

# THE ASYMPTOTICS OF STRING MATCHING PROBABILITIES FOR GAUSSIAN RANDOM SEQUENCES

SHUNSUKE IHARA AND MASASHI KUBO

**Abstract.** Wyner and Ziv (1989) studied the asymptotic properties of recurrence times of stationary ergodic processes, and applied the results to obtain optimal data compression schemes in information transmission. Since then many data compression algorithms based upon string matching of a sequence from an information source with a database have been proposed and studied. In this paper we consider Gaussian stationary processes representing an information source and a database, and study problems of string matching with distortion. We prove theorems concerning the asymptotic behavior of the probability of string matching with distortion and the waiting time for the string matching.

## §1. Main results

Wyner and Ziv [WZ] have proved some asymptotic properties of recurrence times of stationary ergodic finite alphabet processes, and used these properties to obtain insight into the working of the Lempel-Ziv algorithm in information transmission. They showed that  $n^{-1} \log W_n$  converges in probability to  $H$  as  $n \rightarrow \infty$ , where  $W_n$  is the recurrence time of a string with length  $n$  and  $H$  is the entropy of the process. Ornstein and Weiss [OW] extended the result by proving that the convergence is with probability one. Since then many data compression algorithms based upon string matching have been proposed and studied (see [YK] and the references therein). The probability of string matching and the waiting time for string matching have been also studied in ergodic theory [S].

In this paper we consider the lossy case, namely the problems of string matching with distortion. We are interested in the asymptotic behavior of the probability of string matching with distortion and the waiting time for string matching between two independent Gaussian stationary processes.

---

Received November 19, 1997.

Revised January 12, 2001.

2000 Mathematics Subject Classification: 60F10, 60G15, 94A15, 94A34.

Let  $X = \{X_n\}$  and  $Y = \{Y_n\}$  be stationary processes defined on a probability space  $(\Omega, \mathcal{F}, P)$ , representing an information source and a database, respectively. The probability distribution of  $Y_1^n \equiv (Y_1, \dots, Y_n)$  is denoted by  $\mu_Y^n$ . We denote by

$$B_n(x, D) = \left\{ y_1^n \in \mathbf{R}^n; \frac{1}{n} \sum_{k=1}^n |y_k - x_k|^2 < D \right\}$$

the open ball, where  $y_m^n = (y_m, \dots, y_n)$ . We consider the probability

$$\mu_Y^n(B_n(x, D)) = P \left( \frac{1}{n} \sum_{k=1}^n |Y_k - x_k|^2 < D \right)$$

of string matching within distortion  $D > 0$  (with respect to the mean-square error), where  $x = \{x_n\}$  is a realization of  $X$ . We also consider the waiting time for the string matching. For each pair of sequences  $x = \{x_n\}$  and  $y = \{y_n\}$ , the waiting time of length  $n$  is defined as

$$W_n(y, x, D) = \inf \{m \geq 0; y_m^{m+n-1} \in B_n(x, D)\}.$$

Namely,  $W_n(y, x, D)$  is the waiting time until a substring of  $y = \{y_k\}$  first hits the ball  $B_n(x, D)$ . We are interesting in the asymptotic behavior of the probability  $\mu_Y^n(B_n(x, D))$  and the waiting time  $W_n(Y, X, D)$  as  $n \rightarrow \infty$ .

In this paper, we consider the case where  $X = \{X_n\}$  and  $Y = \{Y_n\}$  are mutually independent Gaussian stationary processes. The aim of the paper is to show that the string matching probability goes to zero exponentially fast:

$$\frac{1}{n} \log \mu_Y^n(B_n(x, D)) \longrightarrow -R^*(D),$$

and that the waiting time goes to infinity with the same exponent:

$$\frac{1}{n} \log W_n(Y, X, D) \longrightarrow R^*(D),$$

as  $n \rightarrow \infty$ . The exponent rate  $R^*(D)$  is given in terms of the mutual information and the relative entropy.

We denote by  $D(\mu \parallel \nu)$  the relative entropy (or information divergence) of probability measures, defined as

$$D(\mu \parallel \nu) = \begin{cases} \int \log \frac{d\mu}{d\nu}(x) d\mu(x), & \text{if } \mu \ll \nu, \\ \infty, & \text{otherwise.} \end{cases}$$

The mutual information  $I(\xi, \eta)$  between random variables  $\xi$  and  $\eta$  is defined by

$$I(\xi, \eta) = D(\mu_{\xi\eta} \parallel \mu_\xi \times \mu_\eta),$$

where  $\mu_\xi$ ,  $\mu_\eta$  and  $\mu_{\xi\eta}$  are the probability distributions of  $\xi$ ,  $\eta$  and  $(\xi, \eta)$ , respectively, and  $\mu_\xi \times \mu_\eta$  is the direct product of  $\mu_\xi$  and  $\mu_\eta$ . For two stochastic processes  $\xi = \{\xi_n\}$  and  $\eta = \{\eta_n\}$ , the relative entropy  $\overline{D}(\xi \parallel \eta)$  per unit time is defined by

$$\overline{D}(\xi \parallel \eta) = \liminf_{n \rightarrow \infty} \frac{1}{n} D(\xi_1^n \parallel \eta_1^n),$$

where  $D(\xi_1^n \parallel \eta_1^n) = D(\mu_{\xi_1^n} \parallel \mu_{\eta_1^n})$ . The mutual information  $\overline{I}(\xi, \eta)$  per unit time is defined by

$$\overline{I}(\xi, \eta) = \liminf_{n \rightarrow \infty} \frac{1}{n} I(\xi_1^n, \eta_1^n).$$

