X(t)\}$. We will also assume that $\{X(t)\}$ is a mean-square continuous process which implies that $r(\tau)$ is a continuous real-valued function of τ. Let $(\Omega, \mathscr{F}, \mathscr{P})$ be the probability space over which the random variables $\{X(t): -\infty < t < \infty\}$ are defined. That is, for each $t \in (-\infty, \infty)$, $X(t, \omega)$ is a $\mathscr{F}$-measurable function on Ω. Since the process $\{X(t)\}$ is assumed to be mean-square continuous, therefore continuous in probability, we can assume without loss of generality that the process $\{X(t): -\infty < t < \infty\}$ has been replaced by a standard measurable modification (p. 66 of Doob, 1953). That is, $X(t, \omega)$ is a real-valued (t, ω) function which is $(\mathscr{L} \times \mathscr{F})$-measurable, where $\mathscr{L}$ is the σ-algebra of Lebesgue measurable subsets of $\mathscr{R}$. The necessity of this assumption is measure-theoretic in nature and will be apparent later. We also note that the assumption of mean-square continuity of the process $\{X(t)\}$ is used here.

We are interested in block coding of the source $\{X(t): -\infty < t < \infty\}$, that is, we wish to successively code the sequence of finite sections $X_{kT}^{(k+1)T} =$

$\{X(t): kT \leqslant t \leqslant (k+1)T\}$, $k = \cdots, -1, 0, 1, \ldots$ of the source. Due to the stationarity of the source, it is sufficient to only consider the problem of coding the finite section X_0^T. That is, we can envision a source encoder which maps the set of sample paths of X_0^T into a finite set of codewords, say labeled from 1 to K, and a source decoder which associates with each integer from 1 to K a certain real-valued function on $[0, T]$.[2] The output of the source decoder is regarded as a representation of the finite section X_0^T. Obviously, successive blocks $X_{kT}^{(k+1)T}$ are treated similarly with appropriate shifting of the time variable t.

Thus the problem of coding the finite section X_0^T of the source is the problem of representing the sample paths of X_0^T by one of a finite set of real-valued functions on $[0, T]$. By considering a standard measurable modification of the process $\{X(t): -\infty < t < \infty\}$, we have ensured that the sample paths of X_0^T are in $L_2[0, T]$ with probability one, where $L_2[0, T]$ is the space of all square-integrable real-valued functions on $[0, T]$. Specifically since $X(t, \omega)$ is $(\mathscr{L} \times \mathscr{F})$-measurable and since $\int_0^T E[X^2(t, \omega)]\, dt < \infty$, it follows from the Fubini Theorem (p. 140 of Rudin, 1966) that $\int_0^T X^2(t, \omega)\, dt$ is a well-defined random variable on $(\Omega, \mathscr{F}, \mathscr{P})$ which is finite with probability one. Thus it is reasonable to attempt to represent sample paths of X_0^T by elements in $L_2[0, T]$. Here our prescribed measure of accuracy will be assumed to be the mean-square error per second distortion measure. That is, if a sample path $x(\cdot)$ of X_0^T is represented by a function $y(\cdot)$ in $L_2[0, T]$, a distortion $\rho_T(x, y)$ given by

$$\rho_T(x, y) = \frac{1}{T} \int_0^T [x(t) - y(t)]^2\, dt \tag{11}$$

is incurred.

For any $T > 0$ and $R > 0$, a set $B^T = \{y_1, \ldots, y_K\}$ of elements in $L_2[0, T]$, where $K = \lfloor e^{TR} \rfloor$ will be called a (T, R) source code. Such a code B^T is said to be of rate R and blocklength T. When the code B^T is used to represent the finite section X_0^T of the source, each sample path $x(\cdot)$ is mapped into the codeword $y(\cdot) \in B^T$ which minimizes $\rho^T(x, y)$ and the average distortion incurred is denoted by $\rho_T(B^T)$ and given by

$$\rho_T(B^T) = E[\min_{y \in B^T} \rho_T(X_0^T, y)], \tag{12}$$

where $\rho_T(X_0^T, y) = 1/T \int_0^T [X(t, \cdot) - y(t)]^2\, dt$. We note that $\rho_T(X_0^T, y)$ is a

[2] In considering only the source coding problem, the channel can be assumed to be noiseless.

well-defined random variable since the process $\{X(t)\}$ is measurable. If $\rho_T(B^T) \leqslant D$, we say that the code B^T is D-admissible. The well-known mean-square error rate distortion function $R(D)$ of the time-continuous Gaussian process $\{X(t)\}$ is given parametrically by (Kolmogorov, 1956)

$$D_\theta = \frac{1}{2\pi} \int_{-\infty}^{\infty} \min[\theta, S(\lambda)] \, d\lambda \tag{13}$$

$$R(D_\theta) = \frac{1}{4\pi} \int_{-\infty}^{\infty} \max \left[0, \log \frac{S(\lambda)}{\theta}\right] d\lambda, \tag{14}$$

where $S(\lambda)$ is the power spectral density of the process $\{X(t)\}$ and is given by

$$S(\lambda) = \int_{-\infty}^{\infty} r(\tau) \, e^{-j\lambda\tau} \, d\tau, \qquad \lambda \in (-\infty, \infty). \tag{15}$$

Here, the nonzero portion of the $R(D)$ curve is generated as the parameter θ in (13) and (14) traverses the region $0 < \theta < \Delta = \text{ess sup } S(\lambda)$ and $R(D) = 0$ for $D \geqslant D_{\max} = r(0) = E[X^2(t)]$. In Section **IV**, we will prove the following theorem which is the continuous-time version of Theorem 2.

