This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the ICC 2007 proceedings.

Capacity-Achieving Discrete Signaling over Additive Noise Channels

Anke Feiten
Institute of Theoretical Information Technology
RWTH Aachen University
Sommerfeldstr. 24, 52074 Aachen, Germany
Email: feiten@ti.rwth-aachen.de

Rudolf Mathar
Institute of Theoretical Information Technology
RWTH Aachen University
Sommerfeldstr. 24, 52074 Aachen, Germany
Email: mathar@ti.rwth-aachen.de

Abstract— **Discrete input distributions are capacity-achieving for a variety of noise distributions whenever the input is subject to peak power or other bounding constraints. In this paper, we consider additive noise with arbitrary absolutely-continuous distribution and ask the question what the optimal input distribution over a set of fixed signaling points would be. The capacity-achieving distribution is characterized by constant Kullback Leibler distance between the shifted noise distribution and a certain mixture hereof. As an application, the optimal input distribution for binary symmetric signaling over exponential noise channels is determined. It further follows that in certain symmetric cases the uniform distribution over all signaling points is capacity-achieving.**

I. Introduction

Shannon showed that the scalar additive Gaussian noise channel has a Gaussian capacity-achieving input distribution whenever average power constraints apply. This result was extended to Gaussian vector channels, particularly MIMO channels by [1]. Due to the unlimited support, this input distribution is not realizable in practice. To avoid unbounded power requirements for the transmitter, peak power constraints are added. Interestingly, the capacity achieving input distribution becomes discrete with finite support for the scalar Gaussian channel subject to average and peak power constraints as was demonstrated by [2]. Other channels like Poisson, quadrature Gaussian and additive vector Gaussian were shown to possess a discrete capacity-achieving input distribution under average and peak power constraints as surveyed in [3]. This work and [4] generalize a number of previous papers on the subject by considering conditionally Gaussian vector channels subject to bounded-input constraints described by a bounded and closed support $\mathcal{S} \subset \mathbb{R}^N$. Under certain conditions on $\mathcal{S}$ the capacity achieving distribution is discrete, which includes the previously mentioned channels as special cases.

A related question is investigated in [5]. The optimum constellation of M equiprobable complex signals is sought for an additive Gaussian channel under average power constraints such that the error probability is minimum.

Since for practically relevant cases the capacity-achieving distribution subject to bounded input is discrete, from a practical point of view we may ask the following question. Given a bounded and closed subset $\mathcal{S} \subset \mathbb{R}^N$, further a maximum number of support points M, what is the optimum choice of signaling points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_M \in \mathcal{S}$ and the optimum distribution $(p_1, \dots, p_M)$ such that the input variable $\boldsymbol{X}$ with distribution $P(\boldsymbol{X} = \boldsymbol{x}_i) = p_i$, $i = 1, \dots, M$, maximizes the mutual information between channel input and output, and hence is capacity-achieving in the set of discrete distributions on $\mathcal{S}$ with at most M support points. If M is greater than the number of points of increase of the capacity-achieving bounded-input distribution, an answer is given for cases considered in [3]. In general, however, this seems to be a hard problem, methodologically related to the theory of optimum design of experiments.

In this paper, we address the special case that $\mathcal{S}$ itself consists of finitely many points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_M$ representing the set of inputs that can be generated by the transmitter. We aim at determining the capacity-achieving distribution $(p_1, \dots, p_M)$ for arbitrary additive noise channels. An intuitive characterization of the optimum distribution in terms of Kullback Leibler distances is derived exploiting directional derivatives. Furthermore, explicit solutions are given for certain symmetric noise distributions and corresponding signal constellations.

II. System Model and Prerequisites

We consider the additive noise channel

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{X}$ represents the discrete random input of dimension N with support points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_M \subset \mathbb{R}^N$. The stochastically independent noise vector $\boldsymbol{n}$ is assumed to have (Lebesgue) density $\varphi(\boldsymbol{z})$, $\boldsymbol{z} \in \mathbb{R}^N$, and finite entropy $|H(\boldsymbol{n})| < \infty$. We aim at determining the capacity-achieving probability distribution $\boldsymbol{p} = (p_1, \dots, p_M)$ with $P(\boldsymbol{X} = \boldsymbol{x}_i) = p_i$, $i = 1, \dots, M$.

