RATE DISTORTION OF STATIONARY AND NONSTATIONARY VECTOR GAUSSIAN SOURCES

Venceslav Kafedziski

Department of Electrical Engineering
University "Ss. Cyril and Methodius"
Karpos II bb, 1000 Skopje, Macedonia
E-mail: kafedzi@etf.ukim.edu.mk

ABSTRACT

Rate distortion of certain vector Gaussian sources is addressed. To evaluate rate distortion of a stationary vector Gaussian source and a vector AR Gaussian source, limiting eigenvalue distribution of block Toeplitz matrices is used. To evaluate rate distortion of a vector AR Gaussian source with time varying coefficients, where time variation is described by a stationary, ergodic vector process, random ergodic operator theory is used. To solve this last problem, the limiting eigenvalue distribution of a sequence of matrices of increasing order describing the AR source, is shown to exist. Sufficient conditions are given for each of the three cases for which source coding theorems hold.

1. INTRODUCTION

Rate distortion function dates back to Shannon's original papers [1] but he returned to it and dealt with it exhaustively in 1959 [2]. Meanwhile, Kolmogorov dealt with the computation of rate distortion of stationary Gaussian processes [3] although the terminology used was ε-entropy. Similar results appear in Gallager's classic book [4] where properties of Toeplitz forms from [5] are used. Gallager also proves coding theorem for stationary Gaussian sources. Gray [6] computes the rate distortion of Gaussian autoregressive sources using the limiting eigenvalue distribution of Toeplitz matrices of increasing size. The limiting process is elaborated in a subsequent paper [7], with application to capacity computation of FIR channels with additive Gaussian noise. In [8] Gray extends his results by relaxing the assumptions on the convergence of the AR series. Both [6] and [8] include nonstationary sources with increasing instantaneous power. Rate distortion of vector Gaussian AR sources is treated in [9] by extending the results from [6]. Coding theorem for the stationary case is also stated. Similar techniques as in [9] are used to compute capacity of a vector Gaussian channel [10, 11, 12]. In [13] the limiting eigenvalue distribution of block Toeplitz matrices is derived, following the approach for Toeplitz matrices in [7].

Here we use the results from [13] for limiting eigenvalue distribution of block Toeplitz matrices to derive rate distortion of a stationary vector Gaussian discrete time source. We then move to computing the rate distortion of a vector Gaussian AR source such that we first obtain a block Toeplitz matrix representation of the problem. FInally we compute the rate distortion of a vector Gaussian AR source with time varying coefficients described by a stationary, ergodic vector process using the results from our previous work [14]. In [14] we used the theory of random ergodic operators from [15] to solve certain information theoretic problems, such as capacity, rate distortion and linear coding for classes of non-standard channels and sources. In [14] we extended the results from [15] to operators defined in terms of asymptotically mean stationary (in fact, block stationary) processes. We then applied those results to computing capacity of time varying MIMO FIR channels where the time variation is described by a stationary, ergodic vector process (see also [16]). Here we use similar techniques to compute the rate distortion of a vector Gaussian AR source with time varying coefficients. Note that these sources are nonstationary. We prove coding theorems for all three cases, where the sufficient condition for achievability of rate distortion is that the sources have finite power.

2. BASIC DEFINITIONS

In this Section we formally define rate distortion function of a discrete time random source. We also give the rate distortion of a vector discrete time random Gaussian source. Consider a discrete time source described by a real random process X_i, $i \in \mathbb{Z}$ with process measure m_X. Consider a vector $X^N = [x(0), x(1), \ldots, x(N-1)]'$ of N source samples. The information rate distortion function of $X^N \in \mathbb{R}^N$ is defined as

$$R_N(D) = \inf I(X^N, \hat{X}^N) \tag{1}$$

where the $\inf$ is over all probability distributions $m_{X^N,\hat{X}^N}$ that induce the marginal m_{X^N} and satisfy

$$\mathbb{E}[d_N(X^N,\hat{X}^N)] \leq D \tag{2}$$

where $d_N(X^N,\hat{X}^N)$ is a distortion measure between vectors X^N and $\hat{X}^N$. We use quadratic distortion measure

$$d_N(X^N,\hat{X}^N) = \frac{1}{N}\sum_{i=0}^{N-1}(|X_i - \hat{X}_i|^2) \tag{3}$$