THEOREM 3. *Consider the continuous-time stationary zero mean Gaussian process* $\{X(t), -\infty < t < \infty\}$ *with continuous autocorrelation function* $r(\tau)$, *spectral density function* $S(\lambda)$ *given by* (15) *and mean square error rate distortion function* $R(D)$ *given by* (13) *and* (14). *Then for each* $D \in (0, D_{\max})$, *each* $R \in (R(D), \infty)$ *and each* $\epsilon_1 \in (0, D)$ *and* $\epsilon_2 > 0$, *there exist a* $T_0(D, R, \epsilon_1, \epsilon_2)$ *such that for every* $T \geqslant T_0(D, R, \epsilon_1, \epsilon_2)$, *there exist* (T, R) *source codes* B^T *with average distortion* $\rho_T(B^T)$ *when used to represent the finite segment of the source* X_0^T *bounded by*

$$\rho_T(B^T) \leqslant D + \exp[-T(F(R, D - \epsilon_1) - \epsilon_2)], \tag{16}$$

where for each fixed $D \in (0, D_{\max})$, $F(R, D) > 0$ *and is strictly convex in* R *for* $R > R(D)$. *Moreover for each* $\theta \in (0, \Delta)$ *and* D_θ *given by* (13), $F(R, D_\theta)$ *is given parametrically by*

$$F(R_\rho, D_\theta) = -\rho \frac{\partial F_\infty(\rho, \theta)}{\partial \rho} + F_\infty(\rho, \theta), \tag{17}$$

$$R_\rho = \frac{\partial F_\infty(\rho, \theta)}{\partial \rho}, \tag{18}$$

where $\rho \in (-1, 0)$ *and* $F_\infty(\rho, \theta)$ *is given by*

$$F_\infty(\rho, \theta) = -\frac{1}{4\pi} \int_{-\infty}^{\infty} \max \left[0, \rho \log \left(\frac{\theta(1 + \rho)}{S(\lambda) + \rho\theta}\right)\right] d\lambda. \tag{19}$$

The range of the parameter ρ *in* (17) *and* (18) *corresponds to a range of rates* $R_\rho \in (R(D_\theta), \infty)$ *with* $\partial R_\rho / \partial \rho < 0$. ▮

For $0 < D < D_{\max}$ and any $\epsilon > 0$, Theorem 3 gives the existence of a $(D + \epsilon)$-admissible (T, R) code of rate $R > R(D)$ if T is sufficiently large. As in the discrete-time case, the existence of a D-admissible code of rate $R = R(D)$ is trivial when $D \geqslant D_{\max}$. The continuous-time version of the converse source coding theorem, Theorem 2, also carries over easily. Thus the correct operational significance of the rate distortion function $R(D)$ given by (13) and (14) has been established. Moreover, Theorem 3 shows that the rate of convergence of average distortion $\rho_T(B^T)$ to D for optimum source codes B^T of rate $R > R(D)$ is at least exponential in blocklength T in the block coding of continuous-time mean-square continuous stationary Gaussian sources under square-error distortion criterion.

III. Proof of Theorem 1

The proof of Theorem 1 is based on a random coding argument. For each D in $(0, D_{\max})$ and R in $(R(D), \infty)$ we will exhibit a sequence of ensembles of (n, R) codes B^n for which the code ensemble average distortion satisfies the inequality (6) for sufficiently large n. The proof of the theorem then follows since there must be at least one sequence of codes B^n in this sequence of code ensembles with average distortion $\rho_n(B^n)$ satisfying (6) for sufficiently large n.

For $n \geqslant 1$, let $\{\lambda_k^{(n)}: k = 1, 2,..., n\}$ be the positive eigenvalues of $\mathbf{\Phi}_n$ and let $\mathbf{\Gamma}_n$ be the unitary matrix of orthonormal eigenvectors of $\mathbf{\Phi}_n$ so that $\mathbf{\Phi}_n = \mathbf{\Gamma}_n \mathbf{\Lambda}_n \mathbf{\Gamma}_n{}^T$, where $\mathbf{\Lambda}_n$ is the diagonal matrix $\{\lambda_i^{(n)} \delta_{ij}\}$. For $\theta \in (0, \Delta)$ and $R \in (R(D_\theta), \infty)$ consider an ensemble of (n, R) source codes $B^n = \{\mathbf{Y}_1, ..., \mathbf{Y}_K\}$, where $\mathbf{Y}_k = \mathbf{\Gamma}_n \mathbf{V}_k$ for $1 \leqslant k \leqslant K$ $(K = \lfloor e^{nR} \rfloor)$. Here $\mathbf{V}_K = 0$ and $\{\mathbf{V}_k\}_{k=1}^{K-1}$ are mutually independent identically distributed random vectors independent of the source, each having joint probability distribution $Q_n(\mathbf{v}) = \prod_{k=1}^n Q_{nk}(v_k)$ given by

$$Q_{nk}(v_k) = \begin{cases} \displaystyle\int_{-\infty}^{v_k} \delta(t)\, dt & \text{if } \lambda_k^{(n)} \leqslant \theta. \\ \displaystyle\int_{-\infty}^{v_k} (2\pi(\lambda_k^{(n)} - \theta))^{-1/2} \exp\left(-t^2/2(\lambda_k^{(n)} - \theta)\right) dt & \text{if } \lambda_k^{(n)} > \theta. \end{cases} \tag{20}$$