The mutual information of channel model (1) is computed as

$$\begin{aligned} I(\boldsymbol{X};\boldsymbol{Y}) &= H(\boldsymbol{Y}) - H(\boldsymbol{Y} \mid \boldsymbol{X}) \\ &= H(\boldsymbol{Y}) - H(\boldsymbol{X} + \boldsymbol{n} \mid \boldsymbol{X}) \\ &= H(\boldsymbol{Y}) - H(\boldsymbol{n}), \end{aligned} \tag{2}$$

where H denotes entropy. Observe that $\boldsymbol{Y}$ is absolutely-continuous with density function

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \sum_{i=1}^{M} p_i \varphi_i(\boldsymbol{y}) \tag{3}$$

where $\varphi_i(\boldsymbol{y}) = \varphi(\boldsymbol{y} - \boldsymbol{x}_i)$ denotes the noise density shifted by $\boldsymbol{x}_i$. Hence, the distribution of $\boldsymbol{Y}$ is a mixture of densities φ_i with coefficients p_i.

Since

$$\begin{aligned} H(\boldsymbol{n}) &= -\int \varphi(\boldsymbol{y}) \log \varphi(\boldsymbol{y}) d\boldsymbol{y} \\ &= -\int \varphi_i(\boldsymbol{y}) \log \varphi_i(\boldsymbol{y}) d\boldsymbol{y}, \end{aligned} \tag{4}$$

we may write

$$\begin{aligned} I(\boldsymbol{X};\boldsymbol{Y}) &= H(\boldsymbol{Y}) - H(\boldsymbol{n}) \\ &= \sum_{i=1}^{M} p_i \int \varphi_i(\boldsymbol{y}) \log \frac{\varphi_i(\boldsymbol{y})}{\sum_{j=1}^{M} p_j \varphi_j(\boldsymbol{y})} d\boldsymbol{y} \\ &= \sum_{i=1}^{M} p_i \, D\Big(\varphi_i \Big\| \sum_{j=1}^{M} p_j \varphi_j\Big), \end{aligned} \tag{5}$$

where $D(g\|h) = \int g \log \frac{g}{h}$ denotes the Kullback Leibler distance or, synonymously, the relative entropy between densities g and h, see [6].

In the following we will use directional derivatives to characterize capacity-achieving distributions and give explicit solutions in special symmetric cases. Let f be a concave function with convex domain $\mathcal{C}$, and let $\hat{x}, x \in \mathcal{C}$. The directional derivative of f at $\hat{x}$ in the direction of x is defined as

$$\begin{aligned} Df(\hat{x}, x) &= \lim_{\alpha \to 0+} \frac{1}{\alpha} \Big[f\big((1-\alpha)\hat{x} + \alpha x\big) - f(\hat{x}) \Big] \\ &= \frac{d}{d\alpha} f\big((1-\alpha)\hat{x} + \alpha x\big)\Big|_{\alpha=0+}. \end{aligned}$$

Since f is concave the ratio $\big(f((1-\alpha)\hat{x} + \alpha x) - f(\hat{x})\big)/\alpha$ is monotone increasing with decreasing $1 \geq \alpha \geq 0$, and the directional derivative always exists. Directional derivatives are also employed to determine capacity-achieving distributions for general average p-norm constraints, see [7].

We will furthermore exploit the well known fact that $\hat{x} \in \mathcal{C}$ is a maximizing point of f if and only if $Df(\hat{x}, x) \leq 0$ for all $x \in \mathcal{C}$, see, e.g., [8].

