[10] K. Marton, "Error exponent for source coding with a fidelity criterion," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 197–199, Mar. 1974.

[11] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," in *IRE Nat. Conv. Rec.*, 1959, pp. 142–163.

[12] E. Tuncel and K. Rose, "Error exponents in scalable source coding," *IEEE Trans. Inform. Theory*, vol. 49, pp. 289–296, Jan. 2003.

On the Discreteness of Capacity-Achieving Distributions

Aslan Tchamkerten*, Student Member, IEEE*

***Abstract*—We consider a scalar additive channel $x \rightarrow x + N$ whose input is amplitude constrained. By extending Smith's argument, we derive a sufficient condition on noise probability density functions (pdfs) that guarantee finite support for the associated capacity-achieving distribution(s).**

***Index Terms*—Additive channel, capacity-achieving distribution.**

I. Introduction

Smith [14] investigated capacity-achieving distributions of additive channels whose noise probability density functions (pdfs) decay like a Gaussian tail. Under an input amplitude constraint, he showed that the capacity-achieving distribution was unique, with a finite set of mass points. Oettli [9] stated a criterion that related the noise pdf and input amplitude constraint. He showed that when this criterion is satisfied, along with a unimodality constraint, there exists a binary capacity-achieving distribution. Later Das [2] showed that an additive channel with "heavy tailed" noise admits a finite support capacity-achieving distribution under an average-power constraint on the input.

Several authors [1], [4], [7], [12] have derived specific results on quadrature Gaussian channels, Rayleigh-fading channels, noncoherent additive Gaussian noise channels, and noncoherent Rician-fading channels. Related works can also be found in [5], [8], [10].

We consider additive channels whose inputs are amplitude constrained. For such a setting, we derive a family of noise pdfs whose corresponding capacity-achieving distributions consist of a finite set of mass points. Unlike [9], the discrete character of the capacity-achieving distribution does not depend either on the amplitude constraint or on the unimodality of the noise pdf. Our result also extends the results in [2], [14] to noise pdfs with a wider range of tail decay.

We end this section with some notational conventions. In Section II, we present our result and illustrate it with examples. Section III is devoted to the proofs and, finally, in Section IV we list some directions for future research.

Manuscript received May 14, 2002; revised July 20, 2004. This work was supported in part by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS), a center supported by the Swiss National Science Foundation under Grant 5005-67322.

The author is with the Laboratoire de Théorie de l'Information (LTHI), ISC-I&C, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland (e-mail: aslan.tchamkerten@epfl.ch).

Communicated by P. Narayan, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2004.836662

Capital letters denote random variables (e.g., X) and small letters (e.g., x) denote specific values taken by them. The expectation of X is denoted by $\mathbb{E}X$. A cumulative density function (cdf) is denoted by F. For a sequence of cdfs $\{F_n\}_{n\geq 1}$, $F_n \xrightarrow{w} F$ denotes that F_n converges weakly to F. The entropy of a positive-valued function f is denoted by $H(f)$. For any $k \geq 0$ and any two positive-valued functions L and U, we define

$$H(U)|_k = -\int_{y\leq k} U(y)\log U(y)dy$$

and

$$D(U\|L)|_k = \int_{y\geq k} U(y)\log \frac{U(y)}{L(y)}dy.$$

The real and imaginary parts of an element z in the complex domain $\mathbb{C}$ are denoted by $\Re z$ and $\Im z$, respectively. The boundary of a domain $\mathcal{D} \subseteq \mathbb{C}$ is denoted by $\partial\mathcal{D}$ and its closure by $\overline{\mathcal{D}}$. The open ball centered on $z \in \mathbb{C}$ with radius ε (with respect to the Euclidian distance) is denoted by $B(z;\varepsilon)$.

II. Statement of the Result

We consider a scalar additive noise channel $X \longrightarrow X + N = Y$ with input X, output Y, and noise N independent of X. The cdf F of X is assumed to belong to the set

$$\Omega \triangleq \left\{F : \int_{-A}^{A} dF(x) = 1\right\}$$

where A is any strictly positive constant. Given a noise pdf p_N, the capacity of the corresponding channel is defined by

$$C \triangleq \sup_{F\in\Omega} I(F)$$

with

$$I(F) \triangleq \int_{-A}^{A}\int_{\mathbb{R}} p_N(y-x)\log\frac{p_N(y-x)}{p(y;F)}dydF(x)$$

and where

$$p(y;F) \triangleq \int_{-A}^{A} p_N(y-x)dF(x).$$

We define $\mathcal{N}$ as the set of noise pdfs p_N such that

i. $p_N(y) > 0$ for all $y \in \mathbb{R}$;

ii. there exists $\mu > 0$ such that $\mathbb{E}|N|^\mu < \infty$;

iii. there exists $\delta > 0$ such that p_N admits an analytic extension on $\mathcal{D}_\delta \triangleq \{z \in \mathbb{C} : |\Im z| < \delta\}$;[1]

iv. there exists $k \geq 0$ and two nonincreasing functions $L : [k,\infty) \rightarrow \mathbb{R}_+$ and $U : [k,\infty) \rightarrow \mathbb{R}_+$ such that

   a) for all $z \in \mathcal{D}_\delta$ with $|z| \geq k$,

$$0 < L(|\Re z|) \leq |p_N(z)| \leq U(|\Re z|) \leq 1$$

   b) $H(U)|_k < \infty$

   c) $\int_{y\geq k+x} \frac{U(y-x)^3}{L(y)^2}dy < \infty$ for all $x \in \mathbb{R}_+$.