Here $\delta(t)$ is the Dirac delta function. The expectation over the code ensemble of the average distortion incurred by codes in this ensemble when used to represent the source is called the code ensemble average distortion and is denoted by $\bar{\rho}_n$. We want to show that $\bar{\rho}_n$ satisfied the inequality (6) for sufficiently large n. Since $\mathbf{\Gamma}_n$ is a Euclidean distance preserving transformation, $\rho_n(\mathbf{X}, \mathbf{Y}_k) = 1/n \| \mathbf{X} - \mathbf{Y}_k \|^2 = 1/n \| \mathbf{\Gamma}_n{}^T(\mathbf{X} - \mathbf{Y}_k)\|^2 = 1/n \| \mathbf{U} - \mathbf{V}_k \|^2 = \rho_n(\mathbf{U}, \mathbf{V}_k)$, where $\mathbf{U} = \mathbf{\Gamma}_n{}^T\mathbf{U}$ is a Gaussian random vector with joint probability density $\tilde{p}_n(\mathbf{u}) = \prod_{k=1}^n \tilde{p}_{nk}(u_k)$ and $\tilde{p}_{nk}(u_k)$ is a $N(0, \lambda_k^{(n)})$ probability density for each k. Thus we can carry out our analysis in the transformed $\mathbf{U}$, $\mathbf{V}_k$ spaces. Accordingly

$$\begin{aligned} \bar{\rho}_n &= E[\min_{1 \leqslant k \leqslant K} \rho_n(\mathbf{X}, \mathbf{Y}_k)] \\ &= E[\min_{1 \leqslant k \leqslant K} \rho_n(\mathbf{U}, \mathbf{V}_k)], \end{aligned} \tag{21}$$

where the expectation in (21) is over both source and code ensembles.

Now define the joint conditional probability distribution

$$Q_n'(\mathbf{v} \mid \mathbf{u}) = \prod_{k=1}^n Q'_{nk}(v_k \mid u_k)$$

by

$$Q'_{nk}(v_k \mid u_k) = \begin{cases} \int_{-\infty}^{v_k} \delta(t)\, dt & \text{if } \lambda_k^{(n)} \leqslant \theta, \\ \int_{-\infty}^{v_k} (2\pi\beta_k^{(n)}\theta)^{-1/2} \exp(-(t - \beta_k^{(n)}u_k)^2/2\beta_k^{(n)}\theta)\, dt & \text{if } \lambda_k^{(n)} > \theta, \end{cases} \tag{22}$$

where $\beta_k^{(n)} = 1 - \theta/\lambda_k^{(n)}$. Define the joint conditional probability density $p_n(\mathbf{u} \mid \mathbf{v}) = \prod_{k=1}^n p_{nk}(u_k \mid v_k)$ by

$$p_{nk}(u_k \mid v_k) = \begin{cases} (2\pi\lambda_k^{(n)})^{-1/2} \exp(-u_k{}^2/2\lambda_k^{(n)}) & \text{if } \lambda_k^{(n)} \leqslant \theta, \\ (2\pi\theta)^{-1/2} \exp(-(u_k - v_k)^2/2\theta) & \text{if } \lambda_k^{(n)} > \theta. \end{cases} \tag{23}$$

It follows that

$$p_n(\mathbf{u} \mid \mathbf{v})\, dQ_n(\mathbf{v})\, d\mathbf{u} = \tilde{p}_n(\mathbf{u})\, dQ_n'(\mathbf{v} \mid \mathbf{u})\, d\mathbf{u}. \tag{24}$$

Define the functions $d_n: \mathscr{R}^{nK} \to \mathscr{R}$ and $g_n: \mathscr{R}^{n(K+1)} \to \mathscr{R}$ by

$$d_n(\mathbf{u}, \mathbf{v}_1, \ldots, \mathbf{v}_{K-1}) = \min\{\rho_n(\mathbf{u}, \mathbf{v}_1), \ldots, \rho_n(\mathbf{u}, \mathbf{v}_{K-1}), \rho_n(\mathbf{u}, \mathbf{0})\}, \tag{25}$$

and

$$g_n(\mathbf{u}, \mathbf{v}_0, ..., \mathbf{v}_{K-1}) = \begin{cases} 1 & \text{if } \rho_n(\mathbf{u}, \mathbf{v}_0) < d_n(\mathbf{u}, \mathbf{v}_1, ..., \mathbf{v}_{K-1}), \\ 0 & \text{otherwise,} \end{cases} \tag{26}$$

where $\mathbf{u}, \mathbf{v}_0, ..., \mathbf{v}_{K-1} \in \mathscr{R}^n$. Now since $\int_{\mathscr{R}^n} dQ_n'(\mathbf{v}_0 \mid \mathbf{u}) = 1$ for every $\mathbf{u} \in \mathscr{R}^n$ and since $(1 - g_n) + g_n = 1$, it follows from (21), (24), (25), and (26) that