III. Capacity-Achieving Input Distributions

Candidate distributions are contained in the set of all stochastic vectors of dimension M, namely

$$\mathcal{C} = \{\boldsymbol{p} = (p_1, \ldots, p_M) \mid p_i \geq 0, \sum_{i=1}^{M} p_i = 1\}$$

Obviously, $\mathcal{C}$ is a convex set. The capacity of channel (1) is defined as

$$C = \max_{(p_1,\ldots,p_M)\in\mathcal{C}} I(\boldsymbol{X};\boldsymbol{Y}).$$

By equation (2), distribution $\hat{\boldsymbol{p}} = (\hat{p}_1, \ldots, \hat{p}_M)$ achieves capacity, i.e., maximizes mutual information, if and only if $H(\boldsymbol{Y})$ is maximized by $\hat{\boldsymbol{p}}$. Using density (3) our task is to determine

$$\max_{(p_1,\ldots,p_M)\in\mathcal{C}} -\int \Big(\sum_{i=1}^{M} p_i \, \varphi_i(\boldsymbol{y})\Big) \log \Big(\sum_{i=1}^{M} p_i \, \varphi_i(\boldsymbol{y})\Big) d\boldsymbol{y}.$$

The function

$$f(p_1, \ldots, p_n) = -\int \Big(\sum_{i=1}^{M} p_i \, \varphi_i(\boldsymbol{y})\Big) \log \Big(\sum_{i=1}^{M} p_i \, \varphi_i(\boldsymbol{y})\Big) d\boldsymbol{y}$$

is concave as can be shown by elementary means using the fact that $t \log t$ is convex for positive t. The gradient of $\nabla f = \big(\frac{df}{dp_i}\big)_{i=1,\ldots,M}$ is obtained as

$$\frac{df}{dp_i} = -(\log e)\Big(1 + \int \varphi_i(\boldsymbol{y}) \log \Big(\sum_{j=1}^{M} p_j \varphi_j(\boldsymbol{y})\Big) d\boldsymbol{y}\Big).$$

To compute $\frac{df}{dp_i}$ at $\boldsymbol{p} \in \mathcal{C}$ we interchange differentiation and integration. According to Lemma 16.2 in [9] this is permissible since each partial derivative is dominated by an integrable function as may be seen from the following chain of inequalities. Let $m(\boldsymbol{y}) = \max\{\varphi_1(\boldsymbol{y}), \ldots, \varphi_m(\boldsymbol{y})\}$, which is an integrable function. Then

$$\begin{aligned} \varphi_i(\boldsymbol{y}) \, \Big| \log \Big(\sum_{j=1}^{M} p_j \varphi_j(\boldsymbol{y})\Big) \Big| &\leq \varphi_i(\boldsymbol{y}) \big| \log m(\boldsymbol{y}) \big| \\ &\leq \varphi_i(\boldsymbol{y}) \big| \log m(\boldsymbol{y}) - \log \varphi_i(\boldsymbol{y}) \big| + \varphi_i(\boldsymbol{y}) \big| \log \varphi_i(\boldsymbol{y}) \big| \\ &\leq \varphi_i(\boldsymbol{y}) \Big| \log \frac{m(\boldsymbol{y})}{\varphi_i(\boldsymbol{y})} \Big| + \varphi_i(\boldsymbol{y}) \big| \log \varphi_i(\boldsymbol{y}) \big| \\ &\leq (\log e) \varphi_i(\boldsymbol{y}) \Big| \frac{m(\boldsymbol{y})}{\varphi_i(\boldsymbol{y})} - 1 \Big| + \varphi_i(\boldsymbol{y}) \big| \log \varphi_i(\boldsymbol{y}) \big| \\ &= (\log e) \big| m(\boldsymbol{y}) - \varphi_i(\boldsymbol{y}) \big| + \varphi_i(\boldsymbol{y}) \big| \log \varphi_i(\boldsymbol{y}) \big|. \end{aligned}$$

The last line constitutes an integrable upper bound independent of $\boldsymbol{p}$ since $H(\boldsymbol{n})$ is assumed to exist and to be finite.

The directional derivative of f at $\hat{\boldsymbol{p}}$ in the direction of $\boldsymbol{p}$ now follows from representation

$$\begin{aligned} Df(\hat{\boldsymbol{p}}, \boldsymbol{p}) &= \langle \nabla f(\hat{\boldsymbol{p}}), \boldsymbol{p} - \hat{\boldsymbol{p}} \rangle \\ &= -(\log e) \sum_{i=1}^{M} (p_i - \hat{p}_i) \int \varphi_i(\boldsymbol{y}) \log \Big(\sum_{j=1}^{M} \hat{p}_j \varphi_j(\boldsymbol{y})\Big) d\boldsymbol{y}. \end{aligned}$$

For notational convenience let

$$\begin{aligned} b_i(\hat{\boldsymbol{p}}) &= b_i(\hat{p}_1, \ldots, \hat{p}_M) \\ &= \int \varphi_i(\boldsymbol{y}) \log \Big(\sum_{j=1}^{M} \hat{p}_j \varphi_j(\boldsymbol{y})\Big) d\boldsymbol{y}. \end{aligned}$$

From the concavity of the logarithm it can be easily concluded that $b_i(\boldsymbol{p})$ is a concave function on $\mathcal{C}$ for any $i = 1, \ldots, M$.