Our result stands in the following theorem.

[1]Let $\mathcal{E}$ be a complex open connected set including the real line. A mapping $f : \mathbb{R} \rightarrow \mathbb{R}$ is said to have an analytic extension on $\mathcal{E}$ if there exists a mapping $g : \mathbb{C} \rightarrow \mathbb{C}$ which is analytic on $\mathcal{E}$ and so that $g(x) = f(x)$ for any $x \in \mathbb{R}$.

Theorem 1: Let $X \longrightarrow X+N$ be a scalar additive channel where the input X has cdf in Ω and where N has density $p_N \in \mathcal{N}$. Then

$$C \triangleq \sup_{F\in\Omega} I(F) = \max_{F\in\Omega} I(F) \quad (1)$$

and any capacity-achieving cdf has finite set of points of increase.[2]

Notice, contrary to the result in [9], that the conclusion of the theorem holds irrespectively of the amplitude constraint A.

Let us comment on the class $\mathcal{N}$. The first two conditions are mild requirements. The key and the most stringent requirement is the property of analyticity. The reason for $\mathcal{D}_\delta$ to be a band will be clarified later. Property iv. is primarily to ensure tight upper and lower bounds (from a certain range k) on the noise pdf.

Since Das' result [2] applies to channels with finite average power constraint, we may compare the class of "heavy-tailed" noises with the class $\mathcal{N}$ introduced above. Heavy-tailed noises are defined by a set of three requirements, A1, H1, and H2 in addition to an analytic extension condition (see [2]). A1 is a continuity property while H1 and H2, respectively, ensure that the noise pdf decays slower than a Gaussian but faster than $1/x^2$. Later we shall see examples of noise pdfs in $\mathcal{N}$ that satisfy neither H1 nor H2. The analytic extension condition requires the function

$$x \mapsto h(x;F) = \int_{\mathbb{R}} p_N(y-x)\log p(y;F)dy$$

to admit an analytic extension on a domain that includes the real line for any $F \in \Omega$.[3] This condition is also required in [14] (see the third robustness condition). In our case, the requirements that characterize $\mathcal{N}$ imply that $h(\cdot;F)$ has an analytic extension (see Proposition 3). We now give examples of noise pdfs in $\mathcal{N}$. As we shall see, while these requirements appear to be cumbersome, in many cases they are easily verified.

A. Exponential Noise

Let α be some even positive integer and consider the noise pdf $p_N(y) = ce^{-y^\alpha}$ with $c = c(\alpha)$ the normalization constant.

Clearly property i. is satisfied, property ii. holds for any $\mu > 0$, and condition iii. is trivially verified for any $\delta > 0$. It remains to find U and L that satisfy property iv. Fix some $\delta > 0$. Since for any $z \in \mathcal{D}_\delta$ we have $|z| \le |\Re z| + \delta$, it follows that

$$|ce^{-z^\alpha}| \ge ce^{-|z|^\alpha} \ge ce^{-(|\Re z|+\delta)^\alpha}. \quad (2)$$

On the other hand, since α is even we have

$$\begin{aligned} |ce^{-z^\alpha}| &= |ce^{-(\Re z + i\Im z)^\alpha}| \\ &= \left| ce^{-[(\Re z)^\alpha + \sum_{j=1}^{\alpha/2} (\Re z)^{\alpha-2j}(i\Im z)^{2j}]} \right| \\ &= ce^{-\left[|\Re z|^\alpha\left(1+O\left(\frac{1}{|\Re z|^2}\right)\right)\right]}. \end{aligned} \quad (3)$$

Now fix some constant β with $\frac{2}{3} < \beta < 1$. From (2) and (3), by defining $L(|\Re z|) = ce^{-(|\Re z|+\delta)^\alpha}$ and $U(|\Re z|) = ce^{-\beta|\Re z|^\alpha}$ we have that for $z \in \mathcal{D}_\delta$, $|z|$ large enough

$$L(|\Re z|) \le |ce^{-z^\alpha}| \le U(|\Re z|). \quad (4)$$

Properties iv.b) and iv.c) follow.

B. Polynomial Noise

Let α be any even positive integer and define the pdf $p_N(y) = \frac{c}{1+y^\alpha}$ with $c = c(\alpha)$ the normalization constant. Property i. is clearly satisfied and given α, property ii. holds by choosing $\mu > 0$ sufficiently small. Now p_N has α poles corresponding to the α roots of -1, none of which are on the real line. By choosing δ smaller than the smallest absolute imaginary part of all these roots, property iii. holds. Then, defining $L(|\Re z|) = \frac{c}{(2|\Re z|)^\alpha}$ and $U(|\Re z|) = \frac{c}{(|\Re z|/2)^\alpha}$, one can easily verify that for $z \in \mathcal{D}_\delta$, $|z|$ large enough

$$L(|\Re z|) \le \left|\frac{c}{1+z^\alpha}\right| \le U(|\Re z|).$$

Property iv.b) is immediate and property iv.c) holds since $\alpha \ge 2$.

