Optimum Discrete Signaling over Channels with Arbitrary Noise Distribution

Rudolf Mathar
Inst. for Theor. Information Technology
RWTH Aachen University
D-52056 Aachen, Germany
mathar@ti.rwth-aachen.de

Anke Schmeink
UMIC Research Center
RWTH Aachen University
D-52056 Aachen, Germany
schmeink@umic.rwth-aachen.de

Milan Zivkovic
Inst. for Theor. Information Technology
RWTH Aachen University
D-52056 Aachen, Germany
zivkovic@ti.rwth-aachen.de

Abstract— **General channels with arbitrary noise distributions and a finite set of signaling points are considered in this paper. We aim at finding the capacity-achieving input distribution. As a structural result we first demonstrate that mutual information is a concave function of the input distribution and a convex function of the channel transfer densities. Using the Karush-Kuhn-Tucker theory, capacity achieving distributions are then characterized by constant Kullback-Leibler divergence between each channel transfer density and the mixture hereof built by using the probabilities as weights. If, as a special case, the noise distribution and the signaling points are rotationally symmetric, then the uniform input distribution is optimal. For 2-PAM modulation and certain types of asymmetric noise distributions, including exponential, gamma and Rayleigh, we present extensive numerical evaluations of the optimal input. Furthermore, for 4-QAM we determine the optimal input from a restricted symmetric class of distributions for correlated Gaussian noise.**

I. Introduction and Motivation

It is a classical result of Shannon that the scalar additive Gaussian noise channel subject to average power constraints achieves capacity if the input distribution is Gaussian as well. The survey paper [1] summarizes the extension of this result to complex circularly symmetric Gaussian vector channels, particularly MIMO channels. However, due to its unbounded support this distribution is not realizable in practice. Further research has focused on bounded signaling sets by introducing peak power constraints of different types. Interestingly, the capacity-achieving distribution becomes discrete with finite support in this case, as was shown in [2]. The work [3] gives an overview of previous research on the topic, including Poisson, quadrature Gaussian and additive vector Gaussian noise distributions under average and peak power constraints. This paper also generalizes the problem by considering conditionally Gaussian vector channels subject to bounded-input constraints by some bounded set $\mathcal{S}$. Under certain conditions on $\mathcal{S}$ capacity is achieved for a discrete distribution with finitely many signaling points. In the recent paper [4], a Rayleigh fading channel and average power constraints are considered. It is shown that the support of the capacity-achieving distribution is bounded. However, it is also pointed out that proofs of the important communication-theoretic problem of finite or bounded support by using the identity theorem for holomorphic functions are not rigorous and need further consideration.

A related question of optimal signaling is raised in [5]: what is the optimum constellation of M equiprobably used signaling points for an additive Gaussian channel with average power constraints such that the error probability attains its minimum.

Summarizing the above, for practical purposes it is sufficient to determine a finite signaling constellation of M points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_M$ and the corresponding capacity-achieving input distribution $\boldsymbol{p} = (p_1, \dots, p_M)$. In this paper, we confine ourselves to a fixed given constellation set and merely search for the optimum input distribution. This approach is motivated by practical requirements on the simplicity of the receiver structure. Standard MPAM and MPSK modulation schemes are covered by our model.

Results evolve along the following lines. We first provide some structural results on mutual information, the objective function used in this paper. Concavity as a function of the input distribution, and convexity as a function of arbitrary channel transfer densities is demonstrated in Section II. The optimal input distribution is then characterized by use of the Karush-Kuhn-Tucker (KKT) conditions, see Section III. Finally, in Sections IV and V, this characterization is used to obtain explicit and numerical results for symmetric and arbitrary noise densities for 2-PAM, respectively. Section VI concludes with numerical evaluations of 4-QAM with correlated noise.