Now, the maximum of f is attained at $\hat{\boldsymbol{p}} \in \mathcal{C}$ if and only if the directional derivatives at $\hat{\boldsymbol{p}}$ in any direction $\boldsymbol{p} \in \mathcal{C}$ is non-positive, i.e.,

$$-\sum_{i=1}^{M}(p_i - \hat{p}_i)\, b_i(\hat{\boldsymbol{p}}) \le 0.$$

Hence, $\hat{\boldsymbol{p}}$ is an optimum point if and only if

$$\sum_{i=1}^{M}\hat{p}_i\, b_i(\hat{\boldsymbol{p}}) = \min_{\boldsymbol{p}\in\mathcal{C}} \sum_{i=1}^{M} p_i\, b_i(\hat{\boldsymbol{p}}) = \min_{i=1,\dots,M} b_i(\hat{\boldsymbol{p}}).$$

Equality is obviously achieved if and only if $b_i(\hat{\boldsymbol{p}})$ equals some constant ζ, say, for all i with $\hat{p}_i > 0$. In summary, we have derived the following result.

Proposition 1: Given signaling points $\boldsymbol{x}_1,\dots,\boldsymbol{x}_M$ for the input variable $\boldsymbol{X}$ in channel model (1), distribution $\hat{\boldsymbol{p}}$ is capacity-achieving if and only if

$$\int \varphi_i(\boldsymbol{y}) \log\Big(\sum_{j=1}^{M}\hat{p}_j\varphi_j(\boldsymbol{y})\Big)d\boldsymbol{y} = \zeta \tag{6}$$

for some $\zeta\in\mathbb{R}$ and all indices i with $\hat{p}_i > 0$.

Note that integral (6) is well defined since the measure corresponding to the mixture density always dominates the measure corresponding to φ_i for positive $\hat{p}_i$.

A different way to achieve Proposition 1 would be to use Lagrangian multipliers and exploit the Kuhn Tucker conditions for objective function f and constraints given by $\mathcal{C}$.

The Kullback Leibler distance or relative entropy between φ_i and the mixture of φ_j with coefficients $\hat{p}_1,\dots,\hat{p}_M$ from the optimum $\hat{\boldsymbol{p}}$ is given by

$$\begin{aligned} D\Big(\varphi_i \,\Big\|\, \sum_{j=1}^{M}\hat{p}_j\varphi_j\Big) &= \int \varphi_i(\boldsymbol{y})\log\varphi_i(\boldsymbol{y})d\boldsymbol{y} \\ &\quad - \int \varphi_i(\boldsymbol{y})\log\Big(\sum_{j=1}^{M}\hat{p}_j\varphi_j(\boldsymbol{y})\Big)d\boldsymbol{y}. \end{aligned} \tag{7}$$

The first integral on the right hand side of (7) is independent of $i = 1,\dots,M$ since $\varphi_i(\boldsymbol{y}) = \varphi(\boldsymbol{y}-\boldsymbol{x}_i)$ is a linear shift in argument only. The second term has constant value according to Proposition 1. Hence, for optimum $\hat{\boldsymbol{p}}$ the relative entropy in (7) is constant for all i with positive $\hat{p}_i$. Moreover, from representation (5), the channel capacity is given by this constant Kullback-Leibler distance. In summary, the following result holds.

