INFORMATION AND CONTROL 18, 203–219 (1971)

The Information Capacity of Amplitude- and Variance-Constrained Scalar Gaussian Channels*

JOEL G. SMITH

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91103

The amplitude-constrained capacity of a scalar Gaussian channel is shown to be achieved by a unique discrete random variable taking on a finite number of values. Necessary and sufficient conditions for the distribution of this random variable are obtained. These conditions permit determination of the random variable and capacity as a function of the constraint value. The capacity of the same Gaussian channel subject, additionally, to a nontrivial variance constraint is also shown to be achieved by a unique discrete random variable taking on a finite number of values. Likewise, capacity is determined as a function of both amplitude- and variance-constraint values.

NOMENCLATURE

A	Input amplitude constraint
$C(A)$	Amplitude-constrained capacity
D	Channel uncertainty
E_0	Points of increase of F_0
F	Input probability distribution function
$\mathscr{F}_A$	Space of input distribution functions
F_0	Optimal input distribution
$H(F)$	Output entropy function
$I(F)$	Average mutual information functional
$I'_{F_1}(F)$	Weak derivative of the information functional at F_1
p_N, p_Y	Channel noise and output probability density functions
X, N, Y	Channel input, noise, and output random variables
$d(F_1, F_2)$	Lèvy distance between two distributions F_1 and F_2 in $\mathscr{F}_A$
$h(x; F)$	Marginal entropy density
$i(x; F)$	Marginal information density

* Research conducted in partial satisfaction for the requirements for the degree of Doctor of Philosophy at the University of California, Berkeley, California, and supported by The Joint Services Electronics Program under Contract APOSR 68-1488.

© 1971 by Academic Press, Inc.

643/18/3-1

I. Introduction

Although the literature on the average power-constrained information capacity of continuous channels is extensive (at least for Gaussian channels), the literature on the peak-power or amplitude-limited capacity is rather sparse. Shannon (1948) obtained a loose lower bound for strictly-bandlimited Gaussian white noise channels as well as asymptotic results for large and small ratios of peak signal power to average noise power; Goldman (1953) provided a corrected version. Various mild efforts of mixed success have appeared in the optical and photographic literature dealing with the information capacity of photographic channels (inherently amplitude-limited by the fixed dynamic range of the photographic material). Gallager (1968) has verified the coding theorem and its converse for amplitude-limited channels (memoryless, discrete time only), and derived the capacity for a few simple examples. Nothing, to the author's knowledge, has appeared in the literature concerning the capacity of channels subject to both peak and average power constraints.

This paper determines, separately, the amplitude-constrained capacity and the amplitude- and variance-constrained capacity of a scalar Gaussian channel. Section II develops all preliminary concepts. Section III treats the amplitude-constrained scalar Gaussian channel.

Capacity, as used in this paper, is defined as the supremum of the mutual information functional over the appropriate class of input probability distributions. It is shown that a unique input distribution, called the "optimal" input, exists which achieves capacity. Necessary and sufficient conditions for this optimal input are then obtained by the application of a simple optimization theorem. These conditions are used to establish that the optimal input random variable is discrete taking on a finite number of values. Thus the capacity, for a fixed amplitude limit, is formulated as the maximum of a function of a finite-dimensional vector, the components of which are the points of increase of the optimal probability-distribution function and the corresponding amounts of increase of this function at each point of increase. A computer program of a standard optimization technique is utilized to determine the capacity and the optimal distribution at each of a large number of values of amplitude constraint. The resulting capacity is given graphically as a function of this constraint.

Section IV treats the case where both amplitude and nontrivial variance constraints are imposed. The investigation parallels Section II sufficiently, so that little added effort is involved. The capacity is given graphically as a function of the two constraints.

II. Preliminaries

Consider a scalar additive channel characterized by the expression $Y = X + N$ where X, N, and Y denote the channel input, noise, and output random variables, respectively. Let p_N and p_Y denote, respectively, the probability density functions of N and Y, and F_X (or simple F) denote the probability distribution function of X. The random variable N is assumed to be Gaussian, with zero mean and unit variance. The random variable X is assumed to be constrained to take on values a.s. on $[-A, A]$ for some arbitrary positive value of A.[1] Let $\mathscr{F}_A$ denote the corresponding class of distribution functions F; i.e., F in $\mathscr{F}_A$ implies $F(x) = 0$ for all $x < -A$ and $F(x) = 1$ for all $x > A$. The existence of p_Y follows (Smith, 1969) from the existence of p_N, and furthermore

$$p_Y(y) = \int_{-A}^{A} p_N(y - x)\, dF(x).$$

Conventionally, the average mutual information I between two random variables X and Y is denoted $I(X; Y)$. For an additive noise channel, however, the output random variable Y is the sum of the input and noise random variables X and N. Thus $I(X; Y)$ is (for a fixed channel) a function of the input random variable X only, or, equivalently, a function of the probability distribution function of the input random variable. Occasionally, in the literature, when different input distributions are considered, I is subscripted; e.g., $I_F(X; Y)$, where F is a particular input distribution. Here, $I_F(X; Y)$ is written as $I(F)$, treating the average mutual information as a functional on the space $\mathscr{F}_A$ of probability distribution functions F of the input random variable X.