Note that $I(X^N,\hat{X}^N)$ is the mean $\mathbb{E}_{m_{X^n,\hat{X}^N}}[i(X^N,\hat{X}^N)]$ where $i(X^N,\hat{X}^N)$ is the information density , defined as

$$i(X^N,\hat{X}^N) = \log\frac{dm_{X^N,\hat{X}^N}}{dm_{X^N}m_{\hat{X}^N}}(x^N,\hat{x}^N) \tag{4}$$

The (information) rate distortion function of the source is defined as

$$R(D) = \lim_{N\to\infty}\frac{1}{N}R_N(D) \tag{5}$$

with the constraint (2). Using the definition above, the rate distortion of a zero mean iid Gaussian source with variance σ^2 has been found to be equal to

$$R = \begin{cases} \frac{1}{2}\log\frac{\sigma^2}{D}, & 0 \leq D \leq \sigma^2 \\ 0, & D > \sigma^2 \end{cases} \tag{6}$$

Recall [3, 6] that the rate distortion function of a vector Gaussian source of N Gaussian sources with variances σ_i^2 is given by

$$R_N(D) = \frac{1}{2}\sum_{i=1}^{N}\log\max\left(\frac{\sigma_i^2}{\theta},1\right) \tag{7}$$

where θ is found from

$$D = \frac{1}{N}\sum_{i=1}^{N}\min(\sigma_i^2,\theta) \tag{8}$$

(note that D is normalized by N). This is the basis to derive the rate distortion of a correlated stationary Gaussian source by first performing an eigen decomposition of the source correlation matrix $\Gamma_N = \mathbb{E}[X^N[X^N]']$:

$$\Psi'\Gamma_N\Psi = \mathcal{M} \tag{9}$$

Denote the eigenvalues of the correlation matrix by μ_i, $i = 1,\ldots,N$. Then $\mathcal{M} = \mathrm{diag}(\mu_1,\ldots,\mu_N)$. We transform the vector X^N from the original source by setting

$$U^N = \Psi' X^N \tag{10}$$

and the components of U^N are uncorrelated. To define rate and distortion now U^N and $\hat{U}^N$ are used instead od X^N and $\hat{X}^N$ in (1) and (2) since both equations stay the same. Then, the rate distortion of a vector of size N taken from this source is given by (7)-(8), where σ_i^2 are replaced by μ_i, $i = 1,\ldots,N$ and the rate distortion of the source is obtained by taking the limit (5) and is a well known result [3, 4].

We now define operational rate distortion. We use definitions in [17]. A $(2^{RN},N)$ rate distortion code consists of an encoding function $f_N : \mathbb{R}^N \to \{1,2,\ldots,2^{NR}\}$ and a decoding (reproduction) function $g_N : \{1,2,\ldots,2^{NR}\} \to \mathbb{R}^N$. The distortion associated with the $(2^{RN},N)$ code is defined as

$$\mathbb{E}[d_N(X^N,g_N(f_N(X^N)))] \tag{11}$$

where the expectation is with respect to the probability distribution m_{X^N}. A rate distortion pair (R,D) is said to be achievable if there exists a sequence of $(2^{NR},N)$ rate distortion codes with