Noises with polynomial pdfs as described above are also part of the set of heavy-tailed noises in [2]. An example of a noise pdf in $\mathcal{N}$ which decays slower than heavy-tailed noises is given by $p_N(y) = c\log(2+y^2)/(1+y^2)$. Let

$$L(|\Re z|) = \frac{c\log_2|\Re z|}{(2|\Re z|)^\alpha}$$

and

$$U(|\Re z|) = \frac{2c\log_2|\Re z|}{(|\Re z|/2)^\alpha}.$$

By choosing $\mathcal{D}_\delta$ with $0 < \delta < 1$ we have that for $z \in \mathcal{D}_\delta$, $|z|$ large enough

$$L(|\Re z|) \le \left|\frac{c\log_2(2+z^2)}{1+z^2}\right| \le U(|\Re z|).$$

Properties iv.b) and iv.c) follow.

Also, one can easily construct simple examples of noise pdf in $\mathcal{N}$ that do not satisfy the unimodality constraint required in [9].

III. Proofs of the Results

Lemma 1: Ω is compact in the Levy metric topology and convex.

Proof: See [14]. □

Lemma 2: U and L satisfy

1) $|H(U)|_k| < \infty$
2) $D(U\|L)|_k < \infty$
3) $\int_{y\ge k+x} U(y-x)\log\frac{1}{L(y)}dy < \infty$ for all $x \in \mathbb{R}_+$.

Proof:

1) Since $0 < U(y) \le 1$ for any $y \ge k$, $H(U)|_k \ge 0$ and property iv.b) yields $|H(U)|_k| < \infty$.

2) and 3) Let

$$c \triangleq \int_{y\ge k} U(y)dy.$$

Clearly, $c > 0$ since $U(y) > 0$ for all $y \ge k$. Since

$$0 < L(y) \le U(y) \le 1$$

for all $y \ge k$, for all $x \in \mathbb{R}_+$ we have

$$\begin{aligned} 0 &\le \int_{y\ge k+x} U(y-x)\log\frac{U(y-x)}{L(y)}dy \\ &= -H(U)|_k + \int_{y\ge k+x} U(y-x)\log\frac{1}{L(y)}dy \\ &\overset{(a)}{\le} c\log\left(\frac{1}{c}\int_{y\ge k+x}\frac{U(y-x)^2}{L(y)}dy\right) \\ &\overset{(b)}{<} \infty. \end{aligned} \quad (5)$$

(a) holds by Jensen's inequality. Then, by definition $U(y) \ge L(y)$ for any $y \ge k$. Since U is monotone, $U(y-x) \ge L(y)$ for all $x \ge 0$ and any $y \ge k+x$. Hence property iv.c) yields

$$\int_{y\ge k+x}\frac{U(y-x)^2}{L(y)}dy \le \int_{y\ge k+x}\frac{U(y-x)^3}{L(y)^2}dy < \infty$$

[2]The set of points of increase of a cumulative density function F is defined as

$$S_F \triangleq \{y \in \mathbb{R}:\ F(y+\delta) - F(y-\delta) > 0\,,\ \forall\,\delta > 0\}.$$

In particular, if F admits a density f, S_F is the support of f.

[3]To be precise, the set of heavy-tailed noises is characterized by the requirements A1, A2, H1, and H2 in [2]. If the noise pdf, in addition to satisfying these four conditions has the property that $h(\cdot;F)$ admits an analytic extension on a domain that includes the real line, the corresponding capacity-achieving distribution has finite support. However, we have not mentioned condition A2 since it plays no role in the discrete character of the optimal input distribution but guarantees its uniqueness.

and (b) is justified. Therefore,

$$0 \le -H(U)|_k + \int_{y\ge k+x} U(y-x)\log\frac{1}{L(y)}dy < \infty \tag{6}$$

and since by claim 1. we have $|H(U)|_k < \infty$, it results that

$$\int_{y\ge k+x} U(y-x)\log\frac{1}{L(y)}dy < \infty$$

for any $x \in \mathbb{R}_+$, and Claim 3 is proved. For Claim 2 it suffices to set $x=0$ in (5). □

Lemma 3: For any $p_N \in \mathcal{N}$ the function $p(y;F)$ is continuous in both of its arguments.