We may assume that the Gaussian processes  $X$  and  $Y$  are regular (or purely non-deterministic), namely  $\cap_{n=1}^{\infty} \mathcal{L}_n(X) = \cap_{n=1}^{\infty} \mathcal{L}_n(Y) = \{1\}$  where  $\mathcal{L}_n(X)$  is the subspace of  $L^2(\Omega, P)$  spanned by  $X_k$ ,  $k \leq n$ , and  $\{1\}$  denotes the subspace of  $L^2(\Omega, P)$  which involves only constants. Then it is known that  $X$  and  $Y$  have spectral density functions (SDF's)  $f$  and  $g$ , respectively, such that  $\int_{-\pi}^{\pi} |\log f(\lambda)| d\lambda < \infty$  and  $\int_{-\pi}^{\pi} |\log g(\lambda)| d\lambda < \infty$  (cf. [HH]). We may assume without loss of generality that  $E[X_n] = E[Y_n] = 0$ . Then the covariance matrix of  $X_1^n$  is equal to the  $n$ -dimensional Toeplitz matrix

$$T_n(f) = [t_{jk}(f)]_{j,k=1,\dots,n}, \quad t_{jk}(f) = \int_{-\pi}^{\pi} e^{i(j-k)\lambda} f(\lambda) d\lambda,$$

of the SDF  $f$ . We note that, since  $X$  and  $Y$  are regular, the distributions of  $X_1^n$  and  $Y_1^n$  are non-degenerate, so that the matrices  $T_n(f)$  and  $T_n(g)$  are regular (see [HH]). We put

$$D_0 \equiv E[|X_n - Y_n|^2] = E[X_n^2] + E[Y_n^2].$$

For the process  $X$ , we denote by  $\mathcal{U}_D$  the class of all stochastic processes  $U = \{U_n\}$  satisfying

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n E[|U_k - X_k|^2] \leq D.$$

The following theorem concerning the string matching probability is one of our main results.

**THEOREM 1.** *Let  $X$  and  $Y$  be mutually independent regular Gaussian stationary processes and assume that the SDF's  $f(\lambda)$  and  $g(\lambda)$  are bounded. Then, for any  $0 < D < D_0$ ,*

$$(1) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log P \left( \frac{1}{n} \sum_{k=1}^n |Y_k - x_k|^2 < D \right) = -R^*(D),$$

for  $\mu_X$ -a.e.  $x = \{x_n\}$ , where  $R^*(D)$  is given by

$$(2) \quad R^*(D) = \inf \left\{ \liminf_{n \rightarrow \infty} \frac{1}{n} \left( D(U_1^n \parallel Y_1^n) + I(X_1^n, U_1^n) \right); U \in \mathcal{U}_D \right\}.$$

The asymptotic property (1) has been shown when  $Y = \{Y_n\}$  is an i.i.d. (independent with identical distribution) random sequence [KA, YK] or a Markov chain [YK]. Note that in the previous works (see [YK]) random variables take values in finite sets and the formula (1) is proved by using a combinatorial method, called the method of types. Our method to prove (1) is based upon a large deviation theorem which is stated in Section 2. We also note that our method can be applied for i.i.d. random variables taking values in an arbitrary space to show the property (1), where the distortion is measured not necessarily by squared error distortion. The proof of Theorem 1 will be given in Section 3.

To study the asymptotic behavior of the waiting time, Yang and Kieffer [YK] introduced the following mixing condition. We denote by  $\mathcal{F}_m^n$  the  $\sigma$ -field generated by  $Y_m, \dots, Y_n$ . For each positive integer  $k$ , the number

$$\varphi(k) = \sup \{ |P(G | F) - P(G)|; (G, F) \in \mathcal{G}_k \}$$

is called the mixing coefficient, where

$$\mathcal{G}_k = \{ (G, F); \exists (r, s, t) \text{ such that } r \leq s \leq t - k, G \in \mathcal{F}_r^s, F \in \mathcal{F}_t^\infty \}.$$

We say that  $Y$  has summable mixing coefficients if  $\sum_{k=1}^{\infty} \varphi(k) < \infty$ .

On the asymptotics of the waiting time, we can prove the following theorem.

**THEOREM 2.** *Let  $X$  and  $Y$  be the same processes as in Theorem 1. Assume that the process  $Y$  has summable mixing coefficients. Then, with probability one,*

$$(3) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log W_n(Y, X, D) = R^*(D).$$

In a case where the state space of  $X$  and  $Y$  is discrete, Yang and Kieffer [YK] have shown that, under the mixing condition of Theorem 2, (3) can be derived from (1). Their arguments can be applied for our case without any modification. Hence we can show Theorem 2 as a corollary of Theorem 1.