For sufficiently well-behaved channels, the average mutual information can be written as the difference of two entropy functions: $I(X; Y) = H(Y) - H(Y \mid X)$ or, if a particular input probability distribution function must be specified, $I_F(X; Y) = H_F(Y) - H_F(Y \mid X)$. For an additive channel, the latter term is a constant, denoted here by D, and $H_F(Y)$ then can be denoted $H(F)$. No confusion should arise as this output entropy term is the only entropy term used except for the constant noise entropy D. In addition,

[1] Assume N' has mean μ and variance σ_N^2, X' is constrained a.s. on $[a, b]$, and $Y' = X' + N'$. Let $A \triangleq (b - a)/2\sigma_N$, $X \triangleq [X' - (a + b)/2)]/\sigma_N$, $N \triangleq (N' - \mu)/\sigma_N$, and $Y \triangleq X + N$. It can be shown that $I(X; Y) = I(X'; Y')$. Hence, in the above "normalized" case, A actually represents the ratio signal amplitude/noise standard deviation.

the output density function p_Y, is a (possibly) different function of y for each input distribution function F. Hence, when necessary, $p_Y(y)$ will be written as $p_Y(y; F)$ to indicate the input distribution F that determined the output density function p_Y.

A definition of average mutual information $I(F)$ between channel input and output which suffices for an additive channel characterized by a density function p_N is (Pinsker, 1964) for all F in $\mathscr{F}_A$

$$I(F) \triangleq \int_{-\infty}^{\infty} \int_{-A}^{A} p_N(y - x) \log \frac{p_N(y - x)}{p_Y(y; F)} \, dF(x) \, dy.$$

For noise with finite variance and a bounded density function, I can be written (Ash, 1965) as the difference of two finite entropy functions:

$$I(F) = H(F) - D,$$

where $H(F)$, the output entropy, is, for all F in $\mathscr{F}_A$,

$$H(F) \triangleq - \int_{-\infty}^{\infty} p_Y(y; F) \log p_Y(y; F) \, dy,$$

and D, the noise entropy, is

$$D \triangleq - \int_{-\infty}^{\infty} p_N(z) \log p_N(z) \, dz$$

which is $1/2 \log(2\pi e)$ for this channel. The amplitude-constrained channel capacity C is defined to be

$$C(A) \triangleq \sup_{F \text{ in } \mathscr{F}_A} I(F).$$

The marginal information density and marginal entropy density, are defined, respectively, by

$$i(x; F) \triangleq \int_{-\infty}^{\infty} p_N(y - x) \log \frac{p_N(y - x)}{p_Y(y; F)} \, dy,$$

and

$$h(x; F) \triangleq - \int_{-\infty}^{\infty} p_N(y - x) \log p_Y(y; F) \, dy$$

for all x in $[-A, A]$ and for all F in $\mathscr{F}_A$. Thus (Smith, 1969), the following equations hold for all F in $\mathscr{F}_A$:

$$i(x;F) = h(x;F) - D, \qquad \text{for all } x \text{ in } [-A, A],$$

$$I(F) = \int_{-A}^{A} i(x;F)\, dF(x),$$

and

$$H(F) = \int_{-A}^{A} h(x;F)\, dF(x).$$

The results of this paper rely on a fairly simple bit of optimization theory. The necessary definitions and theory are introduced in a general notation and then a connection drawn to relate this material to the information theory problem of interest.

Let Ω be a convex space, f a function from Ω into the real line $\mathscr{R}$, x_0 a fixed element of Ω, and θ a number in $[0, 1]$. Suppose there exists a map $f'_{x_0} : \Omega \to \mathscr{R}$ such that

$$f'_{x_0}(x) \triangleq \lim_{\theta \downarrow 0} \left\{ \frac{f[(1-\theta)\, x_0 + \theta x] - f(x_0)}{\theta} \right\}, \qquad \text{for all } x \text{ in } \Omega.$$

Then f is said to be weakly differentiable in Ω at x_0, and f'_{x_0} is the weak derivative in Ω at x_0. If f is weakly differentiable in Ω at x_0 for all x_0 in Ω, f is said to be weakly differentiable in Ω, or simply weakly differentiable. Furthermore, f is said (Gallager, 1968) to be convex cap (concave in some references) if for all x_0 and x in Ω, and for all θ in $[0, 1]$,

$$f[(1-\theta)\, x_0 + \theta x] \geqslant (1-\theta) f(x_0) + \theta f(x).$$

Ω is said to be strictly convex-cap when equality holds if and only if $x = x_0$ or $\theta = 0$.