Proposition 2: Distribution $\hat{\boldsymbol{p}}$ is capacity-achieving if and only if

$$D\Big(\varphi_i \,\Big\|\, \sum_{j=1}^{M}\hat{p}_j\varphi_j\Big) = \xi \tag{8}$$

for some $\xi \ge 0$ and all indices i with $\hat{p}_i > 0$. Moreover, the channel capacity amounts to

$$C = \max_{\boldsymbol{p}\in\mathcal{C}} I(\boldsymbol{X};\boldsymbol{Y}) = \xi.$$

Equation (8) has an interesting interpretation. For an input distribution $\hat{\boldsymbol{p}} = (\hat{p}_1,\dots,\hat{p}_M)$ over fixed signaling points $\boldsymbol{x}_1,\dots,\boldsymbol{x}_M$ to be capacity-achieving, the relative entropy between densities $\varphi_i(\boldsymbol{y}) = \varphi(\boldsymbol{y}-\boldsymbol{x}_i)$ and the mixture distribution with weights $\hat{p}_j$ has to be the same for all i with $\hat{p}_i > 0$. Hence, the capacity-achieving distribution $\hat{\boldsymbol{p}}$ places the mixture density $\sum_{j=1}^{M}\hat{p}_j\varphi_j$ somehow in the middle of all shifted densities φ_i with positive coefficient $\hat{p}_i$. Distance is measured by the Kullback Leibler pseudo-metric.

Example 1. Exponential noise, cf. [10].
Let $\lambda > 0$ and consider the exponential noise probability density function

$$\varphi(z) = \lambda\, e^{-\lambda z}\, \mathbb{1}_{(0,\infty)}(z),$$

for the scalar channel $Y = X + n$. Let X have fixed support points $x_1 = -\mu$ and $x_2 = \mu > 0$, selected for signaling with probabilities p_1 and p_2, respectively. Capacity is achieved at points $p_i > 0$, since $p_1 = 0$ or $p_2 = 0$ yields mutual information zero. According to Proposition 1 distribution (p_1, p_2) is capacity achieving if

$$\begin{aligned} &\int \varphi(y+\mu)\ln\big(p_1\varphi(y+\mu)+p_2\varphi(y-\mu)\big)dy \\ &= \int \varphi(y-\mu)\ln\big(p_1\varphi(y+\mu)+p_2\varphi(y-\mu)\big)dy, \end{aligned}$$

concretely for the exponential distribution,

$$\begin{aligned} &\int_{\mu}^{\infty} \lambda e^{-\lambda(y+\mu)}\ln\big(\lambda e^{-\lambda y}(p_1e^{-\lambda\mu}+p_2e^{\lambda\mu})\big)dy \\ &\quad + \int_{-\mu}^{\mu} \lambda e^{-\lambda(y+\mu)}\ln\big(\lambda e^{-\lambda y}p_1e^{-\lambda\mu}\big)dy \\ &= \int_{\mu}^{\infty} \lambda e^{-\lambda(y-\mu)}\ln\big(\lambda e^{-\lambda y}(p_1e^{-\lambda\mu}+p_2e^{\lambda\mu})\big)dy. \end{aligned} \tag{9}$$

Let

$$\beta = \frac{2\lambda\mu}{1-e^{-2\lambda\mu}} \quad \text{and} \quad \alpha = \frac{e^{\beta}-1}{e^{2\lambda\mu}}.$$

Rather tedious algebra gives an explicit solution of equation (9) as

$$\hat{p}_1 = \frac{1}{1+\alpha}, \quad \hat{p}_2 = \frac{\alpha}{1+\alpha}.$$

Channel capacity is given by the according Kullback Leibler distance (8).

Figure 1 shows for $\lambda = 1$ the probabilities p_1 (solid line) and p_2 (dashed line) as a function of the signaling point $\mu \in (0,3)$ for parameter $\lambda = 1$.

IV. Circular Symmetric Noise

Throughout this section we assume symmetric noise density $\varphi(\boldsymbol{y})$ in the sense that

$$\varphi(\boldsymbol{y}) = \varphi(\boldsymbol{T}\boldsymbol{y})$$

for any orthogonal $N\times N$ matrix $\boldsymbol{T}$. Zero mean Gaussian noise with covariance matrix a multiple of the identity matrix is a standard example of this situation.

