THE CAPACITY OF A CHANNEL WITH ARBITRARILY VARYING ADDITIVE GAUSSIAN CHANNEL PROBABILITY FUNCTIONS

RUDOLF AHLSWEDE

COLUMBUS and URBANA

SUMMARY

Let $X = R$ be the "input alphabet" and $Y = R$ be the "output alphabet", where R denotes the set of real numbers. Let $X^t = X$ and $Y^t = Y$ for $t = 1, 2, \ldots, n$

$$X_n = \prod_{t=1}^{n} X^t \quad \text{and} \quad Y_n = \prod_{t=1}^{n} Y^t.$$

For $\sigma^* \geqq \sigma_* \geqq 0$ let $S = \{s \mid \sigma^* \geqq s \geqq \sigma_*\}$ and let

$$f(y \mid x \mid s) = \frac{1}{s\sqrt{(2\pi)}} \exp\left[-\frac{(x-y)^2}{2s^2}\right] \quad \text{for all} \quad x \in X,\ y \in Y,\ s \in S.$$

For every $s_n = (s^1, \ldots, s^n) \in \prod_1^n S$ define $P(\cdot \mid \cdot \mid s_n)$ by

$$P(A \mid x_n \mid s_n) = \int_A \prod_{t=1}^{n} \frac{1}{s^t\sqrt{(2\pi)}} \exp\left[-\frac{(x^t - y^t)}{2(s^t)^2}\right] dy^1, \ldots, dy^n$$

for every $x_n = (x^1, \ldots, x^n) \in X_n$ and every Borel set $A \subset Y_n$. Consider the channel $\mathscr{C}_n = \{P(\cdot \mid \cdot \mid s_n) \in S_n\}$ with transmission probability densities $f(\cdot \mid \cdot \mid s)$ varying arbitrarily from "letter" to "letter".

The author determines the capacity of this channel when the code words satisfy

(a) an average power constraint and

(b) an amplitude constraint.

I. INTRODUCTION

Let R be the set of real numbers and let $X = R$ be the "input alphabet" and $Y = R$ be the "output alphabet" of the channels we shall study below. Let $X^t = X$ and $Y^t =$

$= Y$ for $t = 1, \ldots, n$. By $X_n = \prod_{t=1}^{n} X^t$ we denote the set of input n-sequences (words of length n) and by $Y_n = \prod_{t=1}^{n} Y^t$ we denote the set of output n-sequences. We define an additive Gaussian channel probability density function by

$$f(y \mid x \mid \sigma) = \frac{1}{\sigma\sqrt{(2\pi)}} \exp\left[-\frac{(x-y)^2}{2\sigma^2}\right] \tag{1.1}$$

for all $x \in X$, $y \in Y$.

The probability density functions for words of length n are given by

$$f_\sigma(y_n \mid x_n) = \frac{1}{[\sigma\sqrt{(2\pi)}]^n} \exp\left[-\sum_{t=1}^{n} \frac{(y^t - x^t)^2}{2\sigma^2}\right] \tag{1.2}$$

for all $x_n = (x^1, \ldots, x^n) \in X_n$ and all $y_n = (y^1, \ldots, y^n) \in Y_n$.

Thus the output sequence is obtained from the input sequence by the addition of a sequence of independent Gaussian random variables with mean 0 and variance σ^2. For notational simplicity we avoid the definition of the random variables involved.

The transition probabilities of an additive Gaussian memoryless channel G_σ are defined by

$$P_\sigma(A \mid x_n) = \int_A f_\sigma(y_n \mid x_n)\, dy^1 \ldots dy^n \tag{1.3}$$

for every $x_n \in X_n$ and every Borel set $A \subset Y_n$, $n = 1, 2, \ldots$

A code (n, N) for channel G_σ is a system of pairs $\{(u_i, A_i) \mid i = 1, \ldots, N\}$, where $u_i \in X_n$ for $i = 1, \ldots, N$, and the $A_i (i = 1, \ldots, N)$ are disjoint Borel sets in Y_n. A code (n, N, λ) is a code (n, N) with maximum probability of error $\leqq \lambda$, that is, a code (n, N) satisfying

$$P_\sigma(A_i \mid u_i) \geqq 1 - \lambda \quad \text{for} \quad i = 1, \ldots, N\,. \tag{1.4}$$