$$\begin{aligned}
\tilde{\rho}_n &= \int_{\mathscr{R}^n} d\mathbf{u}\tilde{p}_n(\mathbf{u}) \int_{\mathscr{R}^{n(K-1)}} d_n(\mathbf{u}, \mathbf{v}_1, ..., \mathbf{v}_{K-1}) \left[\prod_{k=1}^{K-1} dQ_n(\mathbf{v}_k)\right] \\
&= \int_{\mathscr{R}^n} d\mathbf{u}\tilde{p}_n(\mathbf{u}) \int_{\mathscr{R}^{nK}} d_n(\mathbf{u}, \mathbf{v}_1, ..., \mathbf{v}_{K-1})[1 - g_n(\mathbf{u}, \mathbf{v}_0, ..., \mathbf{v}_{K-1})] \\
&\quad \times \left[\prod_{k=1}^{K} dQ_n(\mathbf{v}_k)\right] dQ_n'(\mathbf{v}_0 \mid \mathbf{u}) \\
&\quad + \int_{\mathscr{R}^n} d\mathbf{u}\tilde{p}_n(\mathbf{u}) \int_{\mathscr{R}^{nK}} d_n(\mathbf{u}, \mathbf{v}_1, ..., \mathbf{v}_{K-1})\, g_n(\mathbf{u}, \mathbf{v}_0, ..., \mathbf{v}_{K-1}) \\
&\quad \times \left[\prod_{k=1}^{K-1} dQ_n(\mathbf{v}_k)\right] dQ_n'(\mathbf{v}_0 \mid \mathbf{u}) \\
&= \int_{\mathscr{R}^n} d\mathbf{u} \int_{\mathscr{R}^{nK}} d_n(\mathbf{u}, \mathbf{v}_1, ..., \mathbf{v}_{K-1})[1 - g_n(\mathbf{u}, \mathbf{v}_0, ..., \mathbf{v}_{K-1})] \\
&\quad \times p_n(\mathbf{u} \mid \mathbf{v}_0) \left[\prod_{k=0}^{K-1} dQ_n(\mathbf{v}_k)\right] \\
&\quad + \int_{\mathscr{R}^n} d\mathbf{u} \int_{\mathscr{R}^{nK}} d_n(\mathbf{u}, \mathbf{v}_1, ..., \mathbf{v}_{K-1})\, g_n(\mathbf{u}, \mathbf{v}_0, ..., \mathbf{v}_{K-1}) \\
&\quad \times p_n(\mathbf{u} \mid \mathbf{v}_0) \left[\prod_{k=0}^{K-1} dQ_n(\mathbf{v}_k)\right].
\end{aligned} \tag{27}$$

Since $d_n(\mathbf{u}, \mathbf{v}_1, ..., \mathbf{v}_{K-1})[1 - g_n(\mathbf{u}, \mathbf{v}_0, ..., \mathbf{v}_{K-1})] \leqslant \rho_n(\mathbf{u}, \mathbf{v}_0)$, the first integral in (27) can be bounded from above by the following expression

$$\begin{aligned}
&\int_{\mathscr{R}} d\mathbf{u} \int_{\mathscr{R}^n} \rho_n(\mathbf{u}, \mathbf{v}_0)\, p_n(\mathbf{u} \mid \mathbf{v}_0)\, dQ_n(\mathbf{v}_0) \\
&\qquad = \frac{1}{n} \sum_{k=1}^{n} \min(\theta, \lambda_k^{(n)}).
\end{aligned} \tag{28}$$

Note that $d_n(\mathbf{u}, \mathbf{v}_1, \dots, \mathbf{v}_{K-1}) \leqslant \rho_n(\mathbf{u}, \mathbf{0}) = 1/n \sum_{k=1}^{n} u_k^2$. Thus in order to upper bound the second integral in (27), consider for a fixed $(\mathbf{v}_1, \dots, \mathbf{v}_{K-1}) \in \mathscr{R}^{n(K-1)}$ and an arbitrary $A > 0$ the integral

$$\int_{\mathscr{R}^n} d\mathbf{u} \int_{\mathscr{R}^n} d_n(\mathbf{u}, \mathbf{v}_1, \dots, \mathbf{v}_{K-1})\, g_n(\mathbf{u}, \mathbf{v}_0, \dots, \mathbf{v}_{K-1})\, p_n(\mathbf{u} \mid \mathbf{v}_0)\, dQ_n(\mathbf{v}_0)$$

$$\leqslant \frac{1}{n} \sum_{k=1}^{n} \int_{\mathscr{R}^{2n}} u_k^2 g_n(\mathbf{u}, \mathbf{v}_0, \dots, \mathbf{v}_{K-1})\, p_n(\mathbf{u} \mid \mathbf{v}_0)\, dQ_n(\mathbf{v}_0)\, d\mathbf{u}$$

$$= \frac{1}{n} \sum_{k=1}^{n} \Big[\int_{\{u_k^2 \leqslant A\}} u_k^2 g_n(\mathbf{u}, \mathbf{v}_0, \dots, \mathbf{v}_{K-1})\, p_n(\mathbf{u} \mid \mathbf{v}_0)\, dQ_n(\mathbf{v}_0)\, d\mathbf{u}$$

$$+ \int_{\{u_k^2 > A\}} u_k^2 g_n(\mathbf{u}, \mathbf{v}_0, \dots, \mathbf{v}_{K-1})\, p_n(\mathbf{u} \mid \mathbf{v}_0)\, dQ_n(\mathbf{v}_0)\, d\mathbf{u} \Big]$$

$$\leqslant A \int_{\mathscr{R}^{2n}} g_n(\mathbf{u}, \mathbf{v}_0, \dots, \mathbf{v}_{K-1})\, p_n(\mathbf{u} \mid \mathbf{v}_0)\, dQ_n(\mathbf{v}_0)\, d\mathbf{u}$$

$$+ \frac{1}{n} \sum_{k=1}^{n} \int_{\{u_k^2 > A\}} u_k^2 \tilde{p}_{nk}(u_k)\, du_k . \tag{29}$$

By using the inequality $\operatorname{erfc}(u) \leqslant (1/2) \exp(-u^2/2)$, where $\operatorname{erfc}(\cdot)$ is the complementary error function, the integral terms in (29) can be shown to satisfy the following inequality

$$\int_{\{u^2 > A\}} u^2 \tilde{p}_{nk}(u)\, du \leqslant (\Delta + (2\Delta A/\pi)^{1/2}) \exp(-A/2\Delta). \tag{30}$$

