Capacity-Achieving Input Distributions for Some Amplitude-Limited Channels with Additive Noise

W. Oettli

[45]

Abstract: An additive noise channel wherein the noise is described by a piecewise constant probability density is shown to reduce to a discrete channel by means of an explicit construction. In addition, conditions are found which describe a class of continuous, amplitude-limited channels for which the capacity-achieving input distribution is binary.

March 1973

0. It is the purpose of this note to discuss some examples of continuous, amplitude-limited channels with additive noise having the property that the capacity-achieving input distribution can be ascertained a priori to have discrete support. In sections 1 and 2 we first review some well-known optimality conditions for a capacity-achieving input distribution to be used in the remainder of our discussion. In section 3 we consider the case of a piecewise constant noise density. We show by means of an explicit construction that in this case the given channel is equivalent to a certain *discrete* channel. This gives the possibility of characterizing a priori the discrete support of the optimal input distribution. The results of this section were motivated by, and include, results of Färber and Appel concerning the special case of a rectangular noise distribution. In sections 4 and 5 we describe a class of noise distributions which guarantee that the channel assumes capacity via a *binary* input distribution. The approach used in these two sections should be new.

1. We first recall some elementary facts about a discrete, memoryless channel with input symbols x_j, output symbols y_i, and transition probabilities $\pi_{ij} = \text{prob}\{y_i | x_j\}$. For a given input distribution p_j with corresponding output distribution $q_i = \sum_j \pi_{ij} p_j$ the transinformation, $T(p)$, is given as

$$T(p) = \sum_j p_j \sigma_{p,j}, \quad \sigma_{p,j} = \sum_i \pi_{ij} \log \frac{\pi_{ij}}{q_i}.$$

A capacity-achieving input distribution, $\hat{p}$, is such that $T(\hat{p}) = \max_{p \in \mathcal{P}} T(p)$, where $\mathcal{P}$ is the set of all possible input distributions. A necessary and sufficient condition for a capacity-achieving input distribution $\hat{p}$ is that

(1) $\qquad \sigma_{\hat{p},j} = \max_k \sigma_{\hat{p},k}$ for all j with $\hat{p}_j > 0$.

In what follows we only shall need that (1) is a *necessary* condition for a capacity-achieving input distribution $\hat{p}$ of a *discrete* channel.

Condition (1) is easily seen to be equivalent to the optimality condition derived in [1]. Iterative methods for finding a capacity-achieving input distribution may be found in [1], [5].

2. We now consider more generally a (time-discrete) channel of density type [2]. Such a channel is specified by a set X of inputs, provided with a σ-algebra Ξ, a set Y of outputs, provided with a σ-algebra H, and transition densities $\pi(y|x)$ with regard to a given σ-finite measure $\mu(\cdot)$ on H. $\mathcal{P}$ is the family of all probability distributions over Ξ, and for any $P \in \mathcal{P}$ the output density is given by $q(y) = \int_X P(dx)\pi(y|x)$. Furthermore

$$T(P) = \int_X P(dx)\sigma_P(x), \quad \sigma_P(x) = \int_Y \pi(y|x) \log\frac{\pi(y|x)}{q(y)}\mu(dy),$$

and the *capacity* is defined as $C = \sup_{P\in\mathcal{P}} T(P)$. We are interested in cases where capacity is achieved by a distribution $\hat{P} \in \mathcal{P}$ whose support contains only a finite number of points (assuming, thereby, that Ξ comprises at least all one-element subsets of X). For such a distribution $\hat{P}$ with finite support a condition analogous to (1) is given by

$$(2) \qquad \sigma_{\hat{P}}(x) = \max_{\xi\in X} \sigma_{\hat{P}}(\xi) \text{ for all } x \text{ with } \hat{P}(x) > 0,$$

i.e., $\sigma_{\hat{P}}$ assumes its maximum $\hat{P}$-almost everywhere. We only shall need that for a continuous channel (2) is a *sufficient* condition for a capacity-achieving input distribution $\hat{P}$ with *finite* support. Sufficiency is obvious, as the elementary inequality $q\cdot\log(\hat{q}/q) \le \hat{q} - q$ implies

$$\int_X P(dx)[\sigma_P(x) - \sigma_{\hat{P}}(x)] = \int_Y q(y)\log\frac{\hat{q}(y)}{q(y)}\mu(dy) \le 0,$$

and thereby $T(P) \le \int_X P(dx)\sigma_{\hat{P}}(x) \le \max_{\xi\in X}\sigma_{\hat{P}}(\xi) = T(\hat{P})$ for any other input distribution P.