II. Structural Results

Consider a channel with M signaling points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_M \in \mathbb{R}^n$ which are used by the transmitter according to a certain input distribution $\boldsymbol{p} = (p_1, \dots, p_M) \in \mathcal{D}^M$, where the set of all probability distributions with M support points is denoted by

$$\mathcal{D}^M = \{\boldsymbol{p} = (p_1, \dots, p_M) \mid p_i \geq 0, \textstyle\sum_{i=1}^M p_i = 1\}.$$

Let random variable $\boldsymbol{X}$ denote the discrete channel input with distribution $\boldsymbol{p}$. The channel output $\boldsymbol{Y}$ is randomly distorted by noise. Throughout the paper we assume that the distribution of $\boldsymbol{Y}$ given input $\boldsymbol{X} = \boldsymbol{x}_i$ has (Lebesgue) density

$$f(\boldsymbol{y} \mid \boldsymbol{x}_i) = f_i(\boldsymbol{y}), \ \boldsymbol{y} \in \mathbb{R}^n.$$

The AWGN channel $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{n}$ is a special case hereof with $f_i(\boldsymbol{y}) = \varphi(\boldsymbol{y} - \boldsymbol{x}_i)$. Here, φ denotes the density of a Gaussian distribution $N_n(\boldsymbol{0}, \boldsymbol{\Sigma})$.

Mutual information between channel input and output as a function of $\boldsymbol{p} = (p_1, \ldots, p_M)$ and $f_1, \ldots, f_M$ may be written as

$$\begin{aligned} I(\boldsymbol{X};\boldsymbol{Y}) &= I\big(\boldsymbol{p};(f_1,\ldots,f_M)\big) \\ &= H(\boldsymbol{Y}) - H(\boldsymbol{Y} \mid \boldsymbol{X}) \\ &= H\Big(\sum_{i=1}^{M} p_i f_i\Big) - \sum_{i=1}^{M} p_i H(f_i) \\ &= \sum_{i=1}^{M} p_i D\Big(f_i \Big\| \sum_{j=1}^{M} p_j f_j\Big), \end{aligned} \tag{1}$$

where $D(f\|g) = \int f \log \frac{f}{g}$ denotes the Kullback-Leibler divergence between densities f and g.

Let $\mathcal{F}$ denote the set of all Lebesgue densities $f : \mathbb{R}^n \to \mathbb{R}^+$. From the convexity of $t \log t$, $t \geq 0$, it is easily concluded that

$$H\Big(\sum_{i=1}^{M} p_i f_i\Big) \text{ is a concave function of } \boldsymbol{p} \in \mathcal{D}^M. \tag{2}$$

By applying the log-sum inequality (cf. [6]) we also obtain

$$\begin{aligned} &\alpha f_1 \log \frac{f_1}{g_1} + (1-\alpha) f_2 \log \frac{f_2}{g_2} \\ &\quad \geq \big(\alpha f_1 + (1-\alpha) f_2\big) \log \frac{\alpha f_1 + (1-\alpha) f_2}{\alpha g_1 + (1-\alpha) g_2}, \end{aligned}$$

pointwise for any pairs of densities (f_1, g_1), $(f_2, g_2) \in \mathcal{F}^2$. Integrating both sides of the above inequality shows that

$$D(f\|g) \text{ is a convex function of the pair } (f,g) \in \mathcal{F}^2. \tag{3}$$

Applying (2) and (3) to the third and forth line of representation (1), respectively, gives the following.

Proposition 1: Mutual information $I\big(\boldsymbol{p};(f_1,\ldots,f_M)\big)$ is a concave function of $\boldsymbol{p} \in \mathcal{D}^M$ and a convex function of $(f_1, \ldots, f_M) \in \mathcal{F}^M$.

Hence, determining the capacity of the channel for fixed channel transfer densities $f_1, \ldots, f_M$ leads to a concave optimization problem, namely

$$C = \max_{\boldsymbol{p} \in \mathcal{D}^M} I(\boldsymbol{p}; f_1, \ldots, f_M).$$

This problem will be considered in detail in the following section.

III. Capacity-Achieving Input

The capacity-achieving input distribution is the solution of a convex optimization problem. Hence, the solution may be characterized by the Karush-Kuhn-Tucker (KKT) theory.

