Communication of Analog Data from a Gaussian Source Over a Noisy Channel

By A. D. WYNER

(Manuscript received December 20, 1967)

We consider the problem of the transmission of analog data from a Gaussian source over a memoryless channel with capacity C nats per second. The source emits R independent zero mean Gaussian variates per second with variance σ^2. *These digits are block-coded RN at a time into N second channel inputs. The performance criterion is the mean square error. Let* $\epsilon^2(N)$ *be the smallest attainable mean square error with parameter* $N(R, C, \sigma^2$ *fixed). Shannon has shown that* $\epsilon^2(N) \geqq \sigma^2 \exp(-2C/R) \triangleq \epsilon_0^2$ *and* $\epsilon^2(N) \to \epsilon_0^2$ *as* $N \to \infty$. *Hence the ideal error* ϵ_0^2 *is attainable in the limit as the coding delay* $N \to \infty$. *We are concerned with the rate at which* $\epsilon^2(N) \to \epsilon_0^2$, *and our principal result is that* $\epsilon^2(N) - \epsilon_0^2 \leqq O[(\log N/N)^{\frac{1}{2}}]$.

I. INTRODUCTION

We are interested in the following problem. Suppose we have an analog data source which emits a sequence of statistically independent Gaussian variates at a rate of R per second. We wish to transmit this data through a noisy channel of capacity C nats per second. Our problem is the determination of the minimum possible mean-squared-error.

Specifically we shall study the communication system of Figure 1. The output of the analog source is a sequence $X_1, X_2, X_3 \cdots$ of statistically independent Gaussian variates with zero mean and variance σ^2 which appear at the coder input at rate of R per second. After N seconds, $n = NR$ source variates have accumulated at the coder input. Let $\mathbf{X}$ denote this random n-vector. The channel is a discrete memoryless channel* which we assume accepts one input per second, and the coder contains a mapping of $\mathbf{X}$ to an allowable channel input N-vector $\mathbf{S}$. Since it requires N seconds to transmit $\mathbf{S}$, the system can process the data continuously without a "backup" at the coder input.

* Actually our results are valid for a broader class of channels. See the remark after Theorem 2 in Section II.

The decoder examines the channel output N-vector $\mathbf{R}$ and emits a Euclidean n-vector $\mathbf{X}'$ which is hopefully "close" to $\mathbf{X}$. The error criterion which we adopt is (the "mean-squared-error")

$$\epsilon^2 = \frac{1}{n} E \,||\, \mathbf{X} - \mathbf{X}' \,||^2, \tag{1}$$

where "$||\ ||$" is the Euclidean norm and E denotes expectation.

We shall assume that the parameters σ^2, R, and C are held fixed for this entire paper, and denote by $\epsilon^2(N)$, the smallest value of ϵ^2 attainable with parameter N (and therefore $n = RN$). Shannon[1] has shown that

$$(i) \quad \epsilon^2(N) \geqq \sigma^2 \exp(-2C/R) \triangleq \epsilon_0^2\,, \tag{2a}$$

$$(ii) \quad \epsilon^2(N) \to \epsilon_0^2 \quad \text{as} \quad N \to \infty, \tag{2b}$$

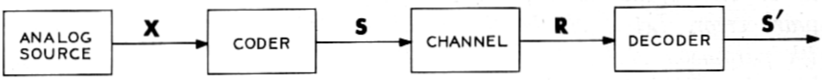

Fig. 1 — Communication System.

so that $\epsilon^2 = \epsilon_0^2$ is attainable in the limit as the delay $N \to \infty$. We are concerned here with the rate at which $\epsilon^2(N)$ approaches the ideal ϵ_0^2, and our principal result is

$$\epsilon^2(N) \leqq \epsilon_0^2\left[1 + \frac{2}{R\sqrt{\beta}}\sqrt{\frac{\log N}{N}} + O\left(\frac{1}{\sqrt{N}}\right)\right], \quad \text{as} \quad N \to \infty, \tag{3}$$

where $\beta > 0$, a parameter related to the channel, is defined in equation (13).

This result is related to a result of D. Sakrison[2] which was done independently.* In fact we have used one of his ideas (Lemma 1 in our paper) to simplify our original proof.