It should be noticed that

$$R(D) \leq R^*(D),$$

where  $R(D)$  is the rate distortion function (per unit time) of the process  $X$  defined as

$$R(D) = \inf\{\bar{I}(X, U); U \in \mathcal{U}_D\}.$$

The rate  $R^*(D)$  can be expressed in terms of SDF's  $f(\lambda)$  and  $g(\lambda)$ . For each  $0 < D < D_0$ , there is a unique constant  $\theta^* < 0$  determined by

$$(4) \quad \int_{-\pi}^{\pi} \frac{g(\lambda)}{1 - 4\pi\theta^*g(\lambda)} d\lambda + \int_{-\pi}^{\pi} \frac{f(\lambda)}{(1 - 4\pi\theta^*g(\lambda))^2} d\lambda = D.$$

We put

$$(5) \quad h(\lambda) = \frac{g(\lambda)}{1 - 4\pi\theta^*g(\lambda)}.$$

Note that  $X = \{X_n\}$  has the spectral representation  $X_n = \int_{-\pi}^{\pi} e^{in\lambda} d\zeta(\lambda)$ , where  $\{d\zeta(\lambda)\}$  is the random spectral measure such that  $E[|d\zeta(\lambda)|^2]/d\lambda = f(\lambda)$ . Since  $\int_{-\pi}^{\pi} h(\lambda)^2 f(\lambda) d\lambda < \infty$ , a process  $\xi = \{\xi_n\}$  defined by

$$\xi_n = -4\pi\theta^* \int_{-\pi}^{\pi} e^{in\lambda} h(\lambda) d\zeta(\lambda)$$

is a Gaussian stationary process with SDF

$$f_{\xi}(\lambda) = 16\pi^2(\theta^*)^2 h(\lambda)^2 f(\lambda).$$

We define a Gaussian stationary process  $U^* = \{U_n^*\}$  by

$$U_n^* = \xi_n + \eta_n,$$

where  $\eta = \{\eta_n\}$  is a Gaussian stationary process, independent of  $X$ , with SDF  $h(\lambda)$ . We can prove the following theorem.

**THEOREM 3.** *Let  $X$  and  $Y$  be the same processes as in Theorem 1. Then, for each  $0 < D < D_0$ ,*

$$(6) \quad R^*(D) = \overline{D}(U^* \parallel Y) + \overline{I}(X, U^*) \\ = \frac{1}{4\pi} \int_{-\pi}^{\pi} \log(1 - 4\pi\theta^*g(\lambda)) d\lambda \\ + \int_{-\pi}^{\pi} \frac{\theta^*g(\lambda)}{1 - 4\pi\theta^*g(\lambda)} d\lambda + 4\pi \int_{-\pi}^{\pi} \frac{(\theta^*)^2 f(\lambda)g(\lambda)}{(1 - 4\pi\theta^*g(\lambda))^2} d\lambda,$$

where  $\theta^* < 0$  is the constant determined by (4).

The proof will be given in Section 3.

## §2. Large deviation theorem

Let  $Z = \{Z_n\}$  be a sequence of random variables. The logarithmic moment generating function associated with  $Z$  is defined as

$$\Lambda_n(\theta) = \log E[\exp(\theta Z_n)], \quad \theta \in \mathbf{R}.$$

For each  $n$ ,  $\Lambda_n(\theta)$  is a convex function. We define the function  $\Lambda(\theta)$  as

$$(7) \quad \Lambda(\theta) = \lim_{n \rightarrow \infty} \frac{1}{n} \Lambda_n(n\theta),$$

if the limit exists. Let  $\mathcal{D}$  be the set of all  $\theta \in \mathbf{R}$  such that the limit (7) exists and the function  $\Lambda(\cdot)$  is of  $C^1$  class in a neighborhood of  $\theta$ , and defined the set  $\mathcal{D}'$  by  $\mathcal{D}' = \{\Lambda'(\theta); \theta \in \mathcal{D}\}$ . In the following we assume that

$$(8) \quad \mathcal{D} \neq \phi, \quad (\mathcal{D}')^\circ \neq \phi,$$

where  $(\mathcal{D}')^\circ$  is the interior of  $\mathcal{D}'$ . The function  $\psi(\theta)$  is defined as

$$\psi(\theta) = \theta\Lambda'(\theta) - \Lambda(\theta), \quad \theta \in \mathcal{D}.$$

The function

$$\Lambda^*(x) = \sup_{\theta \in \mathcal{D}} \{\theta x - \Lambda(\theta)\}$$

is called the Fenchel-Legendre transform of  $\Lambda(\theta)$ . Let us fix  $\theta^* \in \mathcal{D}$  such that  $\Lambda'(\theta^*) \in (\mathcal{D}')^\circ$  and  $\theta^* \neq 0$ , and consider a half-line

$$\Pi = \{x \in \mathbf{R}; \theta^*(x - a^*) > 0\},$$

where  $a^* = \Lambda'(\theta^*)$ . Clearly  $a^* \in \overline{\Pi}$  and

$$\psi(\theta^*) = \Lambda^*(a^*) = \inf\{\Lambda^*(x); x \in \overline{\Pi}\}.$$

For our purpose the following large deviation theorem (cf. [DZ]) is useful.