Distribution $\boldsymbol{p}^* = (p_1^*, \ldots, p_M^*)$ achieves capacity, i.e., maximizes mutual information, if and only if $I(\boldsymbol{X};\boldsymbol{Y})$ is maximized by $\boldsymbol{p}^*$ in the set of all stochastic vectors. By representation (1), we need to solve

$$\begin{aligned} \text{maximize}\Big\{ -\int \Big(\sum_{i=1}^{M} p_i\, f_i(\boldsymbol{y})\Big) \log \Big(\sum_{i=1}^{M} p_i\, f_i(\boldsymbol{y})\Big) d\boldsymbol{y} \\ + \sum_{i=1}^{M} p_i \int f_i(\boldsymbol{y}) \log f_i(\boldsymbol{y}) d\boldsymbol{y} \Big\} \end{aligned}$$

subject to

$$\begin{aligned} &p_i \geq 0, \ i = 1, \cdots, M, \\ &\sum_{i=1}^{M} p_i = 1. \end{aligned}$$

The above is a convex problem since by Proposition 1 the objective function $g(p_1, \ldots, p_M)$ is concave and the constraint set is convex. The Lagrangian is given by

$$\begin{aligned} L(\boldsymbol{p}, \boldsymbol{\lambda}, \nu) = -\int \Big(\sum_{i=1}^{M} p_i\, f_i(\boldsymbol{y})\Big) \log \Big(\sum_{i=1}^{M} p_i\, f_i(\boldsymbol{y})\Big) d\boldsymbol{y} \\ - \sum_{i=1}^{M} p_i \int f_i(\boldsymbol{y}) \log f_i(\boldsymbol{y}) d\boldsymbol{y} \\ + \sum_{i=1}^{M} \lambda_i p_i + \nu\big(\sum_{i=1}^{M} p_i - 1\big). \end{aligned}$$

with the notation $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_M)$. The optimality conditions are (cf. [7, Ch. 5.5.3])

$$\begin{aligned} \frac{\partial L(\boldsymbol{p}, \boldsymbol{\lambda}, \nu)}{\partial p_i} &= 0, \\ p_i, \lambda_i &\geq 0, \\ \lambda_i p_i &= 0, \end{aligned} \tag{4}$$

for all $i = 1, \ldots, M$. Partial derivatives of the Lagrangian w.r.t. p_i are easily obtained as

$$\begin{aligned} \frac{\partial L(\boldsymbol{p}, \boldsymbol{\lambda}, \nu)}{\partial p_i} = -(\log e) - \int f_i(\boldsymbol{y}) \log \Big(\sum_{j=1}^{M} p_j f_j(\boldsymbol{y})\Big) d\boldsymbol{y} \\ + \int f_i(\boldsymbol{y}) \log f_i(\boldsymbol{y}) d\boldsymbol{y} + \lambda_i + \nu, \end{aligned}$$

for $i = 1, \ldots, M$. Hence, (4) leads to the conditions $p_i = 0$ or

$$\int f_i(\boldsymbol{y}) \Big(\log f_i(\boldsymbol{y}) - \log\big(\sum_{j=1}^{M} p_j f_j(\boldsymbol{y})\big)\Big) d\boldsymbol{y} = \log e - \nu.$$

for all $i = 1, \ldots, M$. In summary, we have demonstrated the following result.

Proposition 2: Input distribution $\boldsymbol{p}^*$ is capacity-achieving if and only if

$$D\Big(f_i \Big\| \sum_{j=1}^{M} p_j^* f_j\Big) = \text{const} \tag{5}$$

for all i such that $p_i > 0$. Furthermore, if $H(f_i)$ is independent of i, then $\boldsymbol{p}^*$ is capacity-achieving iff

$$\int f_i(\boldsymbol{y}) \log \Big(\sum_{j=1}^{M} p_j^* f_j(\boldsymbol{y})\Big) d\boldsymbol{y} = \text{const} \tag{6}$$