We henceforth shall restrict ourselves to the amplitude-limited channel with additive noise. In this case $X \subseteq \mathbb{R}$ is a bounded interval; $y=x+z$, $\pi(y|x)=\omega(y-x)$, where z is a given noise random variable having range $Z \subseteq \mathbb{R}$ and density $\omega(z)$ with regard to Lebesgue measure. We obtain then

(3) $\sigma_P(x)=\alpha-\int_{X+Z} \omega(y-x) \log q(y)dy, \quad \alpha=\int_Z \omega(z) \log \omega(z)dz.$

3. We first show that an amplitude-limited channel with additive noise admits a capacity-achieving input distribution with finite support, provided Z is a bounded interval and $\omega(z)$ is piecewise constant between equally spaced points of Z.

To be more specific, let $Z=[0,D]$, $D=n\cdot\Delta$ (n integer), $\omega(z)=$ $=$const. within each interval $[k\cdot\Delta,(k+1)\cdot\Delta]$ of Z. Let $X=[0,S]$, $S=m\cdot\Delta+r$ (m integer, $0 \le r < \Delta$). Let $Y=X+Z=[0,S+D]$. This describes our continuous channel. It is possible to reduce this continuous channel to a discrete channel whose input symbols are points $\xi_j \in X$ and whose output symbols are intervals $\eta_i \subseteq Y$. To this end we define in $[0,S+D]$ points

$$\xi_k = \begin{cases} \Delta\cdot k/2 & (k \text{ even}), \\ r+\Delta\cdot(k-1)/2 & (k \text{ odd}), \end{cases} \quad k=0,1,\ldots,2(m+n)+1,$$

and intervals

$$\eta_k = [\xi_k,\xi_{k+1}], \quad k=0,1,\ldots,2(m+n).$$

Now the capacity of the above continuous channel is achieved by a discrete input distribution whose support is a subset of $\{\xi_j\}_0^{2m+1} \subseteq X$. This can be seen as follows: If we place probabilities $P(\xi_j)$ at the points ξ_j in such a way that $q(y) > 0$, then $\sigma_P(x)$ is continuous, and is linear between any two successive points ξ_j, ξ_{j+1}. In order to establish the optimality of such a distribution it is therefore sufficient to test condition (2) with X replaced by $\{\xi_j\}$. At the points ξ_j, however, the function $\sigma_P(\xi_j)$ equals the components of the vector $\sigma_{P,j}$ corresponding to the discrete channel with inputs $\{\xi_j\}_0^{2m+1}$, outputs $\{\eta_i\}_0^{2(m+n)}$, and transition probabilities $\pi_{ij} = \int_{\eta_i} \omega(y-\xi_j)dy$ under the identification $p_j = P(\xi_j)$. If we have found a capacity-achieving input distribution $\hat{p}$ of this discrete channel, then this distribution satisfies (1) as a necessary condition. The corresponding input distribution $\hat{P}$ of the continuous channel, consequently, satisfies (2) on the restricted set $\{\xi_j\}$; however, because of the piecewise linearity of $\sigma_{\hat{P}}(x)$, $\hat{P}$ satisfies (2) on all of X, thus giving capacity for the continuous channel.

In the special case where $\omega(z)$ is a rectangular density in the interval $[0,D]$ (i.e., $\Delta=D$, $n=1$), the optimal distribution can be given explicitly: Set

$$\hat{P}(\xi_j) = \begin{cases} \rho\cdot(j+1)/2 & (j \text{ odd}), \\ \rho\cdot(2m+2-j)/2 & (j \text{ even}), \end{cases} \qquad j=0,1,\ldots,2m+1,$$

where ρ is a normalizing factor. For this distribution, $\hat{q}(y)$ is on Y periodic with period Δ (on η_k the value of $\hat{q}(y)$ equals $\rho(m+1)/\Delta$ for k even, and $\rho(m+2)/\Delta$ for k odd; η_k and η_{k+1} together have length Δ for all k). Therefore the function $\sigma_{\hat{P}}(x) = \alpha - \frac{1}{\Delta}\int_x^{x+\Delta} \log \hat{q}(y)\, dy$ is independent of x, and the optimality criterion (2) is satisfied. This has also been noted by Färber [3] and Appel [4].