Proof: Pick any $F \in \Omega$. Since p_N is continuous (property iii.) and such that $p_N(\pm\infty)=0$, p_N is bounded by M, say. Hence, for any $y \in \mathbb{R}$

$$\int_{-A}^{+A} p_N(y-x)dF(x) \le M < \infty. \tag{7}$$

Fix some $y \in \mathbb{R}$ and pick any sequence $\{y_n\}_{n\ge1}$ converging to y. The Dominated Convergence Theorem yields

$$\lim_{y_n\to y}\int_{-A}^{A} p_N(y_n-x)dF(x) = \int_{-A}^{+A} p_N(y-x)dF(x). \tag{8}$$

Now for a given y, since p_N is continuous and bounded, the definition of the weak convergence yields $\lim_{n\to\infty} p(y;F_n) = p(y;F)$ whenever $F_n \xrightarrow{w} F$. □

For any $F\in\Omega$, $p_N \in \mathcal{N}$, and $y \in \mathbb{R}$ we have by continuity

$$\begin{aligned}0 &< \min_{x\in[-A,A]} p_N(y-x)\\ &\le p(y;F)\\ &\le \max_{x\in[-A,A]} p_N(y-x)\\ &< \infty.\end{aligned} \tag{9}$$

Lemma 4: For any $F \in \Omega$ and $p_N \in \mathcal{N}$, we have

1) for all $|y| \ge k+A$ and $y \in \mathbb{R}$

$$0 < L(|y|+A) \le p(y;F) \le U(|y|-A) \le 1$$

2) $|H(p(\cdot;F))| < \infty$.

Proof:

1) For any $F \in \Omega$ and $p_N \in \mathcal{N}$

$$\begin{aligned}p(y;F) &\triangleq \int_{-A}^{A} p_N(y-x)dF(x)\\ &\overset{(a)}{\le} \int_{-A}^{A} U(|y-x|)dF(x) \text{ for } |y|\ge k+A\\ &\overset{(b)}{\le} U(|y|-A)\\ &\overset{(c)}{\le} 1\end{aligned} \tag{10}$$

where (a) and (c) hold by property iv.a), and (b) follows from the monotonicity of U. The lower bound is obtained in an analoguous way.

2) By Claim 1, we have

$$\begin{aligned}|H(p(\cdot;F))| &\le \int_{|y|\le k+A} p(y;F)|\log p(y;F)|dy\\ &\quad+\int_{|y|\ge k+A} U(|y|-A)\log\frac{1}{L(|y|+A)}dy\\ &< \infty\end{aligned} \tag{11}$$

where the last inequality holds by (9), the continuity of $p(\cdot;F)$ (Lemma 3) and Claim 3 of Lemma 2. □

Proposition 1: For any $p_N \in \mathcal{N}$ and any sequence $\{F_n\}_{n\ge1}$ in Ω with $F_n \xrightarrow{w} F$, we have

$$\lim_{n\to\infty}\int_{\mathbb{R}} p(y;F_n)\log p(y;F_n)dy = \int_{\mathbb{R}} p(y;F)\log p(y;F)dy. \tag{12}$$

Proof: The proof is based on the Dominated Convergence Theorem. Let $\{F_n\}_{n\ge1}$ be a sequence of cdfs in Ω such that $F_n \xrightarrow{w} F$. We have

$$\begin{aligned}0 &\le \left|\int_{\mathbb{R}} p(y;F)\log p(y;F)dy - \int_{\mathbb{R}} p(y;F_n)\log p(y;F_n)dy\right|\\ &\le \int_{\mathbb{R}} |p(y;F)-p(y;F_n)|\,|\log p(y;F)|\,dy\\ &\quad+\int_{\mathbb{R}} p(y;F_n)\,|\log p(y;F)-\log p(y;F_n)|\,dy.\end{aligned} \tag{13}$$

Using (9) and the continuity of $p(\cdot;F)$ (Lemma 3), the integrands of the two integrals of the right-hand side of (13) can be finitely upperbounded over $[-k-A, k+A]$ independently of F_n.

Now, from Claim 1 of Lemma 4, for any $|y| \ge k+A$ and $F \in \Omega$ we have

$$|p(y;F)-p(y;F_n)|\,|\log p(y;F)| \le U(|y|-A)\log\left(\frac{1}{L(|y|+A)}\right) \tag{14}$$

and

$$p(y;F_n)\,|\log p(y;F)-\log p(y;F_n)| \le U(|y|-A)\log\left(\frac{U(|y|-A)}{L(|y|+A)}\right). \tag{15}$$

From the Claims 1 and 3 of Lemma 2, both right-hand sides of (14) and (15) are integrable over $(-\infty,-k-A]\cup[k+A,\infty)$. By the Dominated Convergence Theorem and the continuity of $p(y;\cdot)$, the right-hand side of the second inequality in (13) goes to zero as $n\to\infty$. Since Claim 2 of Lemma 4 yields $|H(p(\cdot;F))| < \infty$ for any $F\in\Omega$, we conclude that

$$\lim_{n\to\infty}\int p(y;F_n)\log p(y;F_n)dy = \int p(y;F)\log p(y;F)dy. \quad □$$

Proposition 2: I is weakly differentiable[4] over Ω.

[4]Let f be a function from Ω into $\mathbb{R}$. Let $\theta\in[0,1]$ and fix $x_0\in\Omega$. If for all $x\in\Omega$ the limit

$$\lim_{\theta\downarrow 0}\frac{f(x_0+\theta(x-x_0))-f(x_0)}{\theta}$$

exists, f is said to be weakly differentiable in Ω at x_0.