II. STATEMENT AND DISCUSSION OF RESULTS

Following Shannon's technique,[1] we separate the coder into two parts as shown in Figure 2. The first part, called the *source encoder* or *quantizer*, contains a fixed set $\mathcal{S}$ of M Euclidean n-vectors, and associates with each possible input n-vector $\mathbf{X}$ a member of $\mathcal{S}$ (say $\hat{\mathbf{X}}$). Let us denote

* This paper and Sakrison's paper were presented at the International Symposium on Information Theory, San Remo, Italy, September 1967.

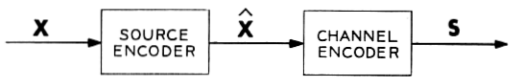

Fig. 2 — Decomposition of the Coder.

the resulting (mean-square) "quantization error" by

$$\epsilon_Q^2 = \frac{1}{n} E \, || \mathbf{X} - \hat{\mathbf{X}} ||^2. \tag{4}$$

The second part of the encoder, called the *channel encoder*, associates with each $\hat{\mathbf{X}}$ (one of the M members of $\mathcal{S}$) a channel input N-vector (say $\mathbf{S}$). The decoder, at the receiving end of the channel, associates with each received N-vector $\mathbf{R}$ one of the members of $\mathcal{S}$ (say $\hat{\mathbf{X}}'$). Let $P_e = \Pr \{\hat{\mathbf{X}}' \neq \hat{\mathbf{X}}\}$ be the probability of a transmission error, and denote the (mean-squared) transmission error by

$$\epsilon_T^2 = \frac{1}{n} E \, || \hat{\mathbf{X}} - \hat{\mathbf{X}}' ||^2. \tag{5}$$

Clearly

$$\epsilon_T^2 \leqq \frac{1}{n} P_e \max_{\mathbf{u}, \mathbf{v} \in \mathcal{S}} || \mathbf{u} - \mathbf{v} ||^2. \tag{6}$$

Further, the overall error ϵ^2 satisfies

$$\epsilon^2 = \frac{1}{n} E \, || \mathbf{X} - \mathbf{X}' ||^2 \leqq (\epsilon_Q + \epsilon_T)^2. \tag{7}$$

Thus we want to make both ϵ_Q^2 and ϵ_T^2 as small as possible.

Consider the parameter M, the number of members of the approximating set $\mathcal{S}$. In the interest of minimizing ϵ_Q^2 we want to make M large. However, in the interest of minimizing P_e and therefore ϵ_T^2, we want to make M small. The proper compromise yields our result, equation (3).

The following theorems indicate just how to choose M. The first is proved in Section III, and the second was proved by C. E. Shannon (Ref. 3, p. 16).

Theorem 1: (Source Encoding). Let $\mathbf{X}$ *be a random n-vector* ($n = 1, 2, \cdots$) *whose components are zero mean Gaussian variates with variance* σ^2. *Let* $R_0 > 0$ *be given and let* $\{a_n\}_1^\infty$ *be a sequence which tends to zero. Then there exists a set* $\mathcal{S}$ *of* M *n-vectors, where*

$$M \leqq \exp [n(R_0 - a_n)], \tag{8}$$

and a mapping f of **X** *into* $\mathcal{S}$, *such that as* $n \to \infty$

$$\frac{1}{n} E \,||\, \mathbf{X} - f(\mathbf{X}) \,||^2 \leqq \sigma^2 e^{-2R_0}\left(1 + 2a_n + \frac{\log n}{n}\right) + O(a_n^2) + O\left(\frac{\log \log n}{n}\right). \tag{9}$$

Furthermore, $\mathcal{S}$ *is such that for all* $\mathbf{u} \in \mathcal{S}$,

$$\frac{1}{n} \,||\, \mathbf{u} \,||^2 \leqq \sigma^2. \tag{10}$$

A special case of some interest in itself is that for which $a_n \equiv 0$. In this case Theorem 1 asserts the existence of a quantization of **X** with $M = \exp(nR_0)$ points and mean-squared quantization error no more than $\sigma^2 \exp(-2R_0)(1 + \log n/n) + O(\log \log n/n)$ as $n \to \infty$. If the channel is noiseless with capacity C nats second, it can transmit e^{CN} messages with no error in the N seconds that it takes for the n-vector **X** to be emitted from the source. Thus if $R_0 = C/R$, $M = \exp(R_0 n) = e^{CN}$, and the members of $\mathcal{S}$ can be reconstructed perfectly by the receiver. The overall error is therefore

$$\epsilon^2 = \epsilon_Q^2 \leqq \sigma^2 \exp(-2C/R)\left(1 + \frac{\log n}{n}\right) + O\left(\frac{\log \log n}{n}\right). \tag{11}$$