4. We shall consider from now on a channel with a signalling interval of length 2s, say $X = [-s, +s]$. An input distribution $\hat{P}$ is called binary, if $\hat{P}(-s) = p$, $\hat{P}(+s) = 1-p$. For binary $\hat{P}$ the optimality condition (2) becomes

(4) $\qquad \sigma_{\hat{P}}(-s) = \sigma_{\hat{P}}(+s) \equiv \kappa,$

(5) $\qquad \sigma_{\hat{P}}(\xi) \le \kappa$ if $-s < \xi < +s$.

In what follows we shall describe a class of amplitude-limited channels with additive noise for which the optimal distribution can be ascertained a priori to be binary. We note that equality (4) can always be attained by an appropriate choice of the weight p. We shall formulate conditions on s and ω which also ensure that the inequality (5) is satisfied, by enforcing suitable functional properties of $\sigma(x)$, such as convexity downwards[†]. We first turn to the case of symmetric noise density.

Proposition 1: Let $\omega(z)$ be symmetric about z=M, increasing[††] in $(-\infty , M]$, decreasing in $[M , +\infty)$, concave in an interval $[M-\phi , M+\phi]$, and eventually constant in $[M-\ell , M+\ell]$ $(0 \leq \ell \leq \phi)$. If $2s \leq \ell+\phi$, then there exists a capacity-achieving input distribution which is binary.

Proof: Set M=0 for convenience. We choose $\hat{P}$ binary with $p=\frac{1}{2}$. The corresponding output density $\hat{q}(y) = \frac{1}{2}\omega(y+s) + \frac{1}{2}\omega(y-s)$ is symmetric about y=0. $\hat{q}$ is increasing in $(-\infty , \ell-s]$, decreasing in $[-\ell+s , +\infty)$, and concave in $[s-\phi , -s+\phi]$ (since it is the sum of two functions with these same properties). The condition $2s \leq \ell+\phi$ implies that these three regions together cover Y. Moreover, the concavity of $\hat{q}$ in $[s-\phi , -s+\phi]$ together with the symmetry about y=0 guarantees that within this interval $\hat{q}$ is increasing up to y=0, and decreasing thereafter. Hence $\hat{q}$ is increasing in $(-\infty , 0]$ and decreasing in $[0 , +\infty)$. The function $\gamma(y) = -\log \hat{q}(y)$ is then decreasing in $(-\infty , 0]$, increasing in $[0 , +\infty)$, and symmetric about y=0. The function $\sigma_{\hat{P}}(x) = \alpha + \int\omega(y-x)\gamma(y)dy$ is therefore symmetric around x=0. We show that it is decreasing in $[-s , 0]$ (condition (5) is then obviously satisfied). Indeed: Let $-s \leq \xi_1 \leq \xi_2 \leq 0$.

$$\sigma(\xi_1) - \sigma(\xi_2) = \int\omega(z)\left[\gamma(z+\xi_1) - \gamma(z+\xi_2)\right]dz$$
$$\equiv \int\omega(z)\left[\rho(z)\right]dz.$$

$\rho(z)$, in view of the symmetry of γ around 0, is skew-symmetric around $K = -(\xi_1+\xi_2)/2$. Also $\rho(z) \leq 0$ for $z \geq K$. For $\zeta \geq 0$ we have

$\omega(K+\zeta) \leq \omega(K-\zeta)$, because of $K \geq 0$; moreover $\rho(K+\zeta) = -\rho(K-\zeta) \leq 0$. Therefore $\omega(K+\zeta)\cdot\rho(K+\zeta) \geq -\omega(K-\zeta)\cdot\rho(K-\zeta)$,

$$\int_{z \geq K} \omega(z)\rho(z)dz \geq -\int_{z \leq K} \omega(z)\rho(z)dz,$$

and consequently $\sigma(\xi_1) - \sigma(\xi_2) \geq 0$. q.e.d.

Examples: a) $\omega(z)$ rectangular of length 1 ($\ell=\phi=1/2$). Binary signalling is optimal if $2s \leq \ell+\phi = 1$. In this case the bound is sharp. b) $\omega(z)$ triangular of length 1 ($\ell=0$, $\phi=1/2$). The proposition gives $2s \leq 0.5$. However, direct verification of (5) shows that binary signalling is optimal up to $2s \approx 0.72$.

5. We now drop the assumption of symmetry. $S = 2s$ is the length of the signalling interval.

Proposition 2: Let $\omega(z)$ be increasing in $(-\infty, M]$, decreasing in $[M, +\infty)$, and concave in $[M-\phi, M+\phi]$. If S satisfies

(6) $\max\{\omega(M-\phi+S), \omega(M+\phi-S)\} \leq \min\{\omega(M-S), \omega(M+S)\}$ (†††),

then there exists a capacity-achieving input distribution which is binary.