Proof: Let F_0 and F be two elements in Ω. For any $0 \le \theta \le 1$ define $F_\theta = F_0 + \theta(F - F_0)$. We shall prove that

$$\lim_{\theta \downarrow 0} \frac{I(F_\theta) - I(F_0)}{\theta} = \int (p(y;F_0) - p(y;F)) \log p(y;F_0) dy. \quad (16)$$

First we prove that the right-hand side of (16) is finite. Claim 1 of Lemma 4 yields

$$\begin{aligned} &\left| \int (p(y;F_0) - p(y;F)) \log p(y;F_0) dy \right| \\ &\quad \le \int_{|y| \le k+A} |(p(y;F_0) - p(y;F)) \log p(y;F_0)| dy \\ &\quad + \int_{|y| \ge k+A} U(|y| - A) \log \frac{1}{L(|y| + A)} dy. \end{aligned} \quad (17)$$

By (9), the continuity of $p(\cdot;F)$, and Claim 3 of Lemma 2, the right-hand side of (17) is finite.

Now, from the proof of the lemma [14, p. 29] we have that

$$\begin{aligned} &\left| \frac{I(F_\theta) - I(F_0)}{\theta} - \int (p(y;F_0) - p(y;F)) \log p(y;F_0) dy \right| \\ &\quad \le \theta \int \frac{[p(y;F) - p(y;F_0)]^2}{p(y;F_0)^2} \\ &\quad \times [2|p(y;F_0) + \theta(p(y;F) - p(y;F_0))| + p(y;F_0)]\, dy \end{aligned} \quad (18)$$

for any $0 \le \theta \le \frac{1}{2}$. We end the proof by showing that the integral on the right-hand side of (18) is finite. This will imply (16) by taking the limit $\theta \downarrow 0$ on both sides of (18). Splitting the integral into two parts we get

$$\begin{aligned} &\int \frac{[p(y;F) - p(y;F_0)]^2}{p(y;F_0)^2} \\ &\quad \times [2|p(y;F_0) + \theta(p(y;F) - p(y;F_0))| + p(y;F_0)] dy \\ &= \int_{|y| \le k+A} \frac{[p(y;F) - p(y;F_0)]^2}{p(y;F_0)^2} \\ &\quad \times [2|p(y;F_0) + \theta(p(y;F) - p(y;F_0))| + p(y;F_0)] dy \\ &\quad + \int_{|y| \ge k+A} \frac{[p(y;F) - p(y;F_0)]^2}{p(y;F_0)^2} \\ &\quad \times [2|p(y;F_0) + \theta(p(y;F) - p(y;F_0))| + p(y;F_0)] dy. \end{aligned} \quad (19)$$

The first integral on the right-hand side of (19) is finite by (9) and the continuity of $p(\cdot;F)$. For the second integral, Claim 1 of Lemma 4 yields

$$\begin{aligned} &\int_{|y| \ge k+A} \frac{[p(y;F) - p(y;F_0)]^2}{p(y;F_0)^2} \\ &\quad \times [2|p(y;F_0) + \theta(p(y;F) - p(y;F_0))| + p(y;F_0)] dy \\ &\quad \le 5 \int_{|y| \ge k+A} \frac{U(|y| - A)^3}{L(|y| + A)^2} dy \\ &\quad < \infty. \end{aligned} \quad (20)$$

□

As we shall see later, the next proposition represents the main point in the proof of the theorem.

Proposition 3: For any $p_N \in \mathcal{N}$ and any $F \in \Omega$, the function $h(\cdot;F) : \mathcal{D}_\delta \to \mathbb{C}$ defined by

$$z \mapsto h(z;F) \triangleq \int_{\mathbb{R}} p_N(y - z) \log(p(y;F)) dy$$

is analytic over $\mathcal{D}_\delta$.

Proof: We prove the proposition by means of Morera's theorem (converse of Cauchy's theorem, e.g., see [11]).

We first show that for any F, $z \mapsto h(z;F)$ is continuous over $\mathcal{D}_\delta$. Pick some z in $\mathcal{D}_\delta$. Then there exists $\varepsilon > 0$ with $\overline{B(z;\varepsilon)} \in \mathcal{D}_\delta$. Let $(z_l)_{l \ge 1}$ be a sequence of complex numbers in $\mathcal{D}_\delta$ converging to z and let t be such that $z_l \in \overline{B(z;\varepsilon)}$ for any $l \ge t$. For any $l \ge t$ we have