Combining (27), (28), (29), and (30), we have for any $A > 0$,

$$\bar{\rho}_n \leqslant \frac{1}{n} \sum_{k=1}^{n} \min(\theta, \lambda_k^{(n)}) + (\Delta + (2\Delta A/\pi)^{1/2}) \exp(-A/2\Delta)$$

$$+ A \int_{\mathscr{R}^n} d\mathbf{u} \int_{\mathscr{R}^{nK}} g_n(\mathbf{u}, \mathbf{v}_0, \dots, \mathbf{v}_{K-1})\, p_n(\mathbf{u} \mid \mathbf{v}_0) \Big[\prod_{k=0}^{K-1} dQ_n(\mathbf{v}_k) \Big]. \tag{31}$$

Now using an argument due to Omura (Lemma 1 of Omura, 1973), the integral term in (31) can be shown to satisfy the following inequality

$$\int_{\mathscr{R}^n} d\mathbf{u} \int_{\mathscr{R}^{nK}} g_n(\mathbf{u}, \mathbf{v}_0, \dots, \mathbf{v}_{K-1})\, p_n(\mathbf{u} \mid \mathbf{v}_0) \Big[\prod_{k=0}^{K-1} dQ_n(\mathbf{v}_k) \Big]$$

$$\leqslant \exp \Big[-n \Big(-\rho R + E_n(\rho, \theta) + O_1\Big(\frac{1}{n}\Big) \Big) \Big], \tag{32}$$

for any $\rho \in (-1, 0)$ where $O_1(1/n) = 1/n \log(1 - e^{-nR})$ and where

$$E_n(\rho, \theta) = -\frac{1}{n} \log \left\{ \int_{\mathscr{R}^n} d\mathbf{u} \left[\int_{\mathscr{R}^n} p_n(\mathbf{u} \mid \mathbf{v})^{1/1+\rho} \, dQ_n(\mathbf{v}) \right]^{1+\rho} \right\}$$

$$= -\frac{1}{n} \sum_{k=1}^{n} \max \left[0, \frac{\rho}{2} \log \left(\frac{\theta(1+\rho)}{\lambda_k^{(n)} + \rho\theta} \right) \right]. \tag{33}$$

Setting $\Delta = n^2$ in (31) and using (32) and the fact that $E_n(\rho, \theta) \leqslant 0$ we have, for any $\rho \in (-1, 0)$,

$$\bar{\rho}_n \leqslant \frac{1}{n} \sum_{k=1}^{n} \min(\theta, \lambda_k^{(n)}) + \exp \left[-n \left(-\rho R + E_n(\rho, \theta) - O_2 \left(\frac{\log n}{n} \right) \right) \right], \tag{34}$$

where

$$O_2 \left(\frac{\log n}{n} \right) = \frac{1}{n} \log[n^2 + (\Delta + (2\Delta/\pi)^{1/2} n) \exp(-n^2(1/2\Delta - R/n))].$$

Now using the well-known Toeplitz Distribution Theorem (p. 64 of Grenander and Szego, 1958), we have in the limit as $n \to \infty$,

$$\frac{1}{n} \sum_{k=1}^{n} \min(\theta, \lambda_k^{(n)}) \to D_\theta \tag{35}$$

and

$$E_n(\rho, \theta) \to E_\infty(\rho, \theta), \tag{36}$$

where D_θ is given by (3) and $E_\infty(\rho, \theta)$ by (9). Thus from (34), (35), and (36) we conclude that for each $\theta \in (0, \Delta)$, $\rho \in (-1, 0)$, $R \in (R(D_\theta), \infty)$ and each $\epsilon_1 > 0$, $\epsilon_2 > 0$ there exists an integer $N_1(\theta, \rho, R, \epsilon_1, \epsilon_2)$ such that for all $n \geqslant N_1$, we have

$$\bar{\rho}_n \leqslant D_\theta + \epsilon_1 + \exp[-n(-\rho R + E_\infty(\rho, \theta) - \epsilon_2)]. \tag{37}$$

For fixed θ and R we can minimize the bound by maximizing the exponent $-\rho R + E_\infty(\rho, \theta)$ over the parameter $\rho \in (-1, 0)$. Thus define for each $\theta \in (0, \Delta)$ and $R \in (R(D_\theta), \infty)$

$$E(R, D_\theta) = \sup_{\rho \in (-1,0)} [-\rho R + E_\infty(\rho, \theta)]. \tag{38}$$

The following lemma gives some useful properties of $E_\infty(\rho, \theta)$. The proof of this lemma is immediate and will be omitted here.

LEMMA 1. *For each* $\theta \in (0, \Delta)$, *we have*

(1) *For each* $\rho \in (-1, 0)$, $\dfrac{\partial E_\infty(\rho, \theta)}{\partial \rho} \geqslant 0$ *and* $\dfrac{\partial^2 E_\infty(\rho, \theta)}{\partial \rho^2} < 0$.