Optimization Theorem

Let f be a continuous, weakly-differentiable strictly convex-cap map from a compact, convex, topological space Ω to $\mathscr{R}$. Define:

$$C \triangleq \sup_{x \text{ in } \Omega} f(x).$$

Then,

(1) $C = \max f(x) = f(x_0)$ for some unique x_0 in Ω.

(2) A necessary and sufficient condition for $f(x_0) = C$ is $f'_{x_0}(x) \leqslant 0$ for all x in Ω.

This basic Optimization Theorem (see Smith, 1969 or Luenberger, 1969 for proof) is valuable in determining the unconstrained optimal element within the convex space. It will also be necessary to determine an element which maximizes the function, subject to an additional constraint. For this purpose the Lagrangian Theorem is quoted below.

Lagrangian Theorem

Let Ω be a convex metric space, and f and g convex-cap functionals on Ω to $\mathscr{R}$, assume there exists an x_1 in Ω such that $g(x_1) < 0$, and let

$$C' \triangleq \sup_{\substack{x \text{ in } \Omega \\ g(x) \leqslant 0}} f(x).$$

If C' is finite, then (Luenberger, 1969) there exists a constant $\lambda \geqslant 0$ such that

$$C' = \sup_{x \text{ in } \Omega} [f(x) - \lambda g(x)].$$

Furthermore, if the supremum in the first equation is achieved by x_0 in Ω and $g(x_0) \leqslant 0$, it is achieved by x_0 in the second equation, and $\lambda g(x_0) = 0$.

The average mutual information between input and output random variables has been formulated as a map from the space $\mathscr{F}_A$ of probability distribution functions F having all points of increase on some finite interval $[-A, A]$. It will be established that $\mathscr{F}_A$ is convex and compact (in the "Levy" metric), and that I is a convex cap, continuous, and weakly differentiable functional in $\mathscr{F}_A$. The amplitude-constrained capacity is the supremum of $I(F)$ over all F in $\mathscr{F}_A$. Thus, the optimization theorem will guarantee the existence of a unique maximizing input distribution and provide necessary and sufficient conditions for achieving this global maximum. Later, an additional variance constraint will be imposed on the input. This will require use of the Lagrangian theorem.

III. Amplitude-Constrained Capacity of a Scalar Additive Gaussian Channel

The amplitude-constrained information capacity is to be determined for the scalar additive Gaussian channel discussed in the preceding section. Proposition 1 establishes that an "optimal input" exists and yields necessary and sufficient conditions for this input. Corollary 1 provides a more usable set of necessary and sufficient conditions. This result and Proposition 2 establish that the optimal input is discrete, taking on a finite number of values (although the number of values will be unknown).

These results permit development of a programming procedure capable of generating the capacity and the optimal input at a large number of values of the constraint A. As a notational convenience, unless necessary for clarification, $\mathscr{F}_A$ and $C(A)$ will be denoted simply as $\mathscr{F}$ and C, respectively.

PROPOSITION 1. *C is achieved by a unique probability distribution function F_0 in $\mathscr{F}$; i.e.,*

$$C = \max_{F \text{ in } \mathscr{F}} I(F) = I(F_0)$$

for some unique F_0 in $\mathscr{F}$. Furthermore, a necessary and sufficient condition for F_0 to achieve capacity is for all F in $\mathscr{F}$

$$\int_{-A}^{A} i(x; F_0)\, dF(x) \leqslant I(F_0).$$

Remark. It suffices to show that $\mathscr{F}$ is convex and compact in some topology, and that $I: \mathscr{F} \to \mathscr{R}$ is strictly convex-cap, continuous, and weakly differentiable in $\mathscr{F}$. Then, the optimization theorem presented in Section II establishes the existence of a unique F_0. The second statement also follows from that theorem by establishing that for all F in $\mathscr{F}$

$$I'_{F_0}(F) = \int_{-A}^{A} i(x; F_0)\, dF(x) - I(F_0).$$

Proof. The convexity of $\mathscr{F}$ is obvious. The compactness of $\mathscr{F}$ in the Lèvy metric topology (see Loève, 1955 or Moran, 1968 for definitions) follows from Helley's Weak Compactness Theorem and from the fact that convergence in the Lèvy metric is equivalent to complete convergence which on a finite interval is equivalent to weak convergence.