PROPOSITION 1. *Assume that the condition (8) is satisfied.*

(i) (Upper bound) *For any measurable set  $A \subset \overline{\Pi}$ ,*

$$(9) \quad \limsup_{n \rightarrow \infty} \frac{1}{n} \log P(Z_n \in A) \leq -\psi(\theta^*).$$

(ii) (Lower bound) *Let  $A$  be an open set such that  $A \cap (\mathcal{D}')^\circ \neq \phi$ . Then, for any  $\theta \in \mathcal{D}$  such that  $\Lambda'(\theta) \in A \cap (\mathcal{D}')^\circ$ ,*

$$(10) \quad \liminf_{n \rightarrow \infty} \frac{1}{n} \log P(Z_n \in A) \geq -\psi(\theta).$$

(iii) *Let  $A \subset \Pi$  be an open set such that  $A \cap (a^* - \delta, a^* + \delta) \neq \phi$  for any  $\delta > 0$ . Then*

$$(11) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log P(Z_n \in A) = \lim_{n \rightarrow \infty} \frac{1}{n} \log P(Z_n \in \overline{A}) = -\psi(\theta^*).$$

### §3. Proof of theorems

Let us introduce some auxiliary notations. It is convenient to use the notion of asymptotic equivalence of sequences  $\{A_n\}$  and  $\{B_n\}$  of matrices, where  $A_n$  and  $B_n$  are  $n$ -dimensional square matrices (cf. [Gr, GS]). We define norms of the matrix by

$$\|A_n\|^2 = \frac{1}{n} \sum_{i,j=1}^n |a_{ij}|^2,$$

$$\|A_n\| \|^2 = \max\{\langle A_n^* A_n x_1^n, x_1^n \rangle; x_1^n \in \mathbf{R}^n, \langle x_1^n, x_1^n \rangle \leq 1\},$$

where  $a_{ij}$  is the  $(i, j)$  component of  $A_n$  and  $\langle x_1^n, y_1^n \rangle = \sum_{k=1}^n x_k y_k$  denotes the inner product in  $\mathbf{R}^n$ . We say that  $\{A_n\}$  and  $\{B_n\}$  are asymptotically equivalent and denote  $A_n \sim B_n$ , if there exists a constant  $K < \infty$  such that  $\|A_n\|, \|B_n\| \leq K$ , for all  $n$ , and

$$\lim_{n \rightarrow \infty} \|A_n - B_n\| = 0.$$

We summarize here some properties on the asymptotic behavior of Toeplitz matrices.

LEMMA 1. (see [Gr, GS, I]) Let  $f$  be a SDF and  $\varphi \in L^2[-\pi, \pi]$ . Then

$$\begin{aligned}\lim_{n \rightarrow \infty} \frac{1}{n} \log |T_n(f)| &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \{2\pi f(\lambda)\} d\lambda, \\ \lim_{n \rightarrow \infty} \frac{1}{n} \text{Tr } T_n(f) &= 2\pi \int_{-\pi}^{\pi} f(\lambda) d\lambda,\end{aligned}$$

and

$$2\pi T_n(\varphi f) \sim T_n(\varphi) T_n(f).$$

Moreover, if  $f(\lambda) \geq a > 0$  for  $\forall \lambda \in [-\pi, \pi]$ ,

$$4\pi^2 (T_n(f))^{-1} \sim T_n(f^{-1}),$$

where  $T_n(f^{-1})$  denotes the Toeplitz matrix corresponding to  $1/f(\lambda)$ .

The following lemma will play a key role in the proof of Theorem 1.

LEMMA 2. Let  $X = \{X_n\}$  be the Gaussian process in Theorem 1. Assume that  $A_n \sim T_n(\varphi)$ , where  $\varphi(\lambda)$  is a bounded function. Then, with probability one,

$$(12) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \langle A_n X_1^n, X_1^n \rangle = \lim_{n \rightarrow \infty} \frac{1}{n} E[\langle A_n X_1^n, X_1^n \rangle] = 2\pi \int_{-\pi}^{\pi} \varphi(\lambda) f(\lambda) d\lambda.$$

The proof of Lemma 2 will be given in Appendix.

We are now in a position to prove Theorem 1.

*Proof of Theorem 1.* Let  $X = \{X_n\}$  and  $Y = \{Y_n\}$  be the processes in Theorem 1. We shall apply the large deviation theorem (Proposition 1) to a random sequence  $Z = \{Z_n\}$  given by

$$Z_n = \frac{1}{n} \sum_{k=1}^n |Y_k - x_k|^2, \quad n = 1, 2, \dots,$$

where a sequence  $x = \{x_n\} \in \mathbf{R}^\infty$  is fixed. We define an interval  $\Theta$  by

$$\Theta = \left(-\infty, \frac{1}{4\pi M}\right),$$

where  $M = \text{ess sup}\{g(\lambda); \lambda \in [-\pi, \pi]\}$ . Let  $\theta \in \Theta$  be fixed for a moment. We denote by  $I_n$  the  $n$ -dimensional identity matrix. Since  $(1 - 4\pi\theta g(\lambda))^{-1}$  is a bounded function, we know that the matrix

$$T_n(g) (T_n(g)^{-1} - 2\theta I_n) = I_n - 2\theta T_n(g) = T_n \left( \frac{1 - 4\pi\theta g}{2\pi} \right)$$



is invertible, and using Lemma 1 we have

$$T_n \left( \frac{1 - 4\pi\theta g}{2\pi} \right)^{-1} \sim \frac{1}{2\pi} T_n \left( \frac{1}{1 - 4\pi\theta g} \right).$$