Let us turn now to the discrete memoryless channel defined by an input set $(1, 2, \cdots, K)$, an output set $(1, 2, \cdots, J)$, and a set of transition probabilities $P(j \mid k)$, $1 \leqq j \leqq J$, $1 \leqq k \leqq K$. Corresponding to each input probability distribution p_k, $1 \leqq k \leqq K$, there is a joint distribution $p(k, j) = P(j \mid k)p_k$ on the product of the input and output sets. Define the random variable (called "information")

$$U(k, j) = \log \left\{ \frac{p(j \mid k)}{\sum_{k'=1}^{K} p_{k'} p(j \mid k')} \right\}, \quad 1 \leqq k \leqq K, \quad 1 \leqq j \leqq J. \tag{12}$$

The channel capacity $C = \max_{\{p_k\}} EU$, where the maximization is performed with respect to all possible input probability distributions. Let $(p_k^*)_{k=1}^K$ be a maximizing input distribution, and let $U^*(k, j)$ be the corresponding information, then define

$$\beta = (2 \operatorname{var} U^*)^{-1} = (2EU^{*2} - 2C^2)^{-1}. \tag{13}$$

Theorem 2: (Shannon): Let $\{b_N\}_{N=1}^{\infty}$ *be a sequence which tends to zero from above. Then there exists an N-dimensional code (for the channel* *described above) with M members*

$$M \geqq \exp\,[N(C - b_N)], \tag{14}$$

and (for any a priori *distribution on the code words)* error probability

$$P_e \leqq k \exp\,(-\beta N b_N^2), \tag{15}$$

where k is independent of N, and β is defined by (13).

Actually we can broaden the class of channels for which our main result (3) holds to include that class of channels for which Theorem 2 holds for some constant β. This broadened class includes the Gaussian channel with signal-to-noise ratio ρ for which $\beta = [(1 + \rho)^2/\rho(2 + \rho)]$ (see equation 74 of Ref. 4).

Theorems 1 and 2 lead us directly to the proper choice of M and our main result. Since the channel encoder must encode each of the members of $\mathcal{S}$ into channel inputs, we equate the M's of Theorems 1 and 2 obtaining (from $n = NR$)

$$R_0 = C/R \quad \text{and} \quad b_N = Ra_n\,. \tag{16}$$

If we then choose

$$b_N = \sqrt{\frac{\log N}{\beta N}}\,, \tag{17}$$

where β is defined in (13) we have from Theorem 1 a quantization error

$$\epsilon_Q^2 \leqq \sigma^2 \exp\,(-2C/R)\left(1 + \frac{2}{R\sqrt{\beta}}\sqrt{\frac{\log N}{N}}\right) + O\left(\frac{\log N}{N}\right), \tag{18}$$

and from (6), (10), and Theorem 2, a transmission error

$$\epsilon_T^2 \leqq 4\sigma^2 P_e \leqq 4\sigma^2 k\,\frac{1}{N}\cdot \tag{19}$$

Thus by combining (7), (18), and (19) we have an over-all mean-squared error

$$\epsilon^2 \leqq \sigma^2 \exp\,(-2C/R)\left(1 + \frac{2}{R\sqrt{\beta}}\sqrt{\frac{\log N}{N}}\right) + O\left(\frac{1}{\sqrt{N}}\right)\cdot \tag{20}$$

This is our result (3).

III. PROOF OF THEOREM 1

We must establish the existence of a mapping f of Euclidean n-space into a set $\mathcal{S}$ of M n-vectors such that $E\,\|\,\mathbf{X} - f(\mathbf{X})\,\|^2/n$ satisfies the upper bound (9). Let $r_o = E \| \mathbf{X} \|$. It can be shown that with very high probability, as $n \to \infty$, $\mathbf{X}$ will be near the surface of the sphere of radius r_o with center at the origin. It turns out to be convenient to establish the existence of a mapping g from the surface S_{r_o} of this sphere to a set $\mathcal{S}$.