$$\begin{aligned} &\int |p_N(y - z_l)| |\log(p(y;F))| dy \\ &\quad \le \int \max_{\xi \in \overline{B(z;\varepsilon)}} |p_N(y - \xi)| |\log p(y;F)| dy \\ &\quad \le \int_{y:|y| \le k+A+|z|+\varepsilon} \max_{\xi \in \overline{B(z;\varepsilon)}} |p_N(y - \xi)| \\ &\qquad \times |\log p(y;F)| dy \\ &\qquad + \int_{y:|y| \ge k+A+|z|+\varepsilon} \max_{\xi \in \overline{B(z;\varepsilon)}} |p_N(y - \xi)| \\ &\qquad \times |\log p(y;F)| dy \\ &\quad \overset{(a)}{\le} M \int_{y:|y| \le k+A+|z|+\varepsilon} |\log p(y;F)| dy \\ &\qquad + \int_{y:|y| \ge k+A+|z|+\varepsilon} \max_{\xi \in \overline{B(z;\varepsilon)}} U(|y - \Re\xi|) \\ &\qquad \times \log \frac{1}{L(|y| + A)} dy \\ &\quad \overset{(b)}{\le} M \int_{y:|y| \le k+A+|z|+\varepsilon} |\log p(y;F)| dy \\ &\qquad + \int_{y:|y| \ge k+A+|z|+\varepsilon} U(|y| - |z| - \varepsilon) \\ &\qquad \times \log \frac{1}{L(|y| + A)} dy \\ &\quad \overset{(c)}{<} \infty \end{aligned} \quad (21)$$

where

$$M = \max_{y:|y| \le k+A+|z|+\varepsilon} \max_{\xi \in \overline{B(z;\varepsilon)}} |p_N(y - \xi)|$$

is finite by compactness and continuity.[5] (a) holds by (9) and property iv.a). Inequality (b) is justified by noting that U is a decreasing function and that for any $\xi \in \overline{B(z;\varepsilon)}$, the triangle inequality yields $|\Re\xi| \le |\xi| \le |z| + \varepsilon$. (c) follows from (9) and Claim 3 of Lemma 2. Applying the Dominated Convergence Theorem we have

$$\lim_{z_l \to z} h(z_l;F) = h(z;F) \quad (22)$$

which proves the continuity of $h(\cdot;F)$ over $\mathcal{D}_\delta$.

[5]Here we see the reason why we restricted the extension domain $\mathcal{D}_\delta$ to be a symmetric band around the real axis. $\mathcal{D}_\delta$ is such that for all $y \in \mathbb{R}$

$$\overline{B(z + y;\varepsilon)} \subset \mathcal{D}_\delta$$

and

$$\overline{B(z + y;\varepsilon)} \subset \mathcal{D}_\delta$$

if and only if $\overline{B(-z + y;\varepsilon)} \subset \mathcal{D}_\delta$. Hence the domain for the calculation of M (see the shaded area in Fig. 1) is included into $\mathcal{D}_\delta$, and thus M is finite.

Consider now a compact triangle[6] $\triangle$ in $\mathcal{D}_\delta$ having perimeter length $|\triangle|$. Let d be the length of the longest side of the triangle and $s \triangleq \min_{z\in\triangle}|z|$. By the triangle inequality $|\xi| \le s+d$ for all $\xi \in \triangle$, an analoguous argument as for continuity yields

$$\begin{aligned}\int_{\partial\triangle} &|h(z;F)|dz \\ &\le \int_{\mathbb{R}} dy \int_{\partial\triangle} |p_N(y-z)||\log p(y;F)|dz \\ &\le |\triangle| M \int_{|y|\le k+A+s+d} |\log p(y;F)|dy \\ &\quad + |\triangle| \int_{|y|\ge k+A+s+d} U(|y|-s-d) \\ &\quad \times \log \frac{1}{L(|y|+A)} dy \\ &< \infty \end{aligned} \tag{23}$$

where

$$M = \max_{y:|y|\le k+A+s+d} \max_{\xi\in\triangle} |p_N(y-\xi)| < \infty.$$

From Fubini's theorem we get

$$\begin{aligned}\int_{\partial\triangle} h(z;F)dz &= \int_{\partial\triangle} dz \int_{\mathbb{R}} p_N(y-z)\log(p(y;F))dy \\ &= \int_{\mathbb{R}} \log(p(y;F))dy \underbrace{\int_{\partial\triangle} p_N(y-z)dz}_{\overset{(a)}{=}0} \\ &= 0 \end{aligned} \tag{24}$$

where (a) holds by analyticity (see also footnote 5). Since $h(\cdot;F)$ is continuous and satisfies (24) for compact triangles, by Morera's theorem we conclude that $h(\cdot;F)$ is analytic over $\mathcal{D}_\delta$. □

Proof of the Theorem: Let $p_N \in \mathcal{N}$. We first prove that $|H(p_N)| < \infty$. By property iv.a) we have

$$\begin{aligned}|H(p_N)| &\le \int_{y:|y|\le k} p_N(y)|\log p_N(y)|dy \\ &\quad + \int_{y:|y|\ge k} U(|y|)\log\frac{1}{L(|y|)}dy \\ &< \infty. \end{aligned} \tag{25}$$

The second inequality holds by continuity of p_N, the fact that $p_N(y) > 0$ for all $y \in \mathbb{R}$, and Claim 3 of Lemma 2.