Hence we can define a matrix  $A_{n,\theta}$  by

$$A_{n,\theta} = (T_n(g)^{-1} - 2\theta I_n)^{-1} = T_n \left( \frac{1 - 4\pi\theta g}{2\pi} \right)^{-1} T_n(g).$$

Then, by using Lemma 1, it is easy to show that

$$2\theta A_{n,\theta} + I_n \sim \frac{1}{\pi} T_n \left( \frac{\theta}{1 - 4\pi\theta g} \right) T_n(g) + T_n \left( \frac{1}{2\pi} \right) \sim \frac{1}{2\pi} T_n \left( \frac{1}{1 - 4\pi\theta g} \right).$$

Therefore, applying Lemma 2, we obtain

$$(13) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \langle (2\theta A_{n,\theta} + I_n)x_1^n, x_1^n \rangle = \lim_{n \rightarrow \infty} \frac{1}{n} E [\langle (2\theta A_{n,\theta} + I_n)X_1^n, X_1^n \rangle] \\ = \int_{-\pi}^{\pi} \frac{f(\lambda)}{1 - 4\pi\theta g(\lambda)} d\lambda,$$

for  $\mu_X$ -a.e.  $x$ . Since the covariance matrix of  $(Y_1, \dots, Y_n)$  is  $T_n(g)$ , we have

$$(14) \quad \Lambda_n(n\theta) = -\frac{1}{2} \log |I_n - 2\theta T_n(g)| + \theta \langle (2\theta A_{n,\theta} + I_n)x_1^n, x_1^n \rangle,$$

where  $\Lambda_n(\theta)$  is the logarithmic moment generating function associated with  $\{Z_n\}$ . It is clear from Lemma 1 that

$$(15) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log |I_n - 2\theta T_n(g)| = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log(1 - 4\pi\theta g(\lambda)) d\lambda.$$

Combine (13), (14) and (15) to get

$$(16) \quad \Lambda(\theta) = \lim_{n \rightarrow \infty} \frac{1}{n} \Lambda_n(n\theta) \\ = -\frac{1}{4\pi} \int_{-\pi}^{\pi} \log(1 - 4\pi\theta g(\lambda)) d\lambda + \int_{-\pi}^{\pi} \frac{\theta f(\lambda)}{1 - 4\pi\theta g(\lambda)} d\lambda,$$

for  $\mu_X$ -a.e.  $x$ . We should note here that, although  $\Lambda_n(\theta)$  depends on the sequence  $x = \{x_n\}$ , the limit  $\Lambda(\theta)$  does not depend on  $x$ . Differentiating the both sides of (16), we can show that  $\Lambda(\theta)$  is differentiable and

$$(17) \quad \Lambda'(\theta) = \int_{-\pi}^{\pi} \frac{g(\lambda)}{1 - 4\pi\theta g(\lambda)} d\lambda + \int_{-\pi}^{\pi} \frac{f(\lambda)}{(1 - 4\pi\theta g(\lambda))^2} d\lambda.$$

Moreover we have

$$\begin{aligned}
 (18) \quad \psi(\theta) &= \theta\Lambda'(\theta) - \Lambda(\theta) \\
 &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \log(1 - 4\pi\theta g(\lambda)) d\lambda \\
 &\quad + \int_{-\pi}^{\pi} \frac{\theta g(\lambda)}{1 - 4\pi\theta g(\lambda)} d\lambda + 4\pi \int_{-\pi}^{\pi} \frac{\theta^2 f(\lambda)g(\lambda)}{(1 - 4\pi\theta g(\lambda))^2} d\lambda.
 \end{aligned}$$

Note that  $D_0 = \int_{-\pi}^{\pi} (f(\lambda) + g(\lambda)) d\lambda = \Lambda'(0)$ . It is clear from (17) that  $\Lambda'(\theta)$  is strictly increasing continuous function. Hence there is a unique  $\theta^* \in \Theta$  such that  $\theta^* < 0$  and  $a^* = \Lambda'(\theta^*) = D$ . It is clear that

$$\Pi = \{x \in \mathbf{R}; \theta^*(x - a^*) > 0\} = (-\infty, D).$$

Applying Proposition 1 for a set  $A = (0, D) \subset \Pi$ , we obtain

$$(19) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log P \left( \frac{Z_n}{n} < D \right) = -\psi(\theta^*).$$

To complete the proof we must prove  $\psi(\theta^*) = R^*(D)$ . Noting (6) and (18), we see that this equation is shown in Theorem 3.  $\square$

We now turn to prove Theorem 3.