$$\lim_{N\to\infty}\mathbb{E}[d_N(X^N,g_N(f_N(X^N)))] \leq D \tag{12}$$

The rate distortion region for a source is the closure of the set of achievable rate distortion pairs (R,D). The (operational) rate distortion function $R(D)$ is the infimum of rates R such that (R,D) is in the rate distortion region of the source for a given distortion D. The main theorem of rate distortion theory states that information and operational rate distortion are equal. Proving that a rate distortion pair computed from information rate distortion function is achievable is in effect proving the positive source coding theorem. Proofs usually use the random coding method and distortion typicality (information and distortion stability). Positive coding theorem for continuous time stationary Gaussian source is proved in [4] (Theorem 9.7.1). Modifications for the discrete time case, which is of interest here, can be easily made. The main ingredient of the proof is to show that Gaussian sources are information and distortion stable. To show this it has to be shown that information density $i(X^N,\hat{X}^N)$ converges to its mean (mutual information $I(X^N,\hat{X}^N)$) and distortion $d_N(X^N,\hat{X}^N)$ converges to its mean (distortion D) in probability. The same applies to $i(U^N,\hat{U}^N)$ and $d_N(U^N,\hat{U}^N)$. Variances of $i(U^N,\hat{U}^N)$ and $d_N(U^N,\hat{U}^N)$ are first computed and Chebyshev inequality is then used to show that $(X^N,\hat{X}^N)$ is stable both in information and distortion (see the proof of Theorem 9.7.1 in [4]). There is another ingredient necessary to prove the positive source coding theorem, and that refers to the choice of the maximum possible distance d_{max} between the vector to encode and the codeword, which appears in the proofs of all positive coding theorems. The role of d_{max} in Gallager's proof is played by $\frac{1}{N}\sum\mu_i$ for the discrete time case. Thus, in order to prove the coding theorem it is sufficient that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mu_i < \infty \tag{13}$$

The expression on the left is equal to the source power, which has to be finite for stationary vector sources. Thus, for all stationary Gaussian sources coding theorem can be proved. Note that some nonstationary sources that meet the condition

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mathbb{E}[|X(n)|^2] < \infty \tag{14}$$

also meet the condition of Gallager's proof. A class of sources that meet (14) is the class of asymptotically mean stationary (AMS) sources described in [18].

3. RATE DISTORTION OF STATIONARY VECTOR GAUSSIAN SOURCES

We consider a stationary vector Gaussian source $\mathbf{x}(n) = [x_1(n)x_2(n)\dots x_r(n)]'$, $n \in \mathbb{Z}$, where x_i, $i = 1, \dots, r$ is the i-th scalar process. We assume a zero mean process, and since the joint pdf is joint Gaussian pdf, the process is completely described by its second order statistics, or its covariance matrix $\mathbf{R}(l) = \mathbb{E}[\mathbf{x}(n+l)\mathbf{x}'(n)]$. $\mathbf{R}(l)$ is an $r \times r$ matrix that depends on the lag l only and has entries $R_{ij}(l) = \mathbb{E}[x_i(n+l)x_j(n)]$, $i, j = 1, \dots r$. We assume that $\sum_l |R_{ij}(l)| < \infty$ for $i, j = 1, \dots, r$. We will determine the rate distortion of this vector source. We start by considering the vector $\mathbf{X}_N = [\mathbf{x}(0)', \mathbf{x}(1)', \dots, \mathbf{x}(N-1)']'$ of size rN, describing time instants from 0 to $N-1$. The covariance matrix of the vector Gaussian source is a block Toeplitz matrix of size $rN \times rN$. Denote this matrix by $\rho_X^N = \mathbb{E}[\mathbf{X}_N \mathbf{X}_N']$:

$$\rho_X^N = \begin{bmatrix} \mathbf{R}(0) & \mathbf{R}(1) & \cdots & \mathbf{R}(N-1) \\ \mathbf{R}(1) & \mathbf{R}(0) & \cdots & \mathbf{R}(N-2) \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{R}(N-1) & \mathbf{R}(N-2) & \cdots & \mathbf{R}(0) \end{bmatrix} \tag{15}$$