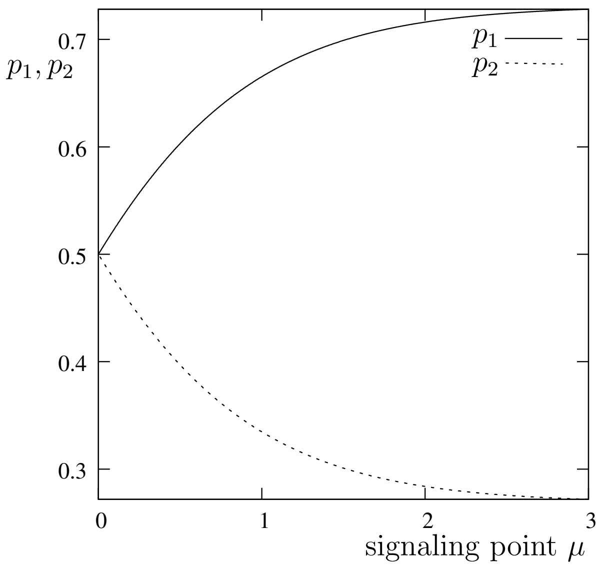

Fig. 1. Capacity-achieving probabilities p_1 and p_2 for signaling points $-\mu$ and μ in case of exponential noise with parameter $\lambda = 1$.

In the case that the signaling points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_M$ are cyclically generated by powers of some orthogonal matrix $\boldsymbol{T}$ the uniform distribution is capacity-achieving as is demonstrated in the following. Such collections of signaling points are sometimes referred to as *Slepian signal set* or *geometrically uniform constellation.* The identity matrix of size N is denoted by $\boldsymbol{I}_N$.

Proposition 3: Let $\boldsymbol{T}$ be orthogonal and such that $\boldsymbol{T}^M = \boldsymbol{I}_N$, and the signaling points be generated as

$$\boldsymbol{x}_i = \boldsymbol{T}^{i-1}\boldsymbol{x}_1, \quad i = 1, \dots, M.$$

Then the uniform distribution $\hat{\boldsymbol{p}} = (\frac{1}{M}, \dots, \frac{1}{M})$ is capacity-achieving.

Proof: We use Proposition 1 and show that for uniform $\hat{p}_i = \frac{1}{M}$ the integral $\int \varphi_i(\boldsymbol{y}) \log\Big(\sum_{j=1}^M \hat{p}_j \varphi_j(\boldsymbol{y})\Big) d\boldsymbol{y}$ is independent of index i. It holds that

$$\begin{aligned}
&\int \varphi(\boldsymbol{y} - \boldsymbol{x}_i) \log\Big(\frac{1}{M}\sum_{j=1}^M \varphi(\boldsymbol{y} - \boldsymbol{x}_j)\Big) d\boldsymbol{y} \\
&= \int \varphi(\boldsymbol{y} - \boldsymbol{T}^{i-1}\boldsymbol{x}_1) \log\Big(\frac{1}{M}\sum_{j=1}^M \varphi(\boldsymbol{y} - \boldsymbol{T}^{i-1}\boldsymbol{x}_j)\Big) d\boldsymbol{y} \\
&= \int \varphi(\boldsymbol{T}^{i-1}\boldsymbol{y} - \boldsymbol{T}^{i-1}\boldsymbol{x}_1) \\
&\qquad \cdot \log\Big(\frac{1}{M}\sum_{j=1}^M \varphi(\boldsymbol{T}^{i-1}\boldsymbol{y} - \boldsymbol{T}^{i-1}\boldsymbol{x}_j)\Big) d\boldsymbol{y} \\
&= \int \varphi(\boldsymbol{y} - \boldsymbol{x}_1) \log\Big(\frac{1}{M}\sum_{j=1}^M \varphi(\boldsymbol{y} - \boldsymbol{x}_j)\Big) d\boldsymbol{y},
\end{aligned}$$