N is called the length of the code. It is easy to see that if no restrictions are placed on the choice of code words arbitrarily large N can be obtained for every n and λ. However, the inputs to a channel are usually required to satisfy certain constraints, depending on the circumstances which lead to the channel model. Two constraints which are useful are an amplitude constraint, that is, for every $u_i = (u_i^1, \ldots, u_i^n)$, $i = 1, \ldots, N$, we have

$$|u_i^t| \leqq A\,, \quad t = 1, \ldots, n\,, \tag{1.5}$$

and an average power constraint E, that is,

$$\sum_{t=1}^{n} (u_i^t)^2 \leqq n \,.\, E \quad \text{for} \quad i = 1, \ldots, N\,. \tag{1.6}$$

We denote the Gaussian channel G_σ with an amplitude constraint by AG_σ and the Gaussian channel G_σ with an average power constraint by EG_σ. The capacity for the channel EG_σ, first deduced by Shannon [8], is

$$\bar{C}_\sigma = \frac{1}{2}\log\left(1 + \frac{E}{\sigma^2}\right). \tag{1.7}$$

A coding theorem and its strong converse for channel AG_σ have been proved by Kemperman ([4], [11] ch. 9).

We introduce now a channel with arbitrarily varying additive Gaussian channel probability functions, which we abbreviate thus: a.v.G. The channel a.v.G. with an amplitude constraint shall be denoted by a.v.AG and the channel a.v.G with an average power constraint shall be denoted by a.v.EG. Let S be a closed interval $[\sigma_*, \sigma^*]$ of non-negative real numbers, and let $S_n = \prod_1^n S$ for $n = 1, 2, \ldots$ For every n-sequence $s_n = (s^1, \ldots, s^n) \in S_n$ we define $f(\cdot|\cdot|\, s_n)$ by

$$f(y_n \mid x_n \mid s_n) = \prod_{t=1}^{n} \frac{1}{s^t\sqrt{(2\pi)}} \exp\left[-\frac{(y^t - x^t)^2}{2(s^t)^2}\right] \tag{1.8}$$

for all $x_n = (x^1, \ldots, x^n) \in X_n$ and all $y_n = (y^1, \ldots, y^n) \in Y_n$, and $P(\cdot|\cdot|\, s_n)$ by

$$P(A \mid x_n \mid s_n) = \int_A f(y_n \mid x_n \mid s_n)\, dy^1 \ldots dy^n \tag{1.9}$$

for every $x_n \in X_n$ and every Borel set $A \subset Y_n$, $n = 1, 2, \ldots$.

The channel a.v.G is defined by the sequence $(\mathscr{C}_n)_{n=1,2,\ldots}$ where

$$\mathscr{C}_n = \{P(\cdot|\cdot|\, s_n) \mid s_n \in S_n\}. \tag{1.10}$$

Suppose that sender and receiver want to communicate over the channel a.v.G without knowing which n-sequences s_n will govern the transmission of any word (input n-sequence). An (n, N) code is an (n, N, λ) code in this case, if

$$P(A_i \mid u_i \mid s_n) \geqq 1 - \lambda, \quad \text{for} \quad i = 1, \ldots, N, \quad \text{and for all} \quad s_n \in S_n. \tag{1.11}$$

A number C is called the capacity of the channel a.v.AG if, for any $\varepsilon > 0$ and any λ, $0 < \lambda < 1$, the following is true for all n sufficiently large:

There exists a code $(n, \exp\{n(C - \varepsilon)\}, \lambda)$ and there does not exist a code $(n, \exp\{n(C + \varepsilon)\}, \lambda)$. Analogously we define the capacity $\bar{C}$ of the channel a.v.EG. We prove in Section 3 that C and $\bar{C}$ exist and we give explicit formulas for them.

The present channel model seems to be very realistic from a practical point of view. In practice the variance of a Gaussian channel is never precisely known, may be different for different input letters and may also vary in time.

In case of finite alphabets, channels with arbitrarily varying channel probability functions were studied in [5], [2], [1]. In case of an output alphabet of size greater than 2, no formula for the capacity is known (see [1], [2]). The channel a.v.G has certain symmetry properties which make it possible to obtain formulas for C and $\bar{C}$ by a rather simple derivation. Our results can easily be extended to the time-continuous case. For the definition of a time-continuous Gaussian channel and for its coding theorem see [3].

II. AUXILIARY RESULTS

We formulate first a result by Kemperman ([4], [11] ch. 9) for the channel AG_σ. In the sequel we will choose the input alphabet X of the channel AG_σ to be the interval $[0, 1]$. It is easy to see that this assumption can be made without loss of generality.