*Proof of Theorem 3.* From the definition of  $U^* = \{U_n^*\}$  we know that the covariance and the cross covariance are given by

$$(20) \quad E[U_{n+m}^* U_m^*] = \int_{-\pi}^{\pi} e^{in\lambda} f_{U^*}(\lambda) d\lambda$$

and

$$(21) \quad E[U_{n+m}^* X_m] = \int_{-\pi}^{\pi} e^{in\lambda} f_{U^*X}(\lambda) d\lambda,$$

where

$$(22) \quad f_{U^*}(\lambda) = f_{\eta}(\lambda) + f_{\xi}(\lambda) = h(\lambda) + 16\pi^2\theta^{*2}h(\lambda)^2 f(\lambda)$$

and

$$(23) \quad f_{U^*X}(\lambda) = -4\pi\theta^*h(\lambda)f(\lambda).$$

Using (20), (21), (22), (23), (5) and (17), we see that

$$\begin{aligned}
 (24) \quad E[|U_n^* - X_n|^2] &= \int_{-\pi}^{\pi} \{f_{U^*}(\lambda) - 2f_{U^*X}(\lambda) + f(\lambda)\} d\lambda \\
 &= \int_{-\pi}^{\pi} \left\{ \frac{g(\lambda)}{1 - 4\pi\theta^*g(\lambda)} + \frac{f(\lambda)}{(1 - 4\pi\theta^*g(\lambda))^2} \right\} d\lambda \\
 &= \Lambda'(\theta^*) \\
 &= D,
 \end{aligned}$$

meaning that

$$(25) \quad U^* \in \mathcal{U}_D.$$

It is known (cf. [I, P2]) that the relative entropy per unit time for Gaussian stationary processes is given by

$$\overline{D}(U^* \parallel Y) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left\{ \frac{f_{U^*}(\lambda)}{g(\lambda)} - 1 - \log \frac{f_{U^*}(\lambda)}{g(\lambda)} \right\} d\lambda.$$

Substitute (22) and (5) into this formula to get

$$\begin{aligned}
 (26) \quad \overline{D}(U^* \parallel Y) \\
 &= \theta^* \int_{-\pi}^{\pi} \frac{g(\lambda)(1 + 4\pi\theta^*f(\lambda) - 4\pi\theta^*g(\lambda))}{(1 - 4\pi\theta^*g(\lambda))^2} d\lambda - \frac{1}{4\pi} \int_{-\pi}^{\pi} \log \frac{f_{U^*}(\lambda)}{g(\lambda)} d\lambda.
 \end{aligned}$$

On the other hand, it is known (see [P1]) that the mutual information per unit time is calculated by

$$\overline{I}(X, U^*) = -\frac{1}{4\pi} \int_{-\pi}^{\pi} \log \left( 1 - \frac{|f_{U^*X}(\lambda)|^2}{f_{U^*}(\lambda)f(\lambda)} \right) d\lambda.$$

Substituting (22) and (23) into this formula, we obtain

$$(27) \quad \overline{I}(X, U^*) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \log \frac{f_{U^*}(\lambda)}{h(\lambda)} d\lambda.$$

It follows from (26), (27), (5) and (18) that

$$\begin{aligned}
 (28) \quad &\overline{D}(U^* \parallel Y) + \overline{I}(U^*, X) \\
 &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \left\{ \frac{4\pi\theta^*g(\lambda)(1 + 4\pi\theta^*f(\lambda) - 4\pi\theta^*g(\lambda))}{(1 - 4\pi\theta^*g(\lambda))^2} \right. \\
 &\quad \left. + \log(1 - 4\pi\theta^*g(\lambda)) \right\} d\lambda \\
 &= \psi(\theta^*).
 \end{aligned}$$

Thus, noting (25), we see that

$$(29) \quad \inf \left\{ \liminf_{n \rightarrow \infty} \frac{1}{n} \left( D(U_1^n \parallel Y_1^n) + I(X_1^n, U_1^n) \right); U \in \mathcal{U}_D \right\} \leq \psi(\theta^*).$$

To prove the converse inequality, for each  $x = \{x_n\}$ , we define a conditional probability distribution  $\nu_n^*(\cdot \mid x)$  on  $\mathbf{R}^n$  by

$$\frac{d\nu_n^*(\cdot \mid x)}{d\mu_Y^n}(y) = \exp(-\Lambda_n(n\theta^*)) \exp\left(\theta^* \sum_{k=1}^n |y_k - x_k|^2\right).$$

Let  $U \in \mathcal{U}_D$  be an arbitrary process. Then

$$\begin{aligned} & \iint_{\mathbf{R}^{2n}} \log \frac{d\nu_n^*(\cdot \mid x)}{d\mu_Y^n}(y) d\mu_{XU}^n(x, y) \\ &= -\Lambda_n(n\theta^*) + \theta^* \iint_{\mathbf{R}^{2n}} \sum_{k=1}^n |y_k - x_k|^2 d\mu_{XU}^n(x, y) \\ &= -\Lambda_n(n\theta^*) + \theta^* E \left[ \sum_{k=1}^n |U_k - X_k|^2 \right], \end{aligned}$$

and

$$\begin{aligned} (30) \quad & \liminf_{n \rightarrow \infty} \frac{1}{n} \iint_{\mathbf{R}^{2n}} \log \frac{d\nu_n^*(\cdot \mid x)}{d\mu_Y^n}(y) d\mu_{XU}^n(x, y) \\ &= -\lim_{n \rightarrow \infty} \frac{1}{n} \Lambda_n(n\theta^*) + \theta^* \limsup_{n \rightarrow \infty} \frac{1}{n} E \left[ \sum_{k=1}^n |U_k - X_k|^2 \right] \\ &\geq -\Lambda(\theta^*) + \theta^* D. \end{aligned}$$