Accordingly we shall construct the mapping f as follows. Let $\mathbf{X}_p = \mathbf{X}(E \| \mathbf{X} \| / \| \mathbf{X} \|) = \mathbf{X}(r_o / \| \mathbf{X} \|)$ be the projection of $\mathbf{X}$ onto the surface S_{r_o}. Let g be a mapping of S_{r_o} to some set $\mathcal{S}$ of n-vectors (g has not yet been found, of course), then $f = g(\mathbf{X}_p)$. The following lemma of D. J. Sakrisan[2], is proved in the Appendix.

Lemma 1: For any mapping g, and $f = g(\mathbf{X}_p)$,

$$E \| \mathbf{X} - f(\mathbf{X}) \|^2 = \text{Var} \| \mathbf{X} \| + E \| \mathbf{X}_p - g(\mathbf{X}_p) \|^2. \tag{21}$$

Since, as we shall see, Var $\| \mathbf{X} \|$ is relatively small, the principal contribution to $E \| \mathbf{X} - f(\mathbf{X}) \|^2$ is $E \| \mathbf{X}_p - g(\mathbf{X}_p) \|^2$.

Our next task is to find the mapping g, and to this end we will establish a lemma concerning the covering of S_{r_o} by spherical caps. First some definitions.

Let $\mathbf{w}$, $\mathbf{z}$ with and without subscripts denote points on S_{r_o}, the surface of a sphere in n-space of radius r_o. Let $\alpha(\mathbf{w}, \mathbf{z})$ be the angle* between $\mathbf{w}$ and $\mathbf{z}$. For $0 \leqq \theta \leqq \pi$, let $\mathcal{C}(\mathbf{w}, \theta) = [\mathbf{z} : \alpha(\mathbf{w}, \mathbf{z}) \leqq \theta]$ be the spherical cap of half angle θ centered at $\mathbf{w}$. Assign the usual "area" measure to S_{r_o}. If $A \subseteq S_{r_o}$ is measurable, let $\mu(A)$ be its measure. In particular, let

$$C_n(\theta) = \mu[\mathcal{C}(\mathbf{w}, \theta)] = \frac{(n-1)\pi^{n-\frac{1}{2}} r_o^{n-1}}{\Gamma\left(\frac{n+1}{2}\right)} \int_0^\theta \sin^{n-2} \varphi \, d\varphi \tag{22}$$

be the area (measure) of a cap of half angle θ†, so that

$$C_n(\pi) = \frac{n\pi^{n/2} r_o^{n-1}}{\Gamma\left(\frac{n+2}{2}\right)} \tag{23}$$

is the area of S_{r_o}. We now state

Lemma 2:‡ Let $\mathbf{X}_p$ be a random vector which is uniformly distributed

* The angle $\alpha(\mathbf{w}, \mathbf{z})$ is defined by $\cos \alpha = (\mathbf{w}, \mathbf{z}) / \| \mathbf{w} \| \, \| \mathbf{z} \|$, where $(\mathbf{w}, \mathbf{z})$ is the inner product and $0 \leq \alpha \leq \pi$.

† It is shown in Ref. 5 that $A_n(r)$, the surface area of an n-sphere of radius r is given by $A_n(r) = n\pi^{n/2} r^{n-1} / \Gamma[(n+2)/2]$. Equation (22) follows from the fact that $C_n(\theta) = \int_0^\theta (r_o d\varphi) A_{n-1} (r_o \sin \varphi)$.

‡ Lemma 2 is related to a result on the covering of the n-sphere in Ref. 6.

on S_{r_o}. Let M (a positive integer) and θ ($0 \leqq \theta \leqq \pi$) be arbitrary. Then (for any dimension n) there exists a set of M points $\{\mathbf{w}_1, \cdots, \mathbf{w}_M\} \subseteq S_{r_o}$ such that

$$Q(\mathbf{w}_1, \cdots, \mathbf{w}_M) \triangleq \Pr\left\{\mathbf{X}_p \notin \bigcup_{j=1}^{M} \mathcal{C}(\mathbf{w}_j, \theta)\right\} \leqq \exp\left[-MC_n(\theta)/C_n(\pi)\right]. \tag{24}$$