The rate distortion of the vector $\mathbf{X}_N$ is given in terms of the eigenvalues μ_i of the matrix ρ_X^N:

$$R_{rN}(D) = \frac{1}{2} \sum_{i=1}^{rN} \log \max[\frac{\mu_i}{\theta}, 1] \tag{16}$$

with

$$D = \frac{1}{rN} \sum_{i=1}^{rN} \min[\mu_i, \theta] \tag{17}$$

In order to determine the rate distortion of the vector source $\mathbf{x}(n)$ we find

$$\lim_{N\to\infty} \frac{1}{rN} R_{rN}(D) \tag{18}$$

This problem reduces to finding the limiting eigenvalue distribution of a sequence of block Toeplitz matrices. From [13] we obtain:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{rN} g(\mu_k) = \int_{-0.5}^{0.5} \sum_{i=1}^{r} g(K_i) df \tag{19}$$

where g is a continuous function. $K_i(f)$, $i = 1, \dots, r$ are the eigenvalues of the spectral matrix $\mathbf{S}(f)$ with entries $S_{ik}(f) = \sum_{l=-\infty}^{\infty} R_{ik}(l) e^{-jl2\pi f}$. $S_{ik}(f)$ for $i = k = 1, \dots, r$ are the power spectral densities of the component processes, and for $i \neq k$ are the cross-spectral densities. Thus, the rate distortion is given by

$$R = \frac{1}{2r} \sum_{i=1}^{r} \int_{-0.5}^{0.5} \log \max[\frac{K_i(f)}{\theta}, 1] df \tag{20}$$

$$D = \frac{1}{r} \sum_{i=1}^{r} \int_{-0.5}^{0.5} \min[K_i(f), \theta] df \tag{21}$$

This rate distortion is achievable. This can be proven using the discrete time version of Theorem 9.7.1 in [4], which can be easily adapted to the vector Gaussian source case. After eigen decomposition of a block of size N of the given vector source we get a vector Gaussian source of dimension rN, and this theorem can be immediately applied, since

$$\lim_{N\to\infty} \frac{1}{rN} \sum_{i=1}^{rN} \mu_i = \lim_{N\to\infty} \frac{1}{rN} \sum_{i=1}^{r} \sum_{n=0}^{N-1} \mathbb{E}[|X_i(n)|^2] < \infty \tag{22}$$

4. RATE DISTORTION OF VECTOR AR SOURCES WITH CONSTANT COEFFICIENTS

Define a vector autoregressive source of r components as:

$$\mathbf{y}(n) = -\sum_{m=1}^{p} C^{(m)} \mathbf{y}(n-m) + \mathbf{z}(n) \tag{23}$$

where $C^{(m)}$ is an $r \times r$ matrix with entries $c_{ij}^{(m)}$, $i, j = 1, \dots r$ and $\mathbf{z}(n)$ is a zero mean vector white Gaussian process, with covariance matrix at lag zero $Q = \mathbb{E}[\mathbf{z}(n)\mathbf{z}'(n)]$. Assuming that Q is symmetric and positive definite, it can be diagonalized by the orthogonal matrix of eigenvectors U

$$U'QU = \Lambda_r \tag{24}$$

where $\Lambda_r = \text{diag}[\lambda_1, \dots, \lambda_r]$. Left-multiplying (23) by U' we obtain:

$$\mathbf{x}(n) = -\sum_{m=1}^{p} D^{(m)} \mathbf{x}(n-m) + \mathbf{w}(n) \tag{25}$$

where the components of $\mathbf{w}(n) = U'\mathbf{z}(n)$ are uncorrelated, the covariance matrix of $\mathbf{w}(n)$ is Λ_r, $\mathbf{x}(n) = U'\mathbf{y}(n)$ and

$$D^{(m)} = U'C^{(m)}U \tag{26}$$

We again consider a block $\mathbf{X}_N$ of N samples of the vector process. The block of noise samples

$$\mathbf{W}_N = [\mathbf{w}(0)', \mathbf{w}(1)', \ldots, \mathbf{w}(N-1)']'$$

can be written in terms of the block of signal samples as:

$$\mathbf{W}_N = \mathbf{D}_N \mathbf{X}_N \tag{27}$$

where $\mathbf{D}_N$ is a block Toeplitz matrix of size $rN \times rN$ with blocks $D^{(m)}$ on the m-th block subdiagonal. We set $D^{(0)} = I_r$, where I_r is the identity matrix of size $r \times r$. Note that $\mathbb{E}[\mathbf{W}_N \mathbf{W}_N^H] = \mathbf{D}_N \rho_X^N \mathbf{D}_N' = \Lambda_N$, where Λ_N is a block diagonal matrix with the block Λ_r on the diagonal repeated N times. Then $\rho_X^N = \mathbf{D}_N^{-1} \Lambda_N \mathbf{D}_N'^{-1}$. Instead in terms of the eigenvalues of the covariance matrix, it is better to express rate distortion in terms of the eigenvalues $\beta_i = 1/\mu_{rN-i}$ of the inverse of the covariance matrix. This matrix is equal to $\mathbf{B}_N = \mathbf{D}_N' \Lambda_N^{-1} \mathbf{D}_N$. Setting $\mathbf{A}_N = \sqrt{\Lambda_N^{-1}} \mathbf{D}_N$, we have $\mathbf{B}_N = \mathbf{A}_N' \mathbf{A}_N$. Here

$$\mathbf{A}_N = \begin{bmatrix} A^{(0)} & 0 & \cdots & 0 & \cdots & 0 \\ A^{(1)} & A^{(0)} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ A^{(p)} & A^{(p-1)} & \cdots & A^{(0)} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & A^{(p)} & \cdots & \cdots & A^{(0)} \end{bmatrix} \tag{28}$$

with $A^{(m)} = \sqrt{\Lambda_r^{-1}} D^{(m)}$ for $m = 0, \ldots, p$, with entries $a_{ik}^{(m)}$, $i, k = 1, \ldots, r$. Note that $\mathbf{B}_N$ is a block Toeplitz matrix except for the lower right corner. Then, it can be shown that when $N \to \infty$ this matrix is equivalent to a block Toeplitz matrix and its eigenvalue distribution is equal to the eigenvalue distribution of this new matrix. Instead of computing the eigenvalues of $\mathbf{B}_N$, we could perform a singular value decomposition (SVD) on $\mathbf{A}_N$ as in [12]. Then the eigenvalues when $N \to \infty$ are distributed as the squares $L_i(f)$, $i = 1, \ldots, r$ of the singular values $\sigma_i(f)$ of the complex $r \times r$ matrix $\mathbf{A}(f)$ with entries $A_{ik}(f) = \sum_{m=0}^{p} a_{ik}^{(m)} e^{-jm2\pi f}$, $i, k = 1, \ldots, r$. The condition for the convergence of the eigenvalues of block Toeplitz matrices that the sequences $a_{ik}^{(m)}$ are absolutely summable (in terms of m) is met, since we consider a finite order AR source. Instead of singular values, it is possible to use the eigenvalues of $\mathbf{A}(f)\mathbf{A}(f)^H$ or $\mathbf{A}(f)^H \mathbf{A}(f)$ in a fashion similar to that in [11], where the capacity of multivariate (MIMO) Gaussian channel with memory is derived. Thus, we arrive at the following parametric rate distortion:

$$R = \bar{\lambda} + \frac{1}{2r} \sum_{i=1}^{r} \int_{-0.5}^{0.5} \log \max[L_i(f), \frac{1}{\theta}] df \tag{29}$$

$$D = \frac{1}{r} \sum_{i=1}^{r} \int_{-0.5}^{0.5} \min[\frac{1}{L_i(f)}, \theta] df \tag{30}$$

where $\bar{\lambda} = \frac{1}{2r} \sum_{i=1}^{r} \log(\lambda_i)$. To obtain the last equation we also used:

$$\begin{aligned} \sum_{i=1}^{rN} \log \max[\frac{1}{\beta_i \theta}, 1] &= \sum_{i=1}^{rN} \log \frac{1}{\beta_i} + \sum_{i=1}^{rN} \log \max[\beta_i, \frac{1}{\theta}] \\ &= N \sum_{i=1}^{r} \log \lambda_i + \sum_{i=1}^{rN} \log \max[\beta_i, \frac{1}{\theta}] \end{aligned}$$

since $\Pi_{i=1}^{rN} \beta_i = \det(B_N) = \det A_N' \det A_N = (\det A_N)^2 = (\Pi_{i=1}^{r} \lambda_i^{-1})^N$.