The convex-cap property follows from the fact that for any F_1 and F_2 in $\mathscr{F}$, and any θ in [0, 1],

$$p_Y[y; (1-\theta)F_1 + \theta F_2)] = (1-\theta)\, p_Y(y; F_1) + \theta p_Y(y; F_2),$$

and (by Lemma 8.3.1 of Ash, 1965)

$$-\int_{-\infty}^{\infty} p_Y(y; F_k) \log p_Y(y; F_\theta)\, dy \qquad \text{is finite for} \quad k = 1, 2,$$

because, then,

$$(1-\theta)\, H(F_1) + \theta H(F_2) \leqslant H[(1-\theta) F_1 + \theta F_2]$$

with equality if and only if $p_Y(y; F_k) = p_Y(y; F_\theta)$. Since the characteristic function of the Gaussian noise-random variable is positive pointwise, then, for arbitrary F_1 and F_2 in $\mathscr{F}$, pointwise equality of $p_Y(y; F_1)$ and $p_Y(y; F_2)$ occurs if and only if the Lèvy metric between F_1 and F_2, denoted $d(F_1, F_2)$, is zero. Thus, the strict convex-cap property holds.

The continuity of $H\colon \mathscr{F} \to \mathscr{R}$ (and hence $I\colon \mathscr{F} \to \mathscr{R}$) follows essentially (Smith, 1969) from the Helly–Bray theorem (Loève, 1955 or Moran, 1968) (which yields that $d(F_n, F) \underset{n}{\to} 0$ implies $p_Y(y; F_n) \underset{n}{\to} p_Y(y; F)$ for arbitrary F_n, F in $\mathscr{F}$), and from the boundedness and continuity of p_Y (which follows from the boundedness and continuity of p_N) and of $-p_Y \log p_Y$.

Finally, it can be established (Smith, 1969) that for arbitrary F_1 and F_2 in $\mathscr{F}$ and θ in [0, 1]

$$\lim_{\theta \downarrow 0} \left\{ \frac{I[(1-\theta) F_1 + \theta F_2] - I(F_1)}{\theta} \right\} = \int_{-A}^{A} i(x; F_1)\, dF_2(x) - I(F_1).$$

Thus, $I\colon \mathscr{F} \to \mathscr{R}$ is weakly differentiable and for all F_1 and F_2 in $\mathscr{F}$

$$I'_{F_1}(F_2) = \int_{-A}^{A} i(x; F_1)\, dF_2(x) - I(F_1). \qquad \text{Q.E.D.}$$

COROLLARY 1. *Let F_0 be an arbitrary probability distribution function in $\mathscr{F}$. Let E_0 denote the points of increase of F_0 on $[-A, A]$. Then, F_0 is "optimal" if and only if*

$$\begin{aligned} i(x; F_0) &\leqslant I(F_0) \qquad \text{for all } x \text{ in } [-A, A], \\ i(x; F_0) &= I(F_0) \qquad \text{for all } x \text{ in } E_0. \end{aligned}$$

Remark. Clearly, if both conditions hold, F_0 must be optimal because the necessary and sufficient condition of Proposition 1 is satisfied. It remains to prove the converse.

Proof. Assume that F_0 is optimal but the first equation of Corollary 1 is not true. Then, there exists x_1 in $[-A, A]$ such that $i(x_1; F_0) > I(F_0)$. Let $F_1(x) \triangleq \mathscr{U}(x - x_1)$ (a unit step function at x_1). Then,

$$\int_{-A}^{A} i(x; F_0)\, dF_1(x) = i(x_1; F_0) > I(F_0).$$

This contradicts Proposition 1. Thus, the first equation is valid. Now, assume that F_0 is optimal, but the second equation is not true. Then, because of the first statement, $i(x; F_0) < I(F_0)$ for all x in E' where E' is some subset of E_0 with positive measure; i.e., $\int_{E'} dF_0(x) = \delta > 0$. Since $\int_{E_0-E'} dF_0(x) = 1 - \delta$ and $i(x; F_0) = I(F_0)$ on $E_0 - E'$, clearly $I(F_0) < I(F_0)$ which is a contradiction. Thus, the second equation is valid. Q.E.D.

PROPOSITION 2. *E_0 is a finite set of points.*

Remark. This proposition says that the optimum random variable is discrete, that the optimal probability distribution function F_0 is simple, and that the capacity C is a function of a finite number of variables. The proof rests on the results of Corollary 1, and two classical theorems: the Identity Theorem of Complex Functions and the Bolzano Weierstrass Theorem.

Proof. In part, Corollary 1 implies that $h(x; F_0) = I(F_0) + D$ on E_0. The extension of $h(x; F_0)$ to the entire complex plane is well-defined:

$$h(z; F_0) \triangleq - \int_{-\infty}^{\infty} p_N(y - z) \log p_Y(y; F_0) \, dy$$

and analytic (Smith, 1969). If E_0 is infinite, then since $E_0 \subset [-A, A]$, E_0 has a limit point by the Bolzano–Weierstrass Theorem (Bartle, 1964) and, hence, $h(z; F_0) = I(F_0) + D$ on the entire complex plane by the Identity Theorem (Knopp, 1945). Thus, in particular, $h(x; F_0) = I(F_0) + D$ on the real line $\mathscr{R}$. It can be shown (Smith, 1969) that this is possible if and only if

$$p_Y(y; F_0) = e^{-I(F_0)-D} \qquad \text{for all } y \text{ in } \mathscr{R}.$$

This follows because $-\log p_Y(y; F_0)$ and $h(x; F_0)$, being locally integrable (Schwartz, 1966), have Fourier transforms at least in the sense of distributions, because the characteristic function of the noise is pointwise positive, and because $I(x; F_0) = I(F_0) + D$ can be written

$$\int_{-\infty}^{\infty} p_N(y - x)[-\log p_Y(y; F_0) - h(x; F_0)] \, dy = 0.$$

Thus, assuming E_0 is not finite leads to the conclusion that the output density function $p_Y(y; F_0)$ is uniform on the real line, which is an obvious impossibility. Q.E.D.