Proof: Let us define the function

$$F(\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_M, \mathbf{z}) = \begin{cases} 1 & \text{if} \quad \alpha(\mathbf{w}_j, \mathbf{z}) > \theta, \quad 1 \leqq j \leqq M, \\ 0 & \text{otherwise.} \end{cases} \tag{25}$$

Then for a fixed set $\mathbf{w}_1, \cdots, \mathbf{w}_M$,

$$\begin{aligned} Q(\mathbf{w}_1, \cdots, \mathbf{w}_M) &= \Pr\left\{\mathbf{X}_p \notin \bigcup_{j=1}^{M} \mathcal{C}(\mathbf{w}_j, \theta)\right\} \\ &= EF(\mathbf{w}_1, \cdots, \mathbf{w}_M, \mathbf{X}_p) \\ &= \frac{1}{C_n(\pi)} \int F(\mathbf{w}_1, \cdots, \mathbf{w}_M, \mathbf{z})\, d\mu(\mathbf{z}). \end{aligned} \tag{26}$$

Now consider a random experiment in which the M points $\mathbf{w}_1, \cdots, \mathbf{w}_M$ are random vectors chosen independently with uniform distribution on S_{r_o}. Q is now a random variable given by

$$Q(\mathbf{W}_1, \mathbf{W}_2, \cdots, \mathbf{W}_M) = \frac{1}{C_n(\pi)} \int F(\mathbf{W}_1, \cdots, \mathbf{W}_M, \mathbf{z})\, d\mu(\mathbf{z}), \tag{27}$$

where upper case $\mathbf{W}$'s represent random vectors. We now compute $EQ(\mathbf{W}_1, \cdots, \mathbf{W}_M)$, the average of Q over all choices of $\mathbf{W}_1, \cdots, \mathbf{W}_M$. We will show that $EQ \leqq \exp[-MC_n(\theta)/C_n(\pi)]$. Since at least one set $\{\mathbf{w}_1, \cdots, \mathbf{w}_M\}$ must satisfy $Q(\mathbf{w}_1, \cdots, \mathbf{w}_M) \leqq EQ$, the lemma will follow.

We can write

$$\begin{aligned} EQ &= E \frac{1}{C_n(\pi)} \int F(\mathbf{W}_1, \cdots, \mathbf{W}_M, \mathbf{z})\, d\mu(\mathbf{z}) \\ &= \frac{1}{C_n(\pi)} \int d\mu(\mathbf{z}) EF(\mathbf{W}_1, \cdots, W_M, \mathbf{z}), \end{aligned} \tag{28}$$

the interchange of expectation and integration being justified by the fact that $F \geqq 0$. As indicated in (28), $EF(\mathbf{W}_1, \cdots, \mathbf{W}_M, \mathbf{z})$ is computed with **z** held fixed. Now

$$EF(\mathbf{W}_1, \cdots, \mathbf{W}_M, \mathbf{z}) = \Pr\{F = 1\} = \Pr \bigcap_{j=1}^{M} \{\alpha(\mathbf{W}_j, \mathbf{z}) > \theta\}$$
$$= \left(1 - \frac{C_n(\theta)}{C_n(\pi)}\right)^M \leqq \exp[-MC_n(\theta)/C_n(\pi)], \tag{29}$$

so that

$$E(Q) \leqq \exp[-MC_n(\theta)/C_n(\pi)] \int \frac{d\mu(\mathbf{z})}{C_n(\pi)} = \exp[-MC_n(\theta)/C_n(\pi)] \tag{30}$$

which concludes the proof.

We now give the construction of g. Let $M = M_n = \exp[n(R_o - a_n)]$ as in Theorem 1, and let $\theta = \theta_n = \text{arc sin} \exp(-R_0 + \delta_n)$, where $\delta_n > 0$ will be specified below. Let $\{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_M\}$ be a set which satisfies (24) for these θ and M. Let $\mathbf{x} \in S_{r_o}$ and say $\mathbf{x} \in \bigcup_{j=1}^{M} \mathcal{C}(\mathbf{w}_j, \theta)$. Let j_o be the smallest index j such that $\mathbf{x} \in \mathcal{C}(\mathbf{w}_j, \theta)$. Then