(2) $\lim_{\rho \to 0} \dfrac{\partial E_\infty(\rho, \theta)}{\partial \rho} = R(D_\theta)$

$$\lim_{\rho \to -1} \frac{\partial E_\infty(\rho, \theta)}{\partial \rho} = \infty.$$

Since $E_\infty(\rho, \theta)$ is strictly concave in ρ for each θ, the maximization in (38) can be performed by differentiating $-\rho R + E_\infty(\rho, \theta)$ with respect to ρ and setting to zero whenever $R(D_\theta) < R < \infty$. Thus we have shown that $E(R, D_\theta)$ is given parametrically in ρ by (7) and (8). From Lemma 1, it is clear that the range of the parameter $\rho \in (-1, 0)$ corresponds to a range of rates $R_\rho \in (R(D_\theta), \infty)$ in (7) and (8). The strict convexity of $E(R, D_\theta)$ in R for $R \in (R(D_\theta), \infty)$ is also immediate. We have thus shown that given any $\theta \in (0, \Delta)$, $R \in (R(D_\theta), \infty)$, $\epsilon_1 > 0$ and $\epsilon_2 > 0$, there exists an integer $N(D_\theta, R, \epsilon_1, \epsilon_2)$ such that for all $n \geqslant N(D_\theta, R, \epsilon_1, \epsilon_2)$.

$$\bar{\rho}_n \leqslant D_\theta + \epsilon_1 + \exp[-n(E(R, D_\theta) - \epsilon_2)]. \qquad (39)$$

Finally the inequality (39) may be written in the form of (6) by setting $D = D_\theta + \epsilon_1$ and appropriately restricting the range of ϵ_1. The proof of Theorem 1 is now complete since there must exist at least one sequence of codes B^n in the sequence of code ensembles whose average distortion $\rho_n(B^n)$ satisfies the inequality (39) when $n \geqslant N(D_\theta, R, \epsilon_1, \epsilon_2)$.

IV. PROOF OF THEOREM 3

The proof of Theorem 3 is also based on a random coding argument. For each D in $(0, D_{\max})$, R in $(R(D), \infty)$ and each $T > 0$ we will introduce an ensemble of (T, R) codes B^T. The expression for code ensemble average distortion is then reduced to one involving only countably-infinite many random variables by using a Karhunen–Loève expansion of the Gaussian source. The results of proof of Theorem 1 are then applied to show that the code ensemble distortion satisfies the inequality (16) for sufficiently large T.

We note that Gallager (p. 486 of Gallager, 1968) has given a proof of the source coding theorem for time-continuous stationary Gaussian sources under mean square error fidelity criterion also using a Karhunen–Loève expansion to reduce the expression for the code ensemble average distortion to one involving countably-infinite many random variables. However the steps provided by Gallager in manipulating the Karhunen–Loève expansion were only formal and were not rigorously justified. We will supply rigorous proofs for these steps here.

Since the source $\{X(t)\}$ is assumed to be mean square continuous, its autocorrelation function $r(\tau)$ is continuous and therefore $\int_0^T \int_0^T |r(t-s)|^2 \, dt \, ds < \infty$. Thus there exist (pp. 131, 145–152 of Liusternik and Sobolev, 1961) a decreasing sequence of numbers $\lambda_1^T \geqslant \cdots \geqslant \lambda_n^T \geqslant \cdots > 0$ and a set of orthonormal functions $\theta_1^T, \theta_2^T,\ldots$ in $L_2[0, T]$ such that

$$\int_0^T r(t-s)\, \theta_i^T(s)\, ds = \lambda_i^T \theta_i(t) \qquad \forall t \in [0, T], \tag{40}$$

for all $i = 1, 2,\ldots$. From the well-known Karhunen–Loeve Theorem (p. 478 of Loève, 1960), we have the following expansion for X_0^T

$$X(t, \omega) = \underset{N\to\infty}{\text{l.i.m.}} \sum_{i=1}^{N} \hat{X}_i(\omega)\, \theta_i^T(t), \tag{41}$$

where the convergence is in the mean square sense uniformly in t for $t \in [0, T]$. Moreover $\{\hat{X}_i\}_{i=1}^{\infty}$ is a sequence of mutually independent Gaussian random variables on $(\Omega, \mathscr{F}, \mathscr{P})$ where $\hat{X}_i$ has zero mean and variance λ_i^T. But since the convergence of the series in (41) also holds in probability for each $t \in [0, T]$, and since the sequence of random variables $\{\hat{X}_i\}_{i=1}^{\infty}$ are mutually independent, it follows from a theorem due to Levy (p. 114 of Chung, 1968) that the convergence of the series in (41) holds with probability one for each $t \in [0, T]$. Now for each integer N, it is clear that $\sum_{i=1}^{N} \hat{X}_i(\omega)\, \theta_i^T(t)$ is a $(\mathscr{L} \times \mathscr{F})$-measurable function. Since by assumption $X(t, \omega)$ is $(\mathscr{L} \times \mathscr{F})$-measurable, it follows that the set

$$\mathscr{N} = \left\{(t, \omega): \sum_{i=1}^{N} \hat{X}_i(\omega)\, \theta_i^T(t) \nrightarrow X(t, \omega) \text{ as } N \to \infty\right\}$$

is a $(\mathscr{L} \times \mathscr{F})$-measurable set. For each $t \in [0, T]$, let $\mathscr{N}_t = \{\omega \in \Omega:(t, \omega) \in \mathscr{N}\}$ and for each $\omega \in \Omega$, let $\mathscr{N}_\omega = \{t \in [0, T]: (t, \omega) \in \mathscr{N}\}$. Then $\mathscr{N}_t \in \mathscr{F}$ and $\mathscr{N}_\omega \in \mathscr{L}$. Since the convergence of the series in (41) holds with probability one for each $t \in [0, T]$, it follows that $\mathscr{P}(\mathscr{N}_t) = 0$ for each $t \in [0, T]$. If m denotes Lebesgue measure, then it follows from the Fubini Theorem that $0 = \int \mathscr{P}(\mathscr{N}_t)\, dt = (m \times \mathscr{P})(\mathscr{N}) = \int m(\mathscr{N}_\omega)\, d\mathscr{P}$ which implies that $m(\mathscr{N}_\omega) = 0$

for almost every ω, that is, with probability one. Thus we can conclude that there exists a $\Lambda \in \mathscr{F}$ such that $\mathscr{P}(\Lambda) = 0$ and such that if $\omega \notin \Lambda$, then

$$X(t, \omega) = \lim_{N\to\infty} \sum_{i=1}^{N} \hat{X}_i(\omega)\, \theta_i^T(t) \text{ for a.e. } t \in [0, T]. \tag{42}$$