A Finite Dimensional Optimization Problem

It has been established that for each fixed amplitude limit A, an optimal input random variable X_0 or, equivalently, an optimal input distribution function F_0 exists which satisfies certain necessary and sufficient conditions. Furthermore, it has been established that X_0 a.s. takes on only a finite number of values. This finite set, denoted E_0, represents, equivalently, the collection of points of increase of F_0 (traditionally called the mass point positions). The optimum values of these points of increase are unknown. The optimum amount of increase (traditionally called the mass point values) of F_0 at each mass point is also unknown. In addition, the number of these points of increase is unknown, but this problem will be ignored momentarily.

Initially, the problem was the determination of an optimal distribution function and the average mutual information generated when that input distribution was used. The problem has essentially been reduced to the determination of a finite number of values. Thus, the capacity, for a fixed amplitude limit, can be formulated as the maximum of a function of a finite-dimensional vector, the components of which are the mass point positions and the mass point values.

Suppose the correct number $n = n_0(A)$ of mass points is known for a particular value of A. Let $x_1, x_2, \ldots, x_n$ denote the mass point positions of an arbitrary input distribution F, and let $Q_1, Q_2, \ldots, Q_n$ denote the corresponding mass point values. Then, F can be written as

$$F(x) = \sum_{i=1}^{n} Q_i \mathscr{U}(x - x_i).$$

Let the vector $\mathbf{Z} = (Z_1, \ldots, Z_{2n})$ consist of components $Z_i = Q_i$ for all $i = 1, 2, \ldots, n$, and $Z_{n+i} = x_i$ for all $i = 1, 2, \ldots, n$. The output density function p_Y will depend upon the vector-value of $\mathbf{Z}$

$$p_Y(y; \mathbf{Z}) = \sum_{i=1}^{n} Q_i p_N(y - x_i).$$

The information map I can be treated as a function of the vector $\mathbf{Z}$:

$$I(\mathbf{Z}) = -\int_{-\infty}^{\infty} p_Y(y; \mathbf{Z}) \log p_Y(y; \mathbf{Z})\, dy - D.$$

Let $\mathscr{R}_0$ denote the region of the $2n$-dimensional Euclidean space in which $\mathbf{Z}$, as defined, must lie. The determination of $\mathscr{R}_0$ is straightforward. If x_i is a mass point position then $|x_i| \leqslant A$ for all $i = 1, 2,..., n$. If Q_i is a mass point value, then $Q_i \geqslant 0$ for all $i = 1, 2,..., n$, and

$$\sum_{i=1}^{n} Q_i = 1.$$

Thus, $\mathscr{R}_0$ is simply the intersection of all the $3n + 1$ restriction sets within which these constraints are satisfied. Then, the capacity C is

$$C = \max_{\mathbf{Z} \text{ in } \mathscr{R}_0} I(\mathbf{Z}).$$

Many optimization algorithms have been implemented as computer programs which solve problems of this form: maximize a known function $I(\mathbf{Z})$ over all vectors $\mathbf{Z} = (Z_1, ..., Z_{2n})$ which lie in a well-defined restriction region $\mathscr{R}_0$. Thus, the determination of the capacity at a fixed value of A is now, in principle, straightforward. Unfortunately, one aspect has been neglected; the correct number of mass points and, hence, the dimensionality of the vector at any fixed amplitude limit is not known. Thus, in practice, the determination of capacity as a function of amplitude limit requires a programming procedure such as is described next.

A Programming Procedure

It can be shown that $i(x; F_0)$ is concave-cup for A sufficiently small ($A \leqslant .1$). This leads to the conclusion, from the necessary and sufficient conditions of Corollary 1, that the optimal set of mass points for A sufficiently small is an equal pair of mass points at the interval extremes $\pm A$. This is intuitively appealing—in a very noisy environment, the mass points are separated as much as is permissible. This result provides a starting point for the program.