$$g(\mathbf{x}) = (\cos\theta)\mathbf{w}_{j_o}, \tag{31}$$

(see Figure 3), and $\| g(\mathbf{x}) - \mathbf{x} \| \leqq r_o \sin\theta$. If $\mathbf{x} \notin \bigcup_{j=1}^{M} \mathcal{C}(\mathbf{w}_j, \theta)$, then let $g(\mathbf{x}) = \mathbf{w}_1$. In this case $\| \mathbf{x} - g(\mathbf{x}) \| \leqq 2r_o$. Hence

$$E \| \mathbf{X}_p - g(\mathbf{X}_p) \|^2 \leqq r_0^2 \sin^2\theta + 4r_0^2 Q(\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_M). \tag{32}$$

Since the set $\{\mathbf{w}_i\}$ satisfies (24),

$$E \|\mathbf{X}_p - g(\mathbf{X}_p)\|^2 \leqq r_o^2 \sin^2\theta_n + 4r_o^2 \exp[-MC_n(\theta)/C_n(\pi)]. \tag{33}$$

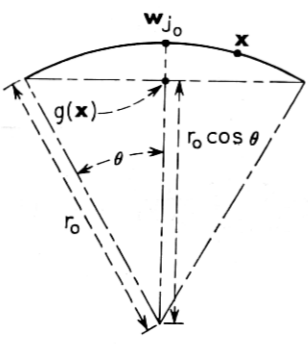

Fig. 3 — Construction of $g(\mathbf{x})$. The solid line is the cap $\mathcal{C}(\mathbf{w}_j, \theta)$.

Combining (33), and Lemma 1, and the fact (proved in the Appendix) that

$$r_o \leqq \sigma\sqrt{n} \quad \text{and} \quad \frac{1}{n}\,\text{var}\,\|\,\mathbf{X}\,\| \leqq 0.92\sigma^2/n, \tag{34}$$

we obtain

$$\frac{1}{n}E\,\|\,\mathbf{X} - f(\mathbf{X})\,\|^2$$
$$\leqq \sigma^2 \sin^2\theta_n + 4\sigma^2 \exp\left[-M_n C_n(\theta_n)/C_n(\pi)\right] + 0.92\sigma^2/n, \tag{35}$$

where $M_n = \exp[n(R_0 - a_n)]$ and $\theta_n = \arcsin\exp(-R_0 + \delta_n)$, where $\delta_n > 0$ is to be specified. Our final step is selection of δ_n so that (9) is satisfied. For θ_n bounded away from $\pi/2$, Shannon (equation 27 of Ref. 4) has shown that as $n \to \infty$,

$$\frac{C_n(\theta_n)}{C_n(\pi)} \sim \frac{1}{\sqrt{2\pi n}}\,\frac{\sin^n\theta_n}{\cos\theta_n\sin\theta_n} > \frac{1}{\sqrt{2\pi n}}\sin^n\theta_n\,, \tag{36}$$

so that (using the definitions of M_n and θ_n) for n sufficiently large

$$M_n\,\frac{C_n(\theta_n)}{C_n(\pi)} \geqq \frac{\exp\left[n(\delta_n - a_n)\right]}{\sqrt{2\pi n}}. \tag{37}$$

We now define δ_n by

$$\delta_n = a_n + \frac{1}{2}\,\frac{\log n}{n} + \frac{\log\log n}{n} + \frac{\log\sqrt{2\pi}}{n}. \tag{38}$$

Then, from (37), for n sufficiently large

$$\exp\left[-M_n C_n(\theta_n)/C_n(\pi)\right] \leqq 1/n. \tag{39}$$

Finally, we have

$$\exp(2\delta_n) = 1 + 2\delta_n + O(\delta_n^2)$$
$$= 1 + 2\delta_n + O(a_n^2) + O\left[\left(\frac{\log n}{n}\right)^2\right]. \tag{40}$$

Combining (35), (39), and (40) we obtain

$$\frac{1}{n}E\,\|\,\mathbf{X} - f(\mathbf{X})\,\|^2$$
$$\leqq \sigma^2 e^{-2R_0}\left(1 + 2a_n + \frac{\log n}{n}\right) + O\left(\frac{\log\log n}{n}\right) + O(a_n^2) \tag{41}$$

which is (9).