Let a be a positive integer and let

$$X(a) = \{(i - \tfrac{1}{2})\, a^{-1} \mid i = 1, \ldots, a\}\,. \tag{2.1}$$

Define $C_\sigma(a)$ as

$$C_\sigma(a) = \max_{\pi} \sum_{x \in X(a)} \pi(x) \int_Y f(y \mid x \mid \sigma) \log \sum_{z \in X(a)} \frac{f(y \mid x \mid \sigma)}{\pi(z) f(y \mid z \mid \sigma)}\, \mathrm{d}y\,. \tag{2.2}$$

(The maximum is taken over all probability distributions on $X(a)$.) $C_\sigma(a)$ is the capacity of the semicontinuous channel (see [11], ch. 8), which we obtain, if we restrict the input alphabet of AG_σ to the set $X(a)$.

Define C_σ by

$$C_\sigma = \sup_a C_\sigma(a)\,. \tag{2.3}$$

We can now formulate

THEOREM K (Kemperman, [4], [11] ch. 9). *Let $0 < \lambda < 1$ and $\varepsilon > 0$ be arbitrary. For all n sufficiently large there exists a code $(n, \exp\{n(C_\sigma - \varepsilon)\}, \lambda)$, and there does not exist a code $(n, \exp(C_\sigma + \varepsilon)\}, \lambda)$, for the channel AG_σ.*

We give now further definitions and state and prove two Lemmas which we shall need in section 3.

The (n, N) code $\{(u_i, A_i) \mid i = 1, \ldots, N\}$ is a strict maximum likelihood code (s.m.l.c.) with respect to $P_\sigma(\cdot \mid \cdot)$ if

$$A_i = \{y_n \mid f_\sigma(y_n \mid u_i) > f_\sigma(y_n \mid u_j) \text{ for } j \neq i\}\,, \quad i = 1, \ldots, N\,. \tag{2.4}$$

$d(\cdot, \cdot)$ shall denote the distance in an n-dimensional Euclidean space E^n.

LEMMA 1 (Shannon [8]). $\{(u_i, A_i) \mid i = 1, \ldots, N\}$ *is a* s.m.l.c. *with respect to* $P_\sigma(\cdot|\cdot)$ *if and only if*

$$A_i = \{y_n \mid d(u_i, y_n) < d(u_j, y_n) \text{ for } j \neq i\}, \quad i = 1, \ldots, N.$$

The u_i's and the y_n's are viewed as elements of the same n-dimensional Euclidean space.

Lemma 1 says that for the channel G_σ maximum likelihood decoding is equivalent to minimal distance decoding.

Proof of Lemma 1. Write $u_i = (u_i^1, \ldots, u_i^n)$ for $i = 1, \ldots, N$. The inequality

$$\frac{1}{[\sigma\sqrt{(2\pi)}]^n} \exp\left[-\sum_{t=1}^{n} \frac{(u_i^t - y^t)^2}{2\sigma^2}\right] > \frac{1}{[\sigma\sqrt{(2\pi)}]^n} \exp\left[-\sum_{t=1}^{n} \frac{(u_j^t - y^t)^2}{2\sigma^2}\right]$$

holds if and only if the inequality $d(u_i, y_n) < d(u_j, y_n)$ holds.

LEMMA 2. *If* $\{(u_i, A_i) \mid i = 1, \ldots, N\}$ *is a* s.m.l.c. *for* $P_{\sigma^*}(\cdot|\cdot)$, *then*

$$P(A_i \mid u_i \mid s_n) \geqq P_{\sigma^*}(A_i \mid u_i) \tag{2.5}$$

for all $i = 1, \ldots, N$ *and all* $s_n \in S_n$.

Proof. Minimal distance decoding results in a partitioning of E^n into n-dimensional polyhedra, or polytopes, around the different signal points, each polyhedron bounded by a finite number (not more than $N - 1$) of $(n - 1)$-dimensional hyperplanes. Denote the polyhedra by $H_1, \ldots, H_N$. A_i equals the interior of H_i, $i = 1, \ldots, N$. From (1.2), (1.3), (1.8) and (1.9) we have that for any i

$$P_{\sigma^*}(A_i \mid u_i) = \int_{A_i} \frac{1}{(\sigma^*)^n (2\pi)^{n/2}} \prod_{t=1}^{n} \exp\left[-\frac{(y^t - u_i^t)^2}{2(\sigma^*)^2}\right] dy^1 \ldots dy^n \tag{2.6}$$

and

$$P(A_i \mid u_i \mid s_n) = \int_{A_i} \frac{1}{(2\pi)^{n/2}} \prod_{t=i}^{n} \exp\left[-\frac{(y^t - u_i^t)^2}{2(s^t)^2}\right] dy^1 \ldots dy^n. \tag{2.7}$$