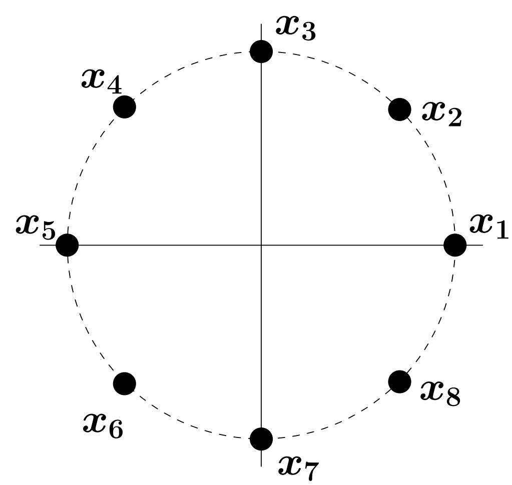

Fig. 2. Octal signal point constellation with uniform capacity-achieving distribution

which is independent of i. In the second line of the above we have used the argument that $\boldsymbol{x}_i = \boldsymbol{T}^{i-1}\boldsymbol{x}_1$, and furthermore that due to $\boldsymbol{T}^M = \boldsymbol{I}_N$ the sum is invariant against cyclic index shift. The third line follows from substituting $\boldsymbol{y}$ by $\boldsymbol{T}^{i-1}\boldsymbol{y}$ and the fact that $\boldsymbol{T}^{i-1}$ is orthogonal with Jacobian equal to 1. Finally, the fourth line ensues from the symmetry of φ. ■

As can be seen from the proof the symmetry assumption on φ can be weakened to invariance merely against the orthogonal matrices $\boldsymbol{T}^i$, $i = 0, \dots, M-1$, i.e., $\varphi(\boldsymbol{y}) = \varphi(\boldsymbol{T}^i\boldsymbol{y})$. In this sense, the symmetry axes of φ have to align the ones of the signaling points.

Example 2. The binary input scalar additive Gaussian channel is defined by

$$Y = X + n$$

with binary input variable X with support points $x_1 = \mu$, $x_2 = -\mu$. Independent noise variable n is normally distributed with zero mean, variance σ^2 and density $\varphi(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-y^2/2\sigma^2}$. Choosing $\boldsymbol{T} = (-1)$ Proposition 3 yields $\hat{p}_1 = \hat{p}_2 = \frac{1}{2}$ as the capacity-achieving input distribution. From Proposition 2 the capacity follows to be

$$\begin{aligned}
C &= D\big(\varphi(y-\mu) \,\|\, \varphi(y-\mu)/2 + \varphi(y+\mu)/2\big) \\
&= -\int \varphi(y-\mu) \log\Big(\frac{1}{2}\,\frac{\varphi(y-\mu) + \varphi(y+\mu)}{\varphi(y-\mu)}\Big) dy \\
&= \log 2 - \int \varphi(y) \log\Big(1 + \frac{\varphi(y+2\mu)}{\varphi(y)}\Big) dy \\
&= \log 2 - \mathrm{E}\big[\log(1 + e^{-W})\big],
\end{aligned}$$

where $W \sim N(2\mu^2/\sigma^2, 4\mu^2/\sigma^2)$ is normally distributed with the given parameters, cf. [11], p.188.

Example 3. Consider a two-dimensional additive Gaussian noise channel with noise covariance matrix $\sigma^2\boldsymbol{I}_2$. Signaling points are generated by the powers of the orthogonal matrix

$$\boldsymbol{T} = \begin{pmatrix} \cos(2\pi/M) & \sin(2\pi/M) \\ -\sin(2\pi/M) & \cos(2\pi/M) \end{pmatrix}.$$

Matrix $\boldsymbol{T}$ represents a counter-clockwise rotation by angle $\alpha = 2\pi/M$. Let $\boldsymbol{x}_1 \in \mathbb{R}^2$ be arbitrary and define the signaling points by

$$\boldsymbol{x}_i = \boldsymbol{T}^{i-1}\boldsymbol{x}_1, \quad i = 1, \ldots, M.$$

Then the assumptions of Proposition 3 are satisfied such that the uniform distribution is capacity-achieving.

Consider for example $M = 8$ and $\boldsymbol{x}_1 = (1, 0)^\mathsf{T}$. Then $\boldsymbol{T} = \begin{pmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix}$. This leads to the octal signal point constellation depicted in Figure 2. Capacity is achieved whenever each point is used with probability $1/8$.

V. Conclusions

Given an additive noise channel with arbitrary noise density and a finite set of signaling points, we have dealt with the question what the capacity-achieving input distribution would be. The optimum distribution has been characterized by possessing constant Kullback Leibler distances between the shifted noise densities and a mixture thereof with optimum probabilities as weights. As an example, we have determined the capacity-achieving distribution for binary equidistant input and exponential noise. In certain symmetric cases, the uniform distribution has been demonstrated to be capacity-achieving. Additive Gaussian vector channels are a special case of the general approach in this paper.