Since for an additive channel, $I(F) = H(p(\cdot;F)) - H(p_N)$, from Proposition 1 we then deduce that I is a concave continuous function over Ω. By Lemma 1, Ω is compact and hence,

$$C = \sup_{F\in\Omega} I(F) = \max_{F\in\Omega} I(F). \tag{26}$$

By Proposition 2, the function I is weakly differentiable. Hence from the Kuhn–Tucker conditions (see [14, Corollary 1]), a necessary and sufficient condition for an input cdf F^* to achieve capacity is to satisfy

$$\int p_N(y-x)\log p(y;F^*)dy \ge -(C+H(p_N)) \tag{27}$$

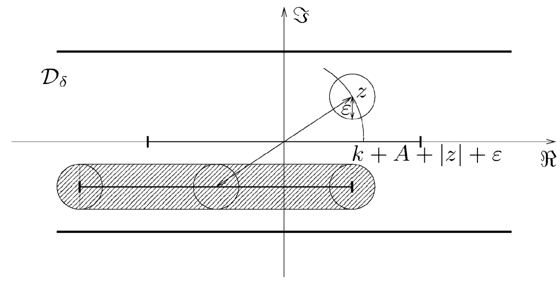

Fig. 1.

for all $x \in [-A, A]$ with equality if x belongs to the set S_{F^*} of points of increase of F^*.

From Proposition 3 we have that $\int p_N(y-z)\log p(y;F^*)dy$ is analytic over a region that includes the real line. Suppose now that S_{F^*} admits an accumulation point. We conclude by the Identity Theorem of Complex Analysis that

$$\int p_N(y-x)\log p(y;F^*)dy = -(C+H(p_N)) \tag{28}$$

for every $x \in \mathbb{R}$. We now show that (28) leads to a contradiction. Let $q \triangleq e^{C+H(p_N)}$ and define

$$\mathcal{S}^+ \triangleq \{y \in \mathbb{R} : \log(qp(y;F^*)) \ge 0\} \tag{29}$$

and

$$\mathcal{S}^- \triangleq \{y \in \mathbb{R} : \log(qp(y;F^*)) \le 0\}. \tag{30}$$

We can restate (28) as

$$\int_{\mathcal{S}^+} p_N(y-x)\log(qp(y;F^*))dy \tag{31}$$

$$= -\int_{\mathcal{S}^-} p_N(y-x)(\log(qp(y;F^*)))dy \tag{32}$$

for every $x \in \mathbb{R}$. By property ii. and Claim 1 of Lemma 4, one can choose some $l \ge k+A$ that satisfies the following two conditions:

$$l^\mu \ge 2\mathbb{E}|N|^\mu$$

and

$$\log(qp(y;F^*)) \le \log(qU(l-A)) < 0$$

for all $|y| \ge l$.

On the one hand, since $\mathcal{S}^+ \subseteq [-l, l]$, by setting $M \triangleq \max_{y\in\mathcal{S}^+}\log(qp(y;F^*))$,[7] it follows that

$$\begin{aligned} 0 &\le \int_{\mathcal{S}^+} p_N(y-x)\log(qp(y;F^*))dy \\ &\le M\int_{-l}^{l} p_N(y-x)dy. \end{aligned} \tag{33}$$

Choosing $x \ge l+k$, property iv.a) yields

$$\int_{-l}^{l} p_N(y-x)dy \le 2lU(x-l) \tag{34}$$

and, therefore, from (33) we get

$$\begin{aligned} 0 &\le \int_{\mathcal{S}^+} p_N(y-x)(\log(qp(y;F^*))dy \\ &\le 2lMU(x-l). \end{aligned} \tag{35}$$

[6]Whenever z_1, z_2, z_3 are three points in $\mathbb{C}$, the compact set

$$\triangle = \{z \in \mathbb{C} : z = sz_1 + tz_2 + uz_3, s+t+u = 1, s,t,u \ge 0\}$$

is called compact triangle.

[7]M exists by continuity of $\log p(\cdot;F)$.

Since $U(x) \to 0$ as $x \to \infty$, the quantity

$$\int_{\mathcal{S}+} p_N(y-x)(\log(qp(y;F^*))dy$$

can be rendered arbitrary small by taking x large enough.

On the other hand, we have

$$-\int_{\mathcal{S}-} p_N(y-x)\log(qp(y;F^*))dy \geq -\int_l^\infty p_N(y-x)(\log(qp(y;F^*))dy. \quad (36)$$

Hence, from the two conditions satisfied by l, we have

$$\begin{aligned} -\int_l^\infty p_N(y-x)(\log(qp(y;F^*))dy &\geq -\log(qU(l-A))\int_l^\infty p_N(y-x)dy \\ &\overset{(a)}{\geq} -\log(qU(l-A))\left(1-\frac{\mathbb{E}|N|^\mu}{l^\mu}\right) \\ &\geq -\frac{1}{2}\log(qU(l-A)) > 0 \end{aligned} \quad (37)$$

where (a) holds for $x \geq 2l$ by Markov's inequality. Therefore, combining (36), (37) and (35), (31) is impossible for x sufficiently large. By contradiction, S_{F^*} admits no accumulation point. Since S_{F^*} is bounded, we conclude by the Bolzano–Weierstrass theorem that S_{F^*} is finite. □

The proof of the theorem follows Smith's argument: from the Kuhn–Tucker conditions derive a contradiction using the analytic extension property of $h(\cdot;F)$. However, in Smith's proof it seems that there is a minor glitch, precisely in the derivation of the contradiction. Let $t(y)$ be a locally integrable function over $\mathbb{R}$ with $T(f)$ as Fourier transform (in the sense of distribution). Let G_N be the characteristic function of the noise pdf and assume that $G_N(f) \neq 0$ for all $f \in \mathbb{R}$. It is claimed that if $\int G_N(f)T(f)S(f)df = 0$ for all Schwartz functions S, the distribution T pointwise is equal to zero (see [14, Appendix A. 6, p. 102]).[8] This implication is not clear. Even in the case we are interested in, i.e., $t(y) = \log(qp(y;F))$ the claim does not seem justified. In the proof we proposed here we used another argument that makes use of the requirement ii. (see (28)) and does not rely on the assumption that the characteristic function of the noise pdf is always nonzero (see fourth robustness condition [14]).