Here we distinguish between two cases: stationary and nonstationary case. If the given AR vector source is stationary the proof of positive coding theorem is the same as that in Section 3. Coding theorem for stationary vector AR source has also been stated in [9]. It is well known that vector AR source is stationary (in fact asymptotically stationary) if the eigenvalues of the matrix

$$\Phi = \begin{bmatrix} -C^{(1)} & -C^{(2)} & \cdot & \cdot\cdot & & -C^{(p)} \\ I_r & 0 & \cdot & \cdot & \cdot & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdot & \cdot & 0 & I_r & 0 \end{bmatrix} \tag{31}$$

all have magnitudes less than one [19, 9]. Coding theorem for a nonstationary scalar AR source is treated in [6] and can be generalized to the nonstationary vector AR source case using similar arguments. Note that terminology nonstationary source in [6, 8] refers to the case where instantaneous signal power increases with time index. For definitions referring to nonstationary scalar AR sources see [20], page 222. These definitions are given in terms of the magnitudes of the roots of the polynomial $1 + a_1\lambda^{-1} + \cdots + a_p\lambda^{-p}$. When we consider vector AR sources the definitions are in terms of the magnitudes of the eigenvalues of matrix Φ as stated previously.

When the vector AR source is stationary, we can start with (20)-(21) to obtain (29)-(30) directly. Obviously (29)-(30) are more general and apply to nonstationary vector AR sources (in the sense defined above) as well.

5. RATE DISTORTION OF VECTOR AR SOURCES WITH STATIONARY ERGODIC COEFFICIENTS

This Section is a generalization of some of the results described in [14]. There we considered an AR source with coefficients that vary in time according to a vector stationary ergodic process. Here we extend those results to the case of a vector AR source. Note that evaluating rate distortion of such source is a problem analogous to the problem of evaluating capacity of a MIMO channel with stationary, ergodic coefficients, which we addressed in [16, 14]. As the main result in this section, we will prove the existence of the rate distortion of this vector source. Note that this vector source is nonstationary, in the sense that the coefficients of the vector AR source vary in time, not in the sense of Gray and Berger, where coefficients do not change in time but instantaneous signal power increases with time. Our model could be useful to model nonstationary signals, such as speech for example, where short frames must be used in LPC coding procedures with constant coefficients. To prove our main result, we first prove the existence of the limiting eigenvalue distribution of a sequence of matrices of increasing size which will be defined shortly.

Our source model is the same as in Section 4 except that AR coefficients change in time according to the stationary, ergodic vector process $\mathbf{c}(n) = c_{ij}^{(m)}(n)$, $n \in \mathbb{Z}$, $i, j = 1, \ldots, r$, $m = 1, \ldots, p$. Sufficient condition to obtain our results is

$$\sigma^2 = \sum_{i=1}^{r} \sum_{j=1}^{r} \sum_{m=1}^{p} \mathbb{E}[|c_{ij}^{(m)}|^2] < \infty \tag{32}$$

We perform the same transformations as in Section 4 on any finite block of r-dimensional samples of size $2N+1$ around the time index 0. Thus, all the matrices involved are of size $(2N+1)r \times (2N+1)r$. Matrix A_{2N+1} is no longer block Toeplitz, since its elements depend on the time index. We consider the infinite matrix operator A:

$$\begin{bmatrix} \cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdots & \cdot & \cdot \\ \cdot & 0 & A^{(p)}(-i) & \cdots & A^{(0)}(-i) & 0 & \cdots & \cdot & \cdot \\ \cdot & \cdot & \ddots & \ddots & \ddots & \ddots & \ddots & \cdot & \cdot \\ \cdot & \cdot & \cdots & 0 & A^{(p)}(i) & \cdots & A^{(0)}(i) & 0 & \cdot \\ \cdot & \cdot & \cdots & \cdots & \cdot & \cdot & \cdots & \cdot & \cdot \end{bmatrix} \tag{33}$$