The necessary and sufficient conditions of Corollary 1 provide a test to determine whether the actual number n equals the correct number $n_0(A)$. To determine $C(A)$ for $A > .1$, increment A by some small amount δ. Using the same two mass points, with the increased value of A, test to see if the necessary and sufficient conditions are satisfied. If the test is valid, this F_0 is still optimum (Corollary 1), and $C(A) = I(F_0)$. Thus, A can be incremented by δ again and the test repeated. Failure of this test at some value of A indicates that this F_0 is no longer optimum and the present number of mass points is no longer sufficient. Thus, the number of mass points must be increased by one, and the distribution function F_0 which maximizes the information rate (subject to the mass point number restriction) is determined.

The test is then repeated. If the necessary and sufficient conditions are satisfied, the number of mass points is correct. If not, the number must be increased by one, and the procedure repeated. The programming procedure is outlined in Fig. 1.

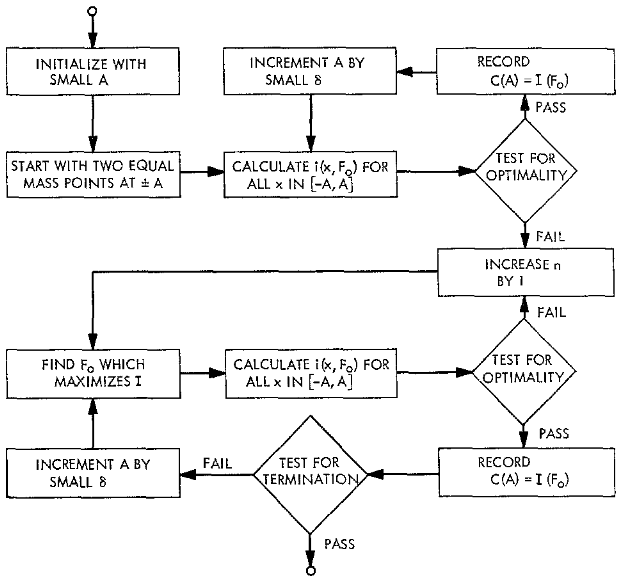

FIG. 1. Outline of programming procedure.

Results

The capacity $C(A)$ of a scalar zero mean unit variance Gaussian channel with input amplitude constraint A is plotted in Fig. 2 in nats/symbol. The correct number $n_0(A)$ of mass points at each A was determined using the programming procedure described. The optimization algorithm used is described in detail by Fiacco and McCormick (1968). (The computer program implementation of this algorithm is available in the University of California Computer Library at Berkeley, California. It is titled, H2 CAL SUMT.) Optimum mass points at selected values of A are shown in Fig. 3.

The asymptote of $C(A)$, as A increases, is the same as the information rate due to a uniformly distributed input. This verifies the argument by Shannon (1948) that for large A,

$$H(X) \simeq H(Y), \qquad \text{and} \qquad C(A) \simeq \log 2A - \log \sqrt{2\pi e}.$$

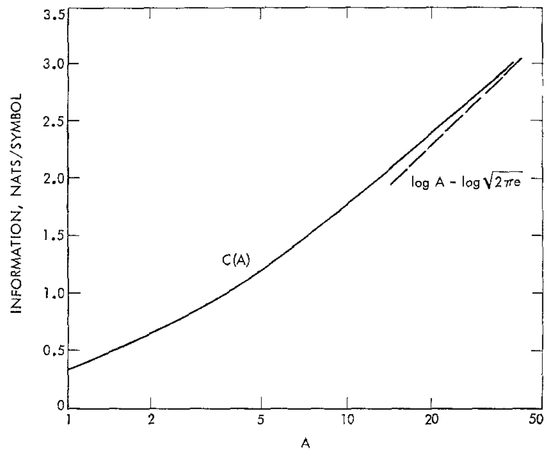

FIG. 2. Amplitude-constrained capacity of a scalar Gaussian channel.

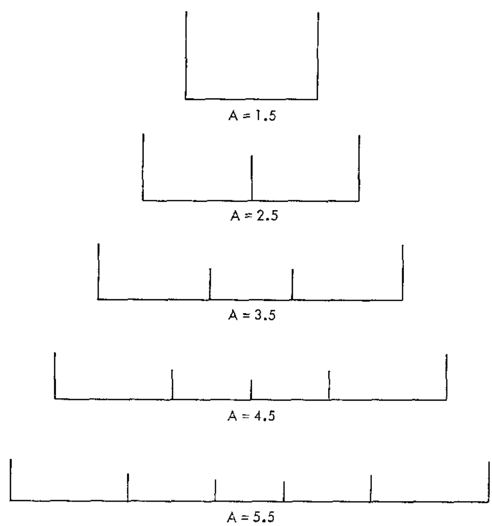

FIG. 3. Optimum mass points at selected values of amplitude constraint A.

This follows from the fact that the entropy of an absolutely continuous random variable a.s. limited to a finite interval is maximized by a uniform distribution over the interval.

Finally, it may be added that these results do not rely strictly on the "Gaussianness" of the noise, but only on its general "smoothness." The capacity of non-Gaussian channels is discussed by Smith (1969).