Now fix a $\theta \in (0, \Delta)$. Since 0 is the only limit point of $\{\lambda_i^T\}_{i=1}^{\infty}$ and $\lambda_1^T \leqslant \Delta$, it follows that there exists an integer $L(T, \theta)$ such that $\lambda_L^T > \theta$. Thus for $\theta \in (0, \Delta)$, $T > 0$ and $R \in (R(D_\theta), \infty)$ consider the ensemble of (T, R) codes $B^T = \{Y_1(t, \omega),..., Y_K(t, \omega)\}$ such that

$$Y_k(t, \omega) = \sum_{i=1}^{L(T,\theta)} V_{ki}(\omega)\, \theta_i^T(t), \qquad t \in [0, T] \tag{43}$$

for $1 \leqslant k \leqslant K$ where $V_{Ki} = 0$ $\forall i$ and where $\{V_{ki}: i \leqslant k \leqslant K - 1, 1 \leqslant i \leqslant L(T, \theta)\}$ are mutually independent Gaussian random variables on $(\Omega, \mathscr{F}, \mathscr{P})$ independent of $\{\hat{X}_i\}_{i=1}^{\infty}$ and such that V_{ki} has mean zero and variance $(\lambda_i^T - \theta)$ for $1 \leqslant k \leqslant K - 1$. Since $Y_k(t, \omega)$ is clearly $(\mathscr{L} \times \mathscr{F})$-measurable it follows that

$$\rho_T(X_0^T, Y_k) = \frac{1}{T} \int_0^T [X(t, \cdot) - Y_k(t, \cdot)]^2\, dt \tag{44}$$

are well defined random variables on $(\Omega, \mathscr{F}, \mathscr{P})$ for $1 \leqslant k \leqslant K$. By using (42) and (44) and the orthonormality of $\{\theta_i^T\}_{i=1}^{\infty}$ we have, for $1 \leqslant k \leqslant K$,

$$\begin{aligned}\rho_T(X_0^T, Y_k) &= \frac{1}{T} \int_0^T \Big[\sum_{i=1}^{\infty} \hat{X}_i \theta_i^T(t) - \sum_{i=1}^{L} V_{ki} \theta_i^T(t)\Big]^2 dt \\ &= \frac{1}{T} \Big\{ \sum_{i=L+1}^{\infty} \hat{X}_i^2 + \sum_{i=1}^{L} (\hat{X}_i - V_{ki})^2 \Big\} \end{aligned} \tag{45}$$

with probability one. Denoting $\hat{\mathbf{X}} = (\hat{X}_i ,..., \hat{X}_L)^T$ and $\mathbf{V}_k = (V_{k1} ,..., V_{kL})^T$ and using (1) and (45) the code ensemble average distortion $\bar{\rho}_T$ can be written as

$$\begin{aligned}\bar{\rho}_T &= E[\min_{1\leqslant k\leqslant K} \rho_T(X_0^T, Y_k)] \\ &= \frac{1}{T} E\Big[\sum_{i=L+1}^{\infty} \hat{X}_i^2\Big] + \Big(\frac{L}{T}\Big) E[\min_{1\leqslant k\leqslant K} \rho_L(\hat{\mathbf{X}}, \mathbf{V}_k)] \\ &= \frac{1}{T} \sum_{i=L+1}^{\infty} \lambda_i^T + \Big(\frac{L}{T}\Big) E[\min_{1\leqslant k\leqslant K} \rho_L(\hat{\mathbf{X}}, \mathbf{V}_k)], \end{aligned} \tag{46}$$

where the expectation is over both source and code ensembles. We can now use a derivation identical to the derivation of (27) to (34) of Proof of Theorem 1 to upper bound the expectation term in (46). This results in the following inequality which is valid for any $\rho \in (-1, 0)$.

$$\bar{\rho}_T \leqslant \frac{1}{T} \sum_{i=L+1}^{\infty} \lambda_i^T + \frac{1}{T} \sum_{i=1}^{L} \theta$$

$$+ \left(\frac{L}{T}\right) \exp\left[-T(-\rho R + F_T(\rho, \theta) - O_3\left(\frac{\log T}{T}\right))\right], \quad (47)$$

where

$$F_T(\rho, \theta) = -\frac{1}{T} \sum_{i=1}^{L} \frac{\rho}{2} \log\left[\frac{\theta(1+\rho)}{\lambda_i^T + \rho}\right]$$

$$= -\frac{1}{T} \sum_{i=1}^{\infty} \max\left[0, \frac{\rho}{2} \log\left(\frac{\theta(1+\rho)}{\lambda_i^T + \rho\theta}\right)\right], \quad (48)$$

and

$$O_3\left(\frac{\log T}{T}\right) = \frac{1}{T} \{\log[T^2 + (\Delta + (2\Delta/\pi)^{1/2} T) \exp(-T^2(1/2\Delta - R/T))]\}.$$

It follows from the Theorem of Kac, Murdock, and Szego (p. 120 of Berger 1971) that there exists a constant $\mathscr{C}(\theta) < \infty$ such that $L(T, \theta)/T \leqslant \mathscr{C}(\theta)$ for all $T > 0$. Thus (47) may be rewritten as

$$\bar{\rho}_T \leqslant \frac{1}{T} \sum_{i=1}^{\infty} \min(\theta, \lambda_i^T) + \exp\left[-T\left(-\rho R + F_T(\rho, \theta) - O_4\left(\frac{\log T}{T}\right)\right)\right], \quad (49)$$

where

$$O_4\left(\frac{\log T}{T}\right) = O_3\left(\frac{\log T}{T}\right) + \frac{1}{T} \log \mathscr{C}(\theta).$$