Applying the affine transformation T_i given by

$$y^t - u_i^t = w^t, \quad t = 1, \ldots, n, \tag{2.8}$$

to (2.6) and (2.7) yields

$$P_{\sigma^*}(A_i \mid u_i) = \int_{T_i A_i} \frac{1}{(\sigma^*)^n (2\pi)^{n/2}} \prod_{t=1}^{n} \exp\left[-\frac{(w^t)^2}{2(\sigma^*)^2}\right] dw^1 \ldots dw^n \tag{2.9}$$

and

$$(2.10)\qquad P(A_i \mid u_i \mid s_n) = \int_{T_iA_i} \frac{1}{(2\pi)^{n/2}} \prod_{t=1}^{n} \frac{1}{s^t} \exp\left[-\frac{(w^t)^2}{2(s^t)^2}\right] dw^1 \ldots dw^n .$$

T_iA_i is the interior of a polyhedron containing the origin. Now we apply the transformation S_i, given by

$$(2.11)\qquad \frac{w^t}{s^t} = \frac{z^t}{\sigma^*}, \quad t = 1, \ldots, n ,$$

to (2.10) and obtain

$$(2.12)\quad P(A_i \mid u_i \mid s_n) = \int_{S_iT_iA_i} \frac{1}{(\sigma^*)^n (2\pi)^{n/2}} \prod_{t=1}^{n} \exp\left[-\frac{z^t}{2(\sigma^*)^2}\right] dz^1 \ldots dz^n .$$

Since $\sigma^*/s^t \geqq 1$, for $t = 1, \ldots, n$, $S_iT_iA_i \supset T_iA_i$. This, (2.9) and (2.12) imply

$$P(A_i \mid u_i \mid s_n) \geqq P_{\sigma^*}(A_i \mid u_i) \quad \text{for} \quad i = 1, \ldots, N .$$

III. THE MAIN RESULTS

This paper started with the observation that the Schalkwijk coding scheme for the channel EG_σ with feedback (see [6], [7], [10]) can be used for the channel a.v.EG with *feedback* and yields a coding theorem with capacity $C_f = \frac{1}{2}\log(1 + E/(\sigma^*)^2)$ that is the capacity of the channel EG_{σ^*}. Only the largest occuring variance $(\sigma^*)^2$ matters.

Theorem 1 below says that this is also true for the channel a.v.EG (without feedback). For the channel AG_σ with feedback there seems to exist until now no optimal coding scheme "similar" to the Schalkwijk scheme for the channel EG_σ with feedback, but still we can determine the capacity for the channel a.v.AG with and without feedback (Corollary and Theorem 2 below).

THEOREM 1 (Coding theorem and strong converse for the channel a.v.EG). *Let* $\bar{C} = \frac{1}{2}\log(1 + E/(\sigma^*)^2)$, $0 < \lambda < 1$ *and* $\varepsilon > 0$, *arbitrary otherwise. For all n sufficiently large*

a) *there exists a code*

$$(n, \exp\{[\bar{C} - \varepsilon]\, n\}, \lambda) ,$$

and

b) *there does not exist a code*

$$(n, \exp\{(\bar{C} + \varepsilon]\, n\}, \lambda)$$

for channel a.v.EG.

Proof. b) is obvious from the definition of the channel a.v.EG and the strong converse for the channel EG_{σ^*} ([11], ch. 9).

We prove now a). The coding theorem for the channel EG_{σ^*} ([8], [11] ch. 9) yields that for all n large enough there exists a s.m.l.c. $\{(u_i, A_i) \mid i = 1, \ldots, K\}$, where $u_i \in X_n$, $A_i \subset Y_n$ for $i = 1, \ldots, K$, such that

$$\sum_{t=1}^{n} (u_i^t)^2 \leqq E \cdot n \quad \text{for} \quad i = 1, \ldots, n, \tag{3.1}$$

$$\frac{1}{K} \sum_{i=1}^{K} P_{\sigma^*}(A_i \mid u_i) \geqq 1 - \frac{\lambda}{2} \tag{3.2}$$

and

$$K \geqq 2 \exp \{[\bar{C} - \varepsilon] n\} . \tag{3.3}$$

(3.2) and (3.3) imply the existence of a subcode $\{(u_{i_1}, A_{i_1}), \ldots, (u_{i_N}, A_{i_N})\}$ satisfying

$$P_{\sigma^*}(A_{i_v} \mid u_{i_v}) \geqq 1 - \lambda \tag{3.4}$$

for $v = 1, \ldots, N$ and

$$N \geqq \exp \{[\bar{C} - \varepsilon] n\} . \tag{3.5}$$

(3.4) and Lemma 2 implies that

$$P(A_{i_v} \mid u_{i_v} \mid s_n) \geqq 1 - \lambda \tag{3.6}$$

for $v = 1, \ldots, N$ and all $s_n \in S_n$.