We also consider the operator $B = A^*A$, where A^* is operator adjoint. Operator B is a Jacobi matrix operator. Note that the vector process $\mathbf{g}$ on its diagonal, super diagonals and sub diagonals with process measure $m_{\mathbf{g}}$ is block stationary ([14]). In [14] we extended the results from [15] for ergodic operators (which assume stationary processes on the diagonals of the Jacobi operator) to operators defined in terms of block stationary processes, by substituting the process measure with its stationary mean $\bar{m}_{\mathbf{g}}$ (for the definition of a stationary mean see [18]). Consider matrix B_{2N+1} of finite size $(2N+1)r \times (2N+1)r$ obtained by truncating the operator B around its entry with indices $(0, 0)$, corresponding to the time index 0. Define the empirical eigenvalue distribution function of the eigenvalues β_i of B_{2N+1} as:

$$F_N(\beta) = \frac{1}{r(2N+1)} (\text{number of eigenvalues } \beta_i \leq \beta) \tag{34}$$

Note that F_N is random, since it depends on the realization of the vector random process $\mathbf{c}(n)$ (or $\mathbf{g}(n)$, which is a function of $\mathbf{c}(n)$). We wish to find the limit of F_N when $N \to \infty$. In [14] we found that the condition for this limit to exist is $\sum_{l=-\infty}^{\infty} \mathbb{E}_{\bar{m}_{\mathbf{g}}}[|b_{l,0}|] < \infty$ where $b_{i,j}$ is the i, j-th entry in the infinite matrix B. This is just the condition in Theorem 4.8 in [15] which we modified for matrix operators defined in terms of AMS processes. It can be easily shown that this condition is met due to condition (32) and due to finite order p of the AR source. The limit F can be expressed as

$$F(\beta) = \mathbb{E}_{\bar{m}_{\mathbf{g}}}[\langle E_B(\beta)e_0, e_0\rangle] = \frac{1}{r} \sum_{j=0}^{r-1} \mathbb{E}_{m_{\mathbf{g}}}[\langle E_B(\beta)e_j, e_j\rangle] \tag{35}$$

where E_B is the resolution of the identity of the operator B, and e_j is the unit vector with the only nonzero entry at index j. Note that the convergence $F_N(\beta) \to F(\beta)$ is convergence a.e. on the sampple space of the process $\mathbf{c}(n)$, (or $\mathbf{g}(n)$) i.e. the probability of the set in the sample space on which the empirical eigenvalue distributions converge to F is equal to 1. Convergence $F_N \to F$ for a given realization of the AR coefficient process $\mathbf{c}(n)$ is convergence in distribution i.e. weak convergence of measures.

Our main result is the formula for rate distortion of the considered vector AR source:

$$R = \bar{\lambda} + \frac{1}{2} \int_0^{\infty} \log \max(\beta, \frac{1}{\theta}) dF(\beta) \tag{36}$$

$$D = \int_0^{\infty} \min(\frac{1}{\beta}, \theta) dF(\beta) \tag{37}$$

To prove this last result we use the fact that $\int h dF_N \to \int h dF$ for functions h that are bounded and for functions h for which there exists g such that $|h(\beta)/g(\beta)| \to 0$ when $\beta \to \pm\infty$ and $\limsup \int |g| dF_N < \infty$ ([21], pp. 163-164). It is sufficient to set $g(\beta) = \beta$ since $\limsup \int_0^{\infty} \beta dF_N = \limsup \sum_i \beta_{N,i} = \limsup \text{Trace} B_{2N+1} < \infty$ due to the assumption (32). Note that R are D are also obtained as limits a.e. since limit $F_N \to F$ is a limit a.e. on the AR coefficient process sample space.

Note that (29)-(30) are consistent with (36)-(37). When the AR coefficients are time non-varying the limiting eigenvalue distribution function is obtained using the arguments in Section 4, and its substitution in (36)-(37) results in (29)-(30).

Coding theorem can be proved when the source power is finite, i.e. when

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{r} \sum_{n=0}^{N-1} \mathbb{E}[|X_i(n)|^2] = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{rN} \mu_i < \infty \tag{38}$$

since Gallager's proof of Theorem 9.7.1 can be reproduced for each realization of the AR coefficient vector process $\mathbf{c}(n)$ in the set of probability 1 on which $F_N \to F$. Information and distortion stability can be proven for all realizations of $\mathbf{c}(n)$ in this set of probability 1.