Inequality (10) follows from (31) and (34) by simply

$$\| f(\mathbf{X}) \| = \| g(\mathbf{X}_p) \| = r_0 \cos \theta < r_0 < \sqrt{n\sigma^2}. \tag{42}$$

APPENDIX

In this appendix we establish some facts about the random vector $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ whose coordinates are independent zero-mean Gaussian variates with variance σ^2.

Proposition:

(*i*) $E \| \mathbf{X} \| \leqq (n\sigma^2)^{\frac{1}{2}}$
(*ii*) $\text{var} \| \mathbf{X} \| \leqq 0.92\sigma^2$

Proof: (*i*) follows from the Schwarz inequality: $[E(1 \cdot \| X \|)]^2 \leqq E \| \mathbf{X} \|^2 E1^2 = E \| \mathbf{X} \|^2 = n\sigma^2$. To establish (*ii*), consider the n-fold joint probability density for the vector $(R, \Phi_1, \cdots, \Phi_{n-1})$, the polar coordinate representation of $\mathbf{X}$*,

$$g(r, \varphi_1, \varphi_2, \varphi_{n-1}) = \exp (r^2/2\sigma^2) r^{n-1} h(\varphi_1, \cdots, \varphi_{n-1}), \tag{43}$$

where $h(\varphi_1, \cdots, \varphi_{n-1}) = (2\pi\sigma^2)^{-n/2} \cos^{n-2} \varphi_1 \cos^{n-3} \varphi_2 \cdots \cos \varphi_{n-2}$.

The marginal distribution for the random variable $R = \| \mathbf{X} \|$ is

$$f(r) = \frac{2}{(2\sigma^2)^{n/2}\Gamma\left(\frac{n}{2}\right)} r^{n-1} \exp(-r^2/2\sigma^2),$$

so that an integration yields

$$E \| \mathbf{X} \| = E(R) = \sqrt{2\sigma^2}\, \Gamma\left(\frac{n+1}{2}\right) / \Gamma\left(\frac{n}{2}\right).$$

Since $E \| \mathbf{X} \|^2 = n\sigma^2$, we have

$$\text{var} \| \mathbf{X} \| = n\sigma^2 \left\{ 1 - \frac{2}{n} \left[\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \right]^2 \right\}. \tag{44}$$

Using the Stirling formula,

$$e^{-u} u^{u-\frac{1}{2}} \sqrt{2\pi} \left(1 + \frac{1}{12u} - \frac{1}{360u^2}\right) \leqq \Gamma(u) \leqq e^{-u} u^{u-\frac{1}{2}} \sqrt{2\pi} \left(1 + \frac{1}{12u}\right),$$

* See, for example, M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, vol. 1, London: Griffin, 1963, Section 11.2.

to underestimate $\Gamma(n + 1/2)$ and overestimate $\Gamma(n/2)$ we obtain

$$\frac{\text{Var} \,||\, \mathbf{X} \,||}{n\sigma^2} \leqq 1 - e^{-1}\left(1 + \frac{1}{n}\right)^n \left(1 - \frac{7}{36n(n+1)} - \frac{1}{90(n+1)^2}\right)^2. \quad (45)$$

Further

$$e^{-1}\left(1 + \frac{1}{n}\right)^n = \exp\left[-1 + n \log(1 + 1/n)\right]$$
$$\geqq \exp\left[-1 + n(1/n - 1/2n^2)\right]$$
$$= \exp(-1/2n) \geqq 1 - \frac{1}{2n},$$

and

$$\frac{1}{36n(n+1)} + \frac{1}{90(n+1)^2} \leqq \frac{1}{n^2}\left(\frac{7}{36} + \frac{1}{90}\right) \leqq \frac{0.21}{n^2}.$$

Hence,

$$\frac{\text{Var} \,||\, \mathbf{X} \,||}{n\sigma^2} \leqq 1 - (1 - 0.5n^{-1})(1 - 0.42n^{-2}) \leqq \frac{0.5}{n} + \frac{0.42}{n^2}$$
$$= \frac{0.5}{n}\left(1 + \frac{0.84}{n}\right) \leqq \frac{0.5(1.84)}{n} = \frac{.92}{n}. \quad (46)$$

This is (*ii*).