Now using the well-known Toeplitz Distribution Theorem (Theorem 4.5.4 of Berger, 1971), we have

$$\lim_{T\to\infty} \frac{1}{T} \sum_{i=1}^{\infty} \min(\theta, \lambda_i^T) = D_\theta \quad (50)$$

and

$$\lim_{T \to \infty} F_T(\rho, \theta) = F_\infty(\rho, \theta), \tag{51}$$

where D_θ is given by (13) and $F_\infty(\rho, \theta)$ by (19). The function $F_\infty(\rho, \theta)$ can be shown to possess all of the properties of $E_\infty(\rho, \theta)$ given in Lemma 1. We define the exponent $F(R, D)$ by

$$F(R, D_\theta) = \sup_{\rho \in (-1, 0)} [-\rho R + F_\infty(\rho, \theta)] \tag{52}$$

for $\theta \in (0, \Delta)$. The proof of Theorem 3 can now be completed using an argument similar to the derivation of (37) and (39) in Proof of Theorem 1.

V. Conclusion

This paper has presented a new version of the source coding theorem for the block coding of stationary Gaussian sources with memory under a square-error distortion criterion. For both time-discrete and time-continuous Gaussian sources, the average square-error distortion of the optimum block source code of rate $R > R(D)$ was shown to decrease at least exponentially in blocklength to D. In both cases the exponent of convergence of average distortion was explicitly derived. These results have application in the study of trade-offs between source encoder complexity and channel capacity requirements in a communication system designed to transmit data from a stationary Gaussian source.

We note that Theorems 1 and 3 can be immediately generalized to include frequency or time-weighted square error distortion criterion defined analogous to that of Sakrison (1968) and Hopkins (1972). Theorem 3 can also be generalized to homogeneous Gaussian random fields under either integral square-error or weighted integral square-error distortion criterion defined by Sakrison and Algazi (1971). Finally we note that an interesting open problem is to determine lower bounds on the rate of convergence of average square-error distortion of the optimum block source code for Gaussian sources.

Received: August 6, 1974; Revised: February 21, 1975

References

Berger, T. (1971), "Rate Distortion Theory—A Mathematical Basis for Data Compression," Prentice-Hall, Englewood Cliffs, NJ.

BLAHUT, R. E. (1972), "An Hypothesis Testing Approach to Information Theory," Ph.D. Dissertation, Electrical Engineering Department, Cornell University, Ithaca, NY.

BUNIN, B. J. AND J. K. WOLF (1971), Convergence to the rate-distortion function for Gaussian sources, *IEEE Trans. Inform. Theory* **PGIT-17**, 65–70.

CHUNG, K. L. (1968), "A Course in Probability Theory," Harcourt, Brace and World, New York.

DOOB, J. L. (1953), "Stochastic Processes," Wiley, New York.

GALLAGER, R. G. (1968), "Information Theory and Reliable Communication," Wiley, New York.

GOBLICK, T. J., JR. (1962), "Coding for a Discrete Information Source with a Distortion Measure," Ph.D. Dissertation, Department of Electrical Engineering, M.I.T., Cambridge, MA.

GRENANDER, U. AND G. SZEGO (1958), "Toeplitz Forms and Their Applications," University of California Press, Berkeley, CA.

HOPKINS, W. D. (1972), "Structural Properties of Rate Distortion Functions," Ph.D. Dissertation, System Science Department, School of Engineering and Applied Science, University of California, Los Angeles, CA.

KING, W. C. (1973), "Block Coding Theorems for Discrete Memoryless Channels and Discrete Memoryless Sources," Ph.D. Dissertation, System Science Department, School of Engineering and Applied Science, University of California, Los Angeles, CA.

KOLMOGOROV, A. N. (1956), On the Shannon theory of information in the case of continuous signals, *IRE Trans. Inform. Theory* **PGIT-2**, 102–108.

LIUSTERNIK, L. A. AND V. J. SOBOLEV (1961), "Elements of Functional Analysis," Ungar, New York.

LOÈVE, M. (1960), "Probability Theory," 2nd Edition, Van Nostrand, Princeton, NJ.

MARTON, K. (1974), Error exponent for source coding with a fidelity criterion, *IEEE Trans. on Inform. Theory* **PGIT-20**, 197–199.

OMURA, J. K. (1973), A coding theorem for discrete-time sources, *IEEE Trans. on Inform. Theory* **PGIT-19**, 490–498.

OMURA, J. K. AND A. SHOHARA (1973), On convergence of distortion for block and tree encoding of symmetric sources, *IEEE Trans. on Inform. Theory* **PGIT-19**, 573–577.

RUDIN, W. (1966), "Real and Complex Analysis," McGraw-Hill, New York.

SAKRISON, D. J. (1968), The rate distortion function of a Gaussian process with a weighted square error criterion, *IEEE Trans. on Inform. Theory* **PGIT-14**, 506–508.

SAKRISON, D. J. AND V. R. ALGAZI (1971), Comparison of line-by-line and two-dimensional encoding of random images, *IEEE Trans. on Inform. Theory* **PGIT-17**, 386–398.

SHANNON, C. E. (1959), Coding theorems for a discrete source with a fidelity criterion, *IRE National Convention Record, New York*, Part 4, 142–163.