Note that condition (38) holds when

$$\int_0^\infty \mu dF_\mu(\mu) < \infty \tag{39}$$

i.e. depends on the properties of $F_\mu(\mu)$, the limiting eigenvalue distribution of the eigenvalues of the covariance matrix. The existence of $F_\beta(\beta)$ implies the existence of $F_\mu(\mu)$ since

$$F_\mu(\mu) = 1 - F_\beta(\frac{1}{\mu}) \tag{40}$$

6. CONCLUSION

We have presented several rate distortion results for vector Gaussian sources, including stationary sources, AR sources with constant coefficients and AR sources with coefficients that vary according to stationary, ergodic processes. Although the first two results are not new, we rederive both results using recently published material on the asymptotic properties of block Toeplitz matrices. The third result is entirely new and is based on the theory of random operators defined in terms of block stationary processes. We also find sufficient conditions for each of the three cases for which coding theorems hold.

7. REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Systems Technical Journal*, vol. 27, no. 4, pp. 379–423,623–656, July 1948.

[2] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," in *IRE National Convention Record*, 1959, number 4, pp. 142–163.

[3] A. N. Kolmogorov, "On the Shannon theory of information in the case of continuous signals," *IRE Transactions on Information Theory*, vol. 2, no. 4, pp. 102–108, December 1956.

[4] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, Inc., 1968.

[5] U. Grenander and G. Szego, *Toeplitz Forms and Their Applications*, University of California Press, 1958.

[6] R. M. Gray, "Information rates of autoregressive processes," *IEEE Transactions on Information Theory*, vol. 16, no. 4, pp. 412–421, July 1970.

[7] R. M. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Transactions on Information Theory*, vol. 18, no. 6, pp. 725–730, November 1972.

[8] R. M. Gray, "On unbounded Toeplitz matrices and nonstationary time series with an application to information theory," *Information and Control*, vol. 24, no. 2, pp. 181–196, February 1974.

[9] W. Toms and T. Berger, "Information rates of stochastically driven dynamic systems," *IEEE Transactions on Information Theory*, vol. 17, no. 1, pp. 113–114, January 1971.

[10] W. Toms and T. Berger, "Capacity and error exponents of a channel modeled as a linear dynamic system," *IEEE Transactions on Information Theory*, vol. 19, no. 1, pp. 124–126, January 1973.

[11] L. H. Brandenburg and A. D. Wyner, "Capacity of the Gaussian channel with memory: The multivariate case," *Bell System Tech. J.*, vol. 53, no. 5, pp. 745–778, May/June 1974.

[12] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Transactions on Communications*, vol. 46, no. 3, pp. 357–366, March 1998.

[13] H. Gazzah, P. A. Regalia, and J.-P. Delmas, "Asymptotic eigenvalue distribution of block Toeplitz matrices and application to blind SIMO channel identification," *IEEE Transactions on Information Theory*, vol. 47, no. 3, pp. 1243–1251, March 2001.

[14] V. Kafedziski, "The applications of random operators in Shannon theory - channel capacity, rate distortion and optimal coding," in *Proceedings of Workshop on Concepts in Information Theory*, 2002, pp. 39–44.

[15] L. Pastur and A. Figotin, *Spectra of Random and Almost-Periodic Operators*, Springer-Verlag, 1992.

[16] V. Kafedziski, "Capacity of time varying single and multiple antenna channels with ISI using random ergodic operators," in *Proceedings of International Symposium on Information Theory*, 2002, p. 479.

[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., 1991.

[18] R. M. Gray, *Probability, Random Processes, and Ergodic Properties*, Springer-Verlag, 1988.

[19] G. C. Reinsel, *Elements of Multivariate Time Series Analysis*, Springer-Verlag, 1993.

[20] T. Berger, *Rate Distortion Theory*, Prentice-Hall, Englewood Cliffs, 1971.

[21] L. Breiman, *Probability*, Addison-Wesley, 1968.