Proof: Set M=0. We choose $\hat{P}$ binary such that equality (4) is satisfied. We then show that under the assumptions made $\sigma_{\hat{P}}(x)$ is convex, so that the inequality (5) is satisfied, too. $\hat{q}(y)$, being a positive linear combination of $\omega(y+s)$ and $\omega(y-s)$, is increasing in $(-\infty, -s]$, decreasing in $[s, +\infty)$, and concave in $[s-\phi, -s+\phi]$. $\gamma(y) = -\log\hat{q}(y)$ is then

decreasing in $(-\infty, -s]$,

convex in $[s-\phi, -s+\phi]$,

increasing in $[s, +\infty)$.

Choose a constant θ between the right- and left-hand sides of (6) and decompose $\omega(z) = \overline{\omega}(z) + \underline{\omega}(z)$, where

$$\overline{\omega}(z) = \max\{\omega(z)-\theta, 0\}, \quad \underline{\omega}(z) = \min\{\omega(z), \theta\}.$$

Because of (6) we have

$\overline{\omega}(z) = 0$ outside $[-\phi+S\ ,\ \phi-S] \equiv A$,

$\underline{\omega}(z) = \theta$ within $[-S\ ,\ +S]$.

The function $a(x) = \int_{x+A} \overline{\omega}(y-x)\gamma(y)dy$ is convex, since for all $x \in [-s\ ,\ +s]$ the domain of integration x+A is contained in the interval $[-\phi+s\ ,\ \phi-s]$ in which γ is convex, and since the convolution of a nonnegative function and a convex function is convex. The function $b(x) = \int \underline{\omega}(y-x)\gamma(y)dy$ is also convex. Direct verification of this is somewhat tedious; it is more convenient to use the right- and left-hand derivatives of b(x): Since $\underline{\omega}$=const. in $[-S\ ,\ +S]$, we have

$$b'(x\pm 0) = -\int_{y \le x-S} d\underline{\omega}(y-x)\gamma(y\pm 0) - \int_{y \ge x+S} d\underline{\omega}(y-x)\gamma(y\pm 0).$$

If x varies in $[-s\ ,\ +s]$, the domain of integration of the first integral is always contained in the domain $y \le -s$ in which $\gamma(y)$ is decreasing. Also $d\underline{\omega}$ in the first integral is always nonnegative because of $y-x \le 0$. Therefore the first term is increasing with x; likewise the second term. The monotonicity of the derivatives of b(x) proves the convexity of b(x). The function $\sigma_{\hat{p}}(x) = \alpha + a(x) + b(x)$ is then convex, by the convexity of a and b. q.e.d.

[1] B. Meister, W. Oettli: On the capacity of a discrete, constant channel. Information and Control 11 (1967), 341 - 351.

[2] K. Jacobs: Über Kanäle vom Dichtetypus. Math. Z. 78 (1962), 151 - 170.

[3] G. Färber: Die Kanalkapazität allgemeiner Übertragungskanäle bei begrenztem Signalwertbereich, beliebigen Signalübertragungszeiten sowie beliebiger Störung. A.E.Ü. (Arch. elektr. Übertragung) 21 (1967), 565 - 574.

[4] U. Appel: Der Einsatz der nichtlinearen Programmierung zur Bestimmung der Kanalkapazität von signalwertbegrenzten Analogkanälen. Diplomarbeit, Institut für Nachrichtentechnik, Technische Hochschule München, 1967.

[5] S. Arimoto: An algorithm for computing the capacity of arbitrary discrete memoryless channels. IEEE Trans. Information Theory IT-18 (1972), 14 - 20.

Footnotes.

(†) $\psi(t)$ is convex if $\psi[(1-\lambda)t_o+\lambda t_1] \leq (1-\lambda)\psi(t_o)+\lambda\psi(t_1)$, $0 \leq \lambda \leq 1$.
ψ is concave if $-\psi$ is convex.

(††) Here and in the following "increasing" does not necessarily mean "strictly increasing"; thus $\omega(z)$ may well be zero outside a given interval Z.

(†††) This condition delimits an interval $0 \leq S \leq S_o$. If ω is symmetric, as in the preceding section, then $S_o = \max\{\ell, \phi/2\}$, thus being smaller than the bound derived in the preceding section, which was $\ell+\phi$.