We may assume that  $D(U_1^n \parallel Y_1^n)$  and  $I(X_1^n, U_1^n)$  are finite. Then we have

$$\begin{aligned} (31) \quad & D(U_1^n \parallel Y_1^n) + I(X_1^n, U_1^n) \\ &= \int_{\mathbf{R}^n} \log \frac{d\mu_U^n}{d\mu_Y^n}(y) d\mu_U^n(y) + \iint_{\mathbf{R}^{2n}} \log \frac{d\mu_{XU}^n}{d\mu_X^n \times \mu_U^n}(x, y) d\mu_{XU}^n(x, y) \\ &= \iint_{\mathbf{R}^{2n}} \log \left( \frac{d\mu_U^n}{d\mu_Y^n}(y) \cdot \frac{d\mu_{XU}^n}{d\mu_X^n \times \mu_U^n}(y, x) \right) d\mu_{XU}^n(x, y) \\ &= \iint_{\mathbf{R}^{2n}} \log \frac{d\mu_{U|X}^n(\cdot \mid x)}{d\mu_Y^n}(y) d\mu_{XU}^n(x, y). \end{aligned}$$

Since the relative entropy is non-negative,

$$\begin{aligned}
 (32) \quad 0 &\leq \iint_{\mathbf{R}^{2n}} \log \frac{d\mu_{U|X}^n(\cdot | x)}{d\nu_n^*(\cdot | x)}(y) d\mu_{U|X}^n(y | x) d\mu_X^n(x) \\
 &= \iint_{\mathbf{R}^{2n}} \log \frac{d\mu_{U|X}^n(\cdot | x)}{d\mu_Y^n}(y) d\mu_{XU}^n(x, y) \\
 &\quad - \iint_{\mathbf{R}^{2n}} \log \frac{d\nu_n^*(\cdot | x)}{d\mu_Y^n}(y) d\mu_{XU}^n(x, y).
 \end{aligned}$$

Therefore, it follows from (30), (31) and (32) that

$$\begin{aligned}
 &\liminf_{n \rightarrow \infty} \frac{1}{n} \{D(U_1^n \| Y_1^n) + I(X_1^n, U_1^n)\} \\
 &\geq \liminf_{n \rightarrow \infty} \frac{1}{n} \iint_{\mathbf{R}^{2n}} \log \frac{d\nu_n^*(\cdot | x)}{d\mu_Y^n}(y) d\mu_{XU}^n(x, y) \\
 &\geq -\Lambda(\theta^*) + \theta^* D \\
 &= \psi(\theta^*).
 \end{aligned}$$

This means that

$$(33) \quad \inf \left\{ \liminf_{n \rightarrow \infty} \frac{1}{n} \left( D(U_1^n \| Y_1^n) + I(X_1^n, U_1^n) \right); U \in \mathcal{U}_D \right\} \geq \psi(\theta^*).$$

The desired result (6) follows from (28), (29) and (33).  $\square$

## Appendix. Proof of Lemma 2

It is clear that

$$(34) \quad E[\langle A_n X_1^n, X_1^n \rangle] = \text{Tr}(A_n T_n(f)).$$

Since

$$A_n T_n(f) \sim T_n(\varphi) T_n(f) \sim 2\pi T_n(\varphi f),$$

we know from (34) that

$$\begin{aligned}
 \lim_{n \rightarrow \infty} \frac{1}{n} E[\langle A_n X_1^n, X_1^n \rangle] &= \lim_{n \rightarrow \infty} \frac{1}{n} \text{Tr} T_n(2\pi\varphi f) \\
 &= 2\pi \int_{-\pi}^{\pi} \varphi(\lambda) f(\lambda) d\lambda.
 \end{aligned}$$

Here we apply the large deviation theorem to the quadratic forms  $\zeta_n = n^{-1}\langle A_n X_1^n, X_1^n \rangle = n^{-1} \sum_{j,k=1}^n a_{jk} X_j X_k$ ,  $n = 1, 2, \dots$ , of the Gaussian process  $\{X_n\}$  (cf. [BD, BGR]). We put

$$\Phi_n(\theta) = \log E[\exp(\theta\zeta_n)].$$

Then we can show equations

$$\Phi_n(n\theta) = -\frac{1}{2} \log |I_n - 2\theta T_n(f)A_n|,$$

$$\Phi(\theta) \equiv \lim_{n \rightarrow \infty} \frac{1}{n} \Phi_n(n\theta) = -\frac{1}{4\pi} \int_{-\pi}^{\pi} \log(1 - 8\pi^2 \theta f(\lambda)\varphi(\lambda)) d\lambda,$$

and

$$\Phi'(\theta) = -\frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{-8\pi^2 f(\lambda)\varphi(\lambda)}{1 - 8\pi^2 \theta f(\lambda)\varphi(\lambda)} d\lambda = 2\pi \int_{-\pi}^{\pi} \frac{f(\lambda)\varphi(\lambda)}{1 - 8\pi^2 \theta f(\lambda)\varphi(\lambda)} d\lambda.$$