THEOREM 2 (Coding theorem and strong converse for the channel a.v.AG). *Let* $C = C_{\sigma^*}$ (*defined in* (2.3)), $0 < \lambda < 1$ *and* $\varepsilon > 0$. *For all n sufficiently large*

a) *there exists a code*

$$(n, \exp \{[C - \varepsilon] n\}, \lambda\}$$

and

b) *there does not exist a code*

$$(n, \exp \{[C + \varepsilon] n\} \lambda)$$

for the channel a.v.AG.

Proof. b) is obvious from the definition of the channel a.v.AG and Theorem K.

For all n large enough Theorem K yields that there exists a s.m.l.c. $\{(u_i, A_i) \mid i = 1, \ldots, K\}$ for the channel AG_{σ^*} such that

$$0 \leqq u_i^t \leqq 1 \quad \text{for} \quad t = 1, \ldots, n; \; i = 1, \ldots, K ; \tag{3.7}$$

(3.8) $$\frac{1}{K}\sum_{i=1}^{K} P_{\sigma^*}(A_i \mid u_i) \geqq 1 - \frac{\lambda}{2}$$

and

(3.9) $$K \geqq 2 \exp\{[C - \varepsilon] n\} .$$

(3.8) and (3.9) imply the existence of a subcode $\{(u_{i_1}, A_{i_1}), \ldots, (u_{i_N}, A_{i_N})\}$ satisfying

(3.10) $$P_{\sigma^*}(A_{i_v} \mid u_{i_v}) \geqq 1 - \lambda \quad \text{for} \quad v = 1, \ldots, N$$

and

(3.11) $$N \geqq \exp\{[C - \varepsilon] n\} .$$

a) follows now from (3.10), Lemma 2 and (3.11).

COROLLARY. *The capacity of the channel* a.v.AG *with feedback equals C.*

Proof. The coding theorem is a consequence of Theorem 2, because feedback could only increase the capacity. The strong converse of the coding theorem follows from the strong converse for channel AG_{σ^*} with feedback ([10]).

REMARK. The main tool for proving Theorems 1, 2 is Lemma 2. This Lemma can also be used to extend the error bounds for G_{σ^*}, obtained in [8], and the results of [9] to a.v.G.

REFERENCES

[1] AHLSWEDE, R.: A note on the existence of the weak capacity for channels with arbitrarily varying channel probability functions and its relation to Shannon's zero error capacity. Ann. of Math. Stat. *41* (1970), 3, 1027—1033.

[2] AHLSWEDE, R., WOLFOWITZ, J.: The capacity of channels with arbitrarily varying channel probability functions and binary output alphabet. Z. Wahrscheinlichkeitstheorie verw. Geb. *15* (1970), 186—194.

[3] ASH, R. B.: Capacity and error bounds for a time continuous Gaussian channel. Information and Control *6* (1963), 14—27.

[4] KEMPERMAN, J. H. B.: On the optimum rate of transmitting information. Ann. of Math. Stat. *40* (1969), 6, 2156—2177.

[5] KIEFER, J., WOLFOWITZ, J.: Channels with arbitrarily varying channel probability functions. Information and Control *5* (1968), 44—54.

[6] SCHALKWIJK, F. P. M, KAILATH, T.: A coding scheme for additive noise channels with feedback. Part I: No band with constraint. IEEE Trans. Inf. Theory *IT-12* (1966), 172—182.

[7] SCHALKWIJK, F. P. M.: A coding scheme for additive noise channels with feedback. Part II: Band limited signals. IEEE Trans. Inform. Theory *IT-12* (1966), 183—189.

[8] SHANNON, C. E.: Probability of error for optimal codes in a Gaussian channel. Bell System Tech. J. *38* (1959), 3, 611—656.

[9] SINAI, Y. G.: The least error and best method of transmitting stationary messages with linear encoding and decoding in the case of Gaussian communication channels (in Russian). Problems of Information Transmission 2, 40—49, Izd. AN SSSR, Moscow 1959.

[10] WOLFOWITZ, J.: Note on the Gaussian channel with feedback and a power constraint. Information and Control *12* (1968), 71—78.

[11] WOLFOWITZ, J.: Coding Theorems of Information Theory. 2nd ed. Springer, Berlin—Heidelberg—New York 1964.

OHIO STATE UNIVERSITY

and

UNIVERSITY OF ILLINOIS