IV. Amplitude- and Variance-Constrained Capacity of a Scalar-Additive Gaussian Channel

For any amplitude limit A and variance limit σ^2, the amplitude- and variance-constrained capacity, denoted $C(A, \sigma^2)$, is[2]

$$C(A, \sigma^2) \triangleq \sup_{\substack{F \text{ in } \mathscr{F}_A \\ \sigma_F^2 \leqslant \sigma^2}} I(F),$$

where

$$\sigma_F^2 \triangleq \int_{-A}^{A} x^2 \, dF(x).$$

Alternately, defining $J: \mathscr{F}_A \to \mathscr{R}$ by $J(F) \triangleq \sigma_F^2 - \sigma^2$ yields by the Lagrangian theorem quoted in Section II, a second expression for capacity; i.e., there exists a nonnegative constant $\lambda = \lambda(A, \sigma^2)$ for $J(F) \leqslant 0$ such that

$$C(A, \sigma^2) = \sup_{F \text{ in } \mathscr{F}_A} [I(F) - \lambda J(F)].$$

Again, where no confusion can arise, $C(A, \sigma^2)$ and $\mathscr{F}_A$ are simply denoted by C and $\mathscr{F}$, respectively.

Proposition 3. *The value C is achieved by a unique input distribution function F_0 in $\mathscr{F}$ satisfying the variance constraint; i.e.,*

$$C(A, \sigma^2) = \max_{F \text{ in } \mathscr{F}} [I(F) - \lambda J(F)] = I(F_0) - \lambda J(F_0).$$

A necessary and sufficient condition for $C = I(F_0)$ is that for some constant $\lambda \geqslant 0$,

$$\int_{-A}^{A} [i(x; F_0) - \lambda x^2] \, dF(x) \leqslant I(F_0) - \lambda\sigma^2 \qquad \text{for all } F \text{ in } \mathscr{F}.$$

Proof. $J: \mathscr{F} \to \mathscr{R}$ is clearly linear and bounded and, hence, convex-cap, continuous, and weakly differentiable in $\mathscr{F}$ with $J'_{F_1}(F_2) = J(F_2) - J(F_1)$. In addition, for any x, such that $|x_1| < \sigma$, letting $F_1(x) \triangleq \mathscr{U}(x - x_1)$ implies that $J(F_1) < 0$. Then by the Lagrangian theorem, since $C(A, \sigma^2)$ is finite, there exists some constant λ such that the second expression for capacity

[2] The second moment constraint is the same as the variance constraint since the "optimal" input must be zero mean because of the symmetry of p_N.

holds. Further, $I - \lambda J$ is strictly convex-cap, continuous, and weakly differentiable. Thus, by the Optimization Theorem there exists a unique distribution function F_0 in $\mathscr{F}$ such that $C(A, \sigma^2) = I(F_0) - \lambda J(F_0)$. The necessary and sufficient condition becomes $I'_{F_0}(F) - \lambda J'_{F_0}(F) \leqslant 0$ for all F in $\mathscr{F}$, or, for all F in $\mathscr{F}$

$$\int_{-A}^{A} [i(x; F_0) - \lambda x^2] \, dF(x) \leqslant I(F_0) - \lambda \sigma_{F_0}^2 .$$

If $\sigma_{F_0}^2 < \sigma^2$, the variance constraint is trivial and the constant λ is zero i.e., $J(F_0) < 0$, but $\lambda J(F_0) = 0$). Thus, the necessary and sufficient condition is established. Q.E.D.

COROLLARY 2. *Let F_0 be an arbitrary probability distribution function in $\mathscr{F}$ satisfying the variance constraint. Let E_0 denote the points of increase of F_0 on $[-A, A]$. Then F_0 is optimal if and only if, for some $\lambda \geqslant 0$,*

$$i(x; F_0) \leqslant I(F_0) + \lambda(x^2 - \sigma^2) \qquad \text{for all } x \text{ in } [-A, A],$$

$$i(x; F_0) = I(F_0) + \lambda(x^2 - \sigma^2) \qquad \text{for all } x \text{ in } E_0 .$$

The proof parallels that of Corollary 1, and, thus, is not included.

PROPOSITION 4. *The value E_0 is a finite set of points.*

Proof. The proof closely parallels the proof of Proposition 2 Assuming, as before, that E_0 is not finite leads to the conclusion that the output density function $p_Y(Y; F_0)$ must be Gaussian with variance $1 + \sigma^2$. It is not possible to achieve this output with an amplitude constrained input, and, hence, E_0 must be finite.

Programming Procedure

The capacity to be determined can now be formulated as the maximum of a function of a finite-dimensional vector. The components and restrictions are as before, except for an added restriction to include the variance constraint.