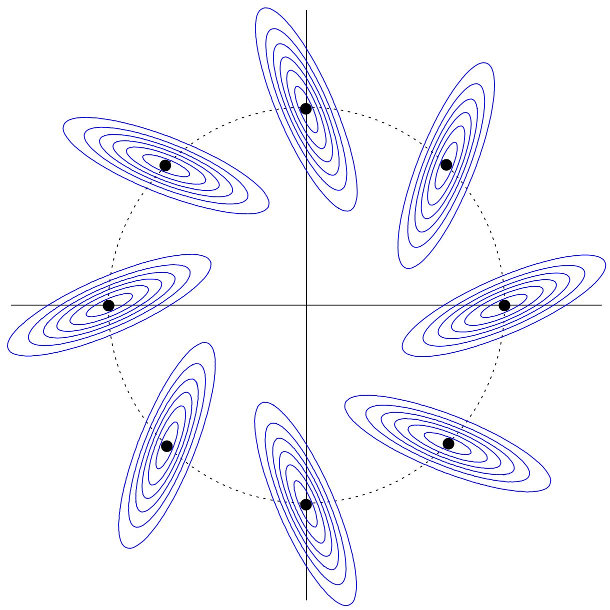

Fig. 1. 8-PSK with contour lines of rotated Gaussian noise distributions, initially with expectation $(1,0)'$, variances $\sigma_1^2 = 2$ and $\sigma_2^2 = 1$, and correlation $\rho = 0.8$.

for all i such that $p_i > 0$.

IV. Symmetric noise distributions

Assume that the channel noise distribution when transmitting symbol $\boldsymbol{x}_i$ is symmetric in the sense that

$$f_i(\boldsymbol{y}) = f_0(\boldsymbol{T}^i \boldsymbol{y}),\ i = 1,\dots,M,$$

for some fixed density f_0 and orthogonal matrix $\boldsymbol{T}$ satisfying

$$\boldsymbol{T}^M = \boldsymbol{I}, \tag{7}$$

the identity matrix. In this case, the uniform distribution $\boldsymbol{p} = (\frac{1}{M},\dots,\frac{1}{M})$ is capacity achieving, as may be seen by the following. For any $i \in \{1,\dots,M\}$ it holds that

$$\begin{aligned} &\int f_i(\boldsymbol{y}) \log\Big(\frac{1}{M}\sum_{j=1}^M f_j(\boldsymbol{y})\Big) d\boldsymbol{y} \\ &= \int f_0(\boldsymbol{T}^i\boldsymbol{y}) \log\Big(\frac{1}{M}\sum_{j=1}^M f_0(\boldsymbol{T}^j\boldsymbol{y})\Big) d\boldsymbol{y} \\ &= \int f_0(\boldsymbol{y}) \log\Big(\frac{1}{M}\sum_{j=1}^M f_0(\boldsymbol{T}^{j-i}\boldsymbol{y})\Big) d\boldsymbol{y} \\ &= \int f_0(\boldsymbol{y}) \log\Big(\frac{1}{M}\sum_{j=1}^M f_0(\boldsymbol{T}^j\boldsymbol{y})\Big) d\boldsymbol{y}, \end{aligned}$$

independent of i. By condition (6) this proves optimality of the uniform distribution.

As an example, let $f_0 \sim N(\boldsymbol{x}_0, \boldsymbol{\Sigma}_0)$ be the density of the n-dimensional Gaussian distribution with expectation $\boldsymbol{x}_0$ and covariance matrix $\boldsymbol{\Sigma}_0$. If some orthogonal matrix $\boldsymbol{T}$ satisfies (7), then f_i may be chosen as

$$f_i(\boldsymbol{y}) \sim N\big(\boldsymbol{T}^i\boldsymbol{x}_0, \boldsymbol{T}^i\boldsymbol{\Sigma}\boldsymbol{T}^{i'}\big),\ i = 1,\dots,M.$$

Obviously, circularly symmetric additive noise, i.e.,

$$f_i(\boldsymbol{y}) = f_0(\boldsymbol{y} - \boldsymbol{x}_i),\ i = 1,\dots,M,$$

with $f_0 \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$, and signaling points generated as $\boldsymbol{x}_i = \boldsymbol{T}^{i'}\boldsymbol{x}_0$, $i = 1,\dots,m$, is an example hereof. Complex MPSK signaling with circularly symmetric noise distribution is a special case, see [8].

The present result shows that correlated noise is also admitted, still having the uniform distribution as capacity-achieving. However, the noise distribution has to be accordingly rotated for each signaling point $\boldsymbol{x}_i = \boldsymbol{T}^{i'}\boldsymbol{x}_0$.