We now give a

Proof of Lemma 1:

$$E \,||\, \mathbf{X} - f(\mathbf{X}) \,||^2 = E \,||\, \mathbf{X} - \mathbf{X}_p + \mathbf{X}_p - g(\mathbf{X}_p) \,||^2$$
$$= E \,||\, \mathbf{X} - \mathbf{X}_p \,||^2 + E \,||\, \mathbf{X}_p - g(\mathbf{X}_p) \,||^2$$
$$+ 2E(\mathbf{X} - \mathbf{X}_p \,,\, \mathbf{X}_p - g(\mathbf{X}_p)), \quad (47)$$

where $(\mathbf{u}, \mathbf{v})$ is the inner product of the n-vectors $\mathbf{u}$ and $\mathbf{v}$. Now $||\mathbf{X} - \mathbf{X}_p \,||^2 = \{||\, \mathbf{X} \,|| - E \,||\, \mathbf{X} \,||\}^2$, so that

$$E \,||\, \mathbf{X} - \mathbf{X}_p \,||^2 = \text{var} \,||\, \mathbf{X} \,||. \quad (48)$$

Further, the inner product

$$((\mathbf{X} - \mathbf{X}_p \,,\, \mathbf{X}_p - g(\mathbf{X}_p))$$
$$= (\mathbf{X}, \mathbf{X}_p) - (\mathbf{X}, g(\mathbf{X}_p)) - (\mathbf{X}_p \,,\, \mathbf{X}_p) + (\mathbf{X}_p \,,\, g(\mathbf{X}_p))$$
$$= ||\, \mathbf{X} \,||\, r_0 - r_0^2 - ||\, \mathbf{X} \,||\, ||\, g(\mathbf{X}_p) \,|| \cos \alpha_1 + r_0 \,||\, g(\mathbf{X}_p) \,|| \cos \alpha_2 \,, \quad (49)$$

where α_1 is the angle between $\mathbf{X}$ and $g(\mathbf{X}_p)$, and α_2 is the angle between $\mathbf{X}_p$ and $g(\mathbf{X}_p)$. Since $\mathbf{X}$ and $\mathbf{X}_p$ are colinear, $\alpha_1 = \alpha_2 = \alpha(\mathbf{X}_p)$ a function of $\mathbf{X}_p$. Now from (43) we see that the random variable $R = \| \mathbf{X} \|$ and the vector $(\Phi_1 , \cdots \Phi_{n-1})$ are statistically independent. Since $\mathbf{X}_p$, depends only on $\Phi, \cdots \Phi_{n-1}$ and not on R, we conclude that $\| \mathbf{X} \|$ is independent of $g(\mathbf{X}_p)$ and $\alpha(\mathbf{X}_p)$. Thus from (49)

$$E(\mathbf{X} - \mathbf{X}_p , \mathbf{X}_p - g(\mathbf{X}_p))$$
$$= r_0 E \| \mathbf{X} \| - r_0^2 - E \| \mathbf{X} \| E[\| g(\mathbf{X}_p) \| \cos \alpha(\mathbf{X}_p)]$$
$$+ r_0 E[\| g(\mathbf{X}_p) \| \cos \alpha(\mathbf{X}_p)] = 0. \tag{50}$$

Equations (47), (48) and (50) imply Lemma 1.

REFERENCES

1. Shannon, C. E., "Coding Theorems for a Discrete Source with a Fidelity Criterion," *1959 IRE Convention Record* Part 4, pp. 142–163.
2. Sakrisan, D. J., "A Geometric Treatment of the Problem of Source Encoding a Gaussian Randon Variable," Tech. Rep., Space Sciences Laboratory, University of California, Berkeley, 1967; also presented at the Int. Symp. Inform. Theory, San Remo, Italy, September 1967.
3. Shannon, C. E., "Certain Results in Coding Theory for Noisy Channels," Inform. Control, *1,* (September 1957), pp. 6–25.
4. Shannon, C. E., "Probability of Error for Optimal Codes in the Gaussian Channel," B.S.T.J., *38,* No. 5 (May 1959), pp. 611–656.
5. Wozencraft, J. and Jacobs, I., *Principles of Communication Engineering,* New York: John Wiley and Sons, Inc., 1965, pp. 355–357.
6. Wyner, A. D., "Random Packings and Coverings of the Unit n-Sphere, B.S.T.J., *46,* No. 9 (November 1967), pp. 2111–2118.