The necessary and sufficient conditions of Corollary 2 provide a test, comparable to the test discussed in Section III, to determine whether the actual number n of mass points equals, at any pair (A, σ^2), the correct number $n_0(A, \sigma^2)$. If $n = n_0(A, \sigma^2)$, then the optimization algorithm will produce a distribution function F_0 which satisfies the necessary and sufficient condition for the constant λ determined by solving the second of the two equations of the necessary and sufficient condition at any x_0 in E_0, except

$x_0 = \sigma$. Further, $I(F_0)$ will equal $C(A, \sigma^2)$. If $n < n_0(A, \sigma^2)$, the optimization algorithm will produce a distribution function F_0 which will not satisfy the above test, for λ calculated as above, and $I(F_0)$ will be less than $C(A, \sigma^2)$.

The starting point can be established by showing that for A sufficiently small and A^2/σ^2 sufficiently small the optimum set of mass points are a pair of equal mass points at $\pm A$ and one at the origin. This third mass point has enough mass to satisfy the variance constraint.

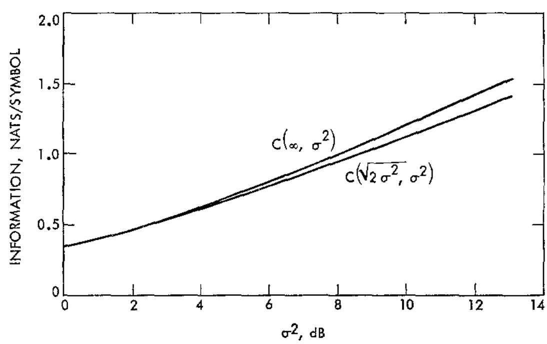

FIG. 4. Amplitude- and variance-constrained capacity of a scalar Gaussian channel.

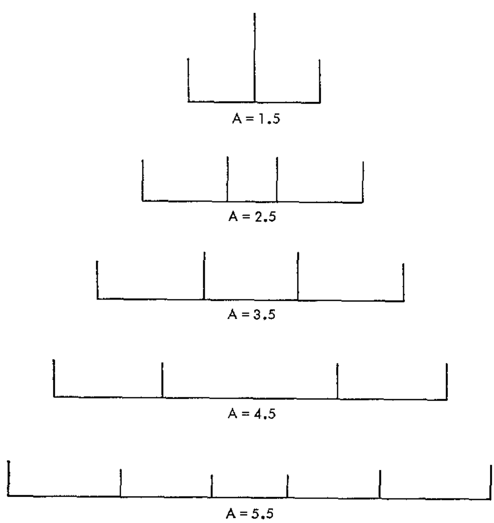

FIG. 5. Optimum mass points at selected values of amplitude constraint A with $\sigma = A/\sqrt{2}$.

Capacity $C(A, \sigma^2)$ is plotted in Fig. 4 in nats/symbol as a function of σ^2 in dB for a fixed ratio of $A^2/\sigma^2 = 2$. Optimum mass points at selected values of A are shown in Fig. 5. A plot of the curve $\log \sqrt{1 + \sigma^2}$ is also included in Fig. 4. This represents the limiting curve as $A^2/\sigma^2 \to \infty$. The interesting result is that the curves are so close. For example, restricting the peak power A^2 to no more than twice the average power σ^2, for $\sigma^2 \leqslant 10$ dB results in less than a 7% loss in capacity. Furthermore, only five mass points are required at $\sigma^2 = 10$ dB to achieve $C(\sqrt{2\sigma^2}, \sigma^2)$. Note that any greater ratios of A^2/σ^2 will yield curves lying between the two curves given.

RECEIVED: April 6, 1970; REVISED: November 5, 1970

REFERENCES

ASH, R. B. (1965), "Information Theory," Interscience, New York.

BARTLE, R. G. (1964), "Elements of Real Analysis," John Wiley & Sons, Inc., New York.

FIACCO, A. V. AND MCCORMICK, G. P. (1968), "Nonlinear Programming," John Wiley & Sons, Inc., New York.

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

GOLDMAN, S. (1953), "Information Theory," Prentice-Hall, Englewood Cliffs, N. J.

KNOPP, K. (1945), "Theory of Functions," Dover, New York.

LOÈVE, M. (1955), "Probability Theory," Van Nostrand, New York.

LUENBERGER, D. G. (1969), "Optimization by Vector Space Methods," John Wiley & Sons, New York.

MORAN, P. A. (1968), "Introduction to Probability Theory," Clarendon, Oxford.

PINSKER, M. S. (1964), "Information and Information Stability of Random Variables and Processes," Holden-Day, San Francisco, Calif.

SCHWARTZ, L. (1966), "Mathematics for the Physical Sciences," Hermann, Paris.

SHANNON, C. E. (1948), A mathematical theory of communication, *Bell System Tech. J.* **27**, Part II, 623–656.

SMITH, J. G. (1969), "On the Information Capacity of Peak and Average Power Constrained Gaussian Channels," Ph.D. dissertation, Dept. Elec. Engrg., Univ. California, Berkeley, Calif.