IV. Conclusion

We exhibited a family of noise pdfs for which the corresponding channel has discrete capacity-achieving distribution(s), thereby extending [9], [14], [2]. At this point, some issues related to the discrete character of optimal inputs are suitable for investigation. The optimal distribution for an additive Gaussian channel with bounded input has discrete support, while when the input has only power limitation, the optimal distribution is Gaussian. It would therefore be interesting to study the sensitivity of the discrete nature of optimal distributions to input constraints. For instance, if F_n^* denotes the optimal input distribution of a Gaussian channel whose input's cdf is restricted to the set $\{F : \int_{\mathbb{R}} |x|^n dF(x) < \infty\}$, then F_2^* admits a Gaussian density. On the other hand, Smith's result suggests that F_∞^* has a finite set of points of increase. What happens between $n=2$ and $n=\infty$?

Another issue might be to exhibit, if any, a (simple) example of a non-Gaussian additive channel, with input average power constraint, whose optimal distribution is not discrete. One way to tackle the problem would be to start with a Gaussian additive channel then modify the Gaussianity of the noise, and see how the corresponding optimal input reacts in terms of support.

Oettli's work, which is based on convex analysis, is of particular interest since it derives general results concerning the discreteness of capacity-achieving distribution(s) without the use of the analytic extension arguments.[9] It might be interesting to extend the results we have so far to a wider class of additive channels. From previous works it seems that Gaussian noise with average power constraint is a very special case, and that a large class of other combinations of noise–input constraints yield discrete optimal distribution(s) (even for non-purely additive channels [1]).

[8] In [15 (proof of Proposition 2)], the author also refers to that claim.

[9] In particular the noise pdf might be discontinuous.

Acknowledgment

The author is grateful to İ. E. Telatar for many valuable discussions, and K. Viswanathan for comments on the final manuscript.

References

[1] I. C. A. Faycal, M. D. Trott, and S. Shamai (Shitz), "The capacity of discrete-time memoryless rayleigh fading channels," *Trans. Inform. Theory*, vol. 47, pp. 1290–1301, May 2001.

[2] A. Das, "Capacity-achieving distributions for non-Gaussian additive noise channels," in *Proc. IEEE Int. Symp. Information Theory*, Sorrento, Italy, June 2000, p. 432.

[3] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[4] M. C. Gursoy, H. V. Poor, and S. Verdú, "On the capacity-achieving distribution of the noncoherent rician fading channel," in *Proc. Canadian Workshop on Information Theory*, Waterloo, ON, Canada, May 18–21, 2003.

[5] P. Hou, B. J. Belzer, and T. R. Fisher, "On the capacity of the Partially coherent additive white Gaussian noise channel," in *Proc. IEEE Int. Symp. Information Theory*, Yokohama, Japan, June/July 2003, p. 372.

[6] D. G. Luenberger, *Optimization by Vector Space Methods*. New York: Wiley, 1969.

[7] M. Katz and S. Shamai (Shitz), "On the capacity-achieving distribution of the discrete-time noncoherent additive white Gaussian noise channel," in *Proc. IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, June/July 2002, p. 165.

[8] R. Nuriyev and A. Anastasopoulos, "Capacity characterization for the non-coherent block independent AWGN channel," in *Proc. IEEE Int. Symp. Information Theory*, Yokohama, Japan, June/July 2003, p. 373.

[9] W. Oettli, "The capacity-achieving input distribution for some amplitude-limited channels with additive noise," *IEEE Trans. Inform. Theory*, vol. IT–20, pp. 372–374, May 1974.

[10] R. Palanki, "On the capacity achieving distributions of some fading channels," in *Proc. 40th Annu. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2002, pp. 337–346.

[11] R. Remmert, *Theory of Complex Functions* New York, 1998.

[12] S. Shamai (Shitz) and I. Bar-David, "The capacity of average and peak-power limited quadrature Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1060–1071, July 1995.

[13] A. N. Shiryayev, *Probability* Berlin, Germany, 1984.

[14] J. G. Smith, "On the information capacity of peak and average power constrained Gaussian channels," Ph.D. dissertation, Dept. Elec. Eng., Univ. Calif., Berkeley, CA, 1969.

[15] ——, "The information capacity of amplitude and variance constrained scalar Gaussian channel," *Inform. Contr.*, vol. 18, pp. 203–219, 1971.

[16] K. Yosida, *Analysis* Berlin, Germany, 1980.