We note here that

$$\Phi'(0) = 2\pi \int_{-\pi}^{\pi} f(\lambda)\varphi(\lambda) d\lambda = \lim_{n \rightarrow \infty} \frac{1}{n} E[\langle A_n X_1^n, X_1^n \rangle].$$

Let  $\Psi(\theta) = \theta\Phi'(\theta) - \Phi(\theta)$ . Then

$$\Psi(\theta) = -\frac{1}{4\pi} \int_{-\pi}^{\pi} \left\{ \frac{-8\pi^2 \theta f(\lambda)\varphi(\lambda)}{1 - 8\pi^2 \theta f(\lambda)\varphi(\lambda)} - \log(1 - 8\pi^2 \theta f(\lambda)\varphi(\lambda)) \right\} d\lambda.$$

Let  $\delta > 0$  be an arbitrary small number. There exists  $\theta_1 < 0$  such that  $\Phi'(\theta_1) = \Phi'(0) - \delta$ . Applying Proposition 1, we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P(\zeta_n < \Phi'(0) - \delta) = -\Psi(\theta_1).$$

Similarly there exists  $\theta_2 > 0$  such that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P(\zeta_n > \Phi'(0) + \delta) = -\Psi(\theta_2).$$

Therefore, for the events  $G_n = \{|\zeta_n - \Phi'(0)| > \delta\}$ ,  $n = 1, 2, \dots$ , we have

$$(35) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log P(G_n) = -c,$$

where  $c = \min(\Psi(\theta_1), \Psi(\theta_2)) > 0$ . It is easily seen from (35) that the summation  $\sum_{n=1}^{\infty} P(G_n)$  is finite. Thus, by using Borel-Cantelli Lemma, we have

$$P\left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} G_k^c\right) = 1.$$

This means that, for almost all  $\omega$ , there exists  $n_0 = n_0(\omega)$  such that

$$|\zeta_n - \Phi'(0)| \leq \delta, \quad \forall n \geq n_0.$$

Since  $\delta > 0$  is arbitrary, we conclude that

$$\lim_{n \rightarrow \infty} \zeta_n = \Phi'(0) = 2\pi \int_{-\pi}^{\pi} \varphi(\lambda) f(\lambda) d\lambda,$$

with probability one. Thus the proof of (12) is complete.  $\square$

#### REFERENCES

- [BD] W. Bryc and A. Dembo, *Large deviations for quadratic functionals of Gaussian processes*, J. Theoretical Prob., **10** (1997), 307–332.
- [BGR] B. Bercu, F. Gamboa and A. Rouault, *Large deviations for quadratic forms of Gaussian stationary processes*, Stochastic Process Appl., **71** (1997), 75–90.
- [DZ] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Jones and Bartlett Pub., Boston, MA, 1992.
- [Gr] R. M. Gray, *Toeplitz and circulant matrices*, Stanford Univ., Tech. Report, 6504-1 (1977).
- [GS] V. Grenander and G. Szegő, *Toeplitz Forms and Their Applications*, Univ. of California Press, Berkeley, 1958.
- [HH] T. Hida and M. Hitsuda, *Gaussian Processes*, Amer. Math. Soc., Providence, Rhode Island, 1993.
- [I] S. Ihara, *Information Theory for Continuous Systems*, World Scientific, Singapore, 1993.
- [KA] H. Koga and S. Arimoto, *Asymptotic properties of a lossy data compression algorithm with common database*, Technical Report of IEICE, **IT-118** (1993).
- [OW] D. S. Ornstein and B. Weiss, *Entropy and data compression schemes*, IEEE Trans. Inform. Theory, **IT-39** (1993), 78–83.
- [P1] M. S. Pinsker, *Amount of information about a Gaussian stationary random process contained in a second process which is stationarily correlated with it*, Dokl. Akad. Nauk SSSR, **99** (1954), 213–216.
- [P2] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes.*, Holden-Day, San Francisco, 1964.

- [S] P. C. Shields, *The Ergodic Theory of Discrete Sample Paths*, Amer. Math. Soc., Providence, Rhode Island, 1996.
- [WZ] A. D. Wyner and J. Ziv, *Some asymptotic properties of the entropy of a stationary ergodic data source with applications to data compression*, IEEE Trans. Inform. Theory, **IT-35** (1989), 1250–1258.
- [YK] E.-H. Yang and J. C. Kieffer, *On the performance of data compression algorithms based upon string matching*, IEEE Trans. Inform. Theory, **44** (1998), 47–65.

Shunsuke Ihara  
*School of Informatics and Sciences*  
*Nagoya University*  
*Nagoya 464-8601*  
*Japan*  
`ihara@math.nagoya-u.ac.jp`

Masashi Kubo  
*Faculty of Education*  
*Tokoha Gakuen University*  
*Shizuoka 420-0911*  
*Japan*  
`kubo@tokoha-u.ac.jp`