Acknowledgment

This work was partially supported by Deutsche Forschungsgemeinschaft grant Ma1184/14-1.

References

[1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, Nov. 1999.

[2] J. G. Smith, "The information capacity of amplitude and variance-constrained scalar Gaussian channels," *Inf. Contr.*, vol. 18, no. 6, pp. 203–219, 1971.

[3] T. H. Chan, S. Hranilovic, and F. R. Kschischang, "Capacity-achieving probability measure for conditionally Gaussian channels with bounded inputs," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 2073–2088, June 2005.

[4] T. H. Chan and F. R. Kschischang, "On the discreteness of the capacity-achieving probability measure of conditional Gaussian channels," in *Proceedings IEEE International Symposium on Information Theory, ISIT 2004*, Chicago, June 2004, p. 347.

[5] G. J. Foschini, R. D. Gitlin, and S. B. Weinstein, "Optimization of two-dimensional signal constellations in the presence of Gaussian noise," *IEEE Transactions on Communications*, vol. 22, no. 1, pp. 28–38, January 1974.

[6] T. Cover and J. Thomas, *Elements of Information Theory*. New York: John Wiley, 1991.

[7] A. Feiten and R. Mathar, "Water-filling is the limiting case of a general capacity maximization principle," in *Proc. IEEE International Symposium on Information Theory (ISIT'06)*, Seattle, Washington, July 2006, pp. 1282–1286.

[8] P. Whittle, "Some general points in the theory of optimal experimental design," *J. Roy. Statist. Soc. B*, vol. 35, pp. 123–130, 1973.

[9] H. Bauer, *Maß- und Integrationstheorie*. Berlin, New York: Walter de Gruyter, 1990.

[10] A. Tchamkerten, "On the discreteness of capacity-achieving distributions," *IEEE Transactions on Information Theory*, vol. 50, no. 11, pp. 2773–2778, Nov. 2004.

[11] D. J. C. MacKay, *Information Theory, Inference, and Learning Theory*. Cambridge: Cambridge University Press, 2003.

Capacity-Achieving Discrete Signaling over Additive Noise Channels

Anke Feiten
Institute of Theoretical Information Technology
RWTH Aachen University
Sommerfeldstr. 24, 52074 Aachen, Germany
Email: feiten@ti.rwth-aachen.de

Rudolf Mathar
Institute of Theoretical Information Technology
RWTH Aachen University
Sommerfeldstr. 24, 52074 Aachen, Germany
Email: mathar@ti.rwth-aachen.de

I. Correction

Due to a programming error, Figure 1 in the original paper is mistaken. It has to be replaced by the following graph.

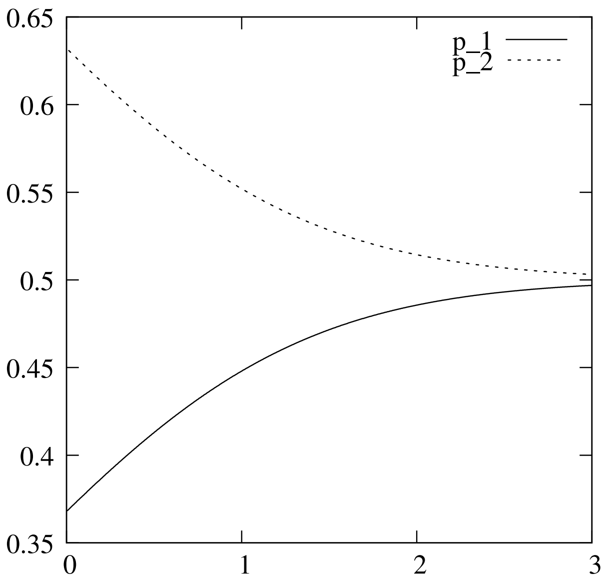

Fig. 1. Capacity-achieving probabilities p_1 and p_2 for signaling points $-\mu$ and μ in case of exponential noise with parameter $\lambda = 1$.