8-PSK with the signaling constellation depicted in Fig. 1 is an example hereof. $\boldsymbol{x}_0 = \boldsymbol{x}_8 = (1,0)'$ is the initial point and $\boldsymbol{T} = \begin{pmatrix} \cos\pi/4 & \sin\pi/4 \\ -\sin\pi/4 & \cos\pi/4 \end{pmatrix}$ the orthogonal rotation with $\boldsymbol{T}^8 = \boldsymbol{I}$. Fig. 1 also shows the contour lines of the Gaussian noise distribution with covariance matrix $\begin{pmatrix} 2 & 1.6 \\ 1.6 & 1 \end{pmatrix}$.

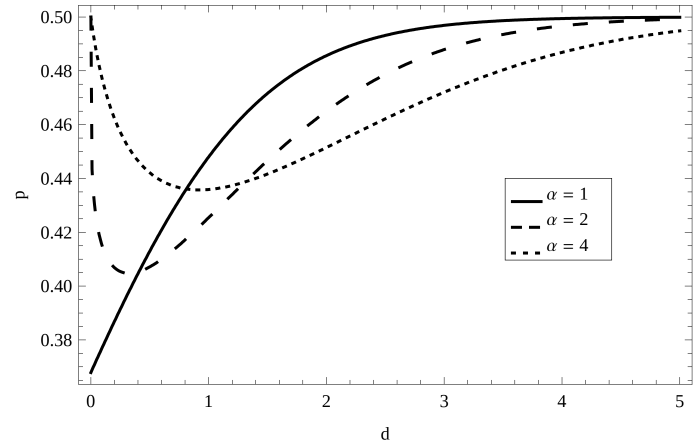

Fig. 2. Optimum probability for selecting signaling point $-d$ with 2-PAM. The noise distribution is $\Gamma(\alpha,\lambda)$ with parameters $\lambda = 1$ and $\alpha \in \{1,2,4\}$.

V. 2-PAM with Skew Noise Distributions

Pulse amplitude modulation with two signaling points (2-PAM) $x_1 = -d$ and $x_2 = d$, $d > 0$, is considered in this section. If the noise distribution is symmetric, then from Section IV the uniform distribution $p_1 = p_2 = 1/2$ follows to be capacity-achieving. Simply choose $M = 2$ and $\boldsymbol{T} = (-1) \in \mathbb{R}$.

We now assume additive noise in the form that

$$f_1(y) = f(y-d) \quad \text{and} \quad f_2(y) = f(y+d)$$

for some fixed noise density f. Optimality condition (6) for $p_1 = p$ and $p_2 = (1-p)$, $0 \le p \le 1$ then leads to the problem of finding p such that

$$\int f_1 \log\big(pf_1 + (1-p)f_2\big) = \int f_2 \log\big(pf_1 + (1-p)f_2\big) \tag{8}$$

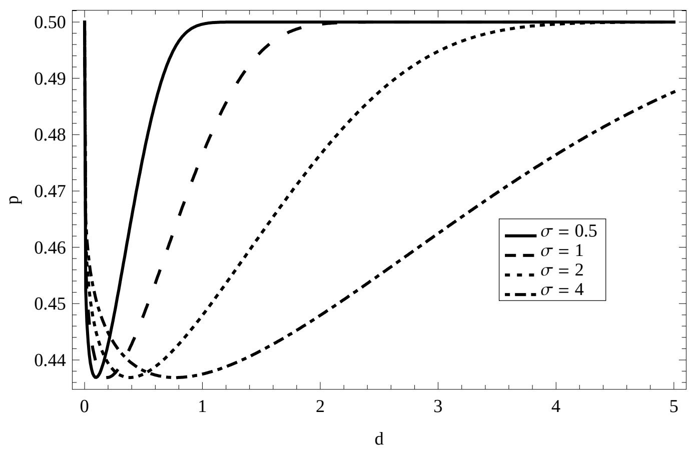

Fig. 3. Optimum probability for selecting signaling point $-d$ with 2-PAM. The noise distribution is Rayleigh with parameter $\sigma \in \{0.5, 1, 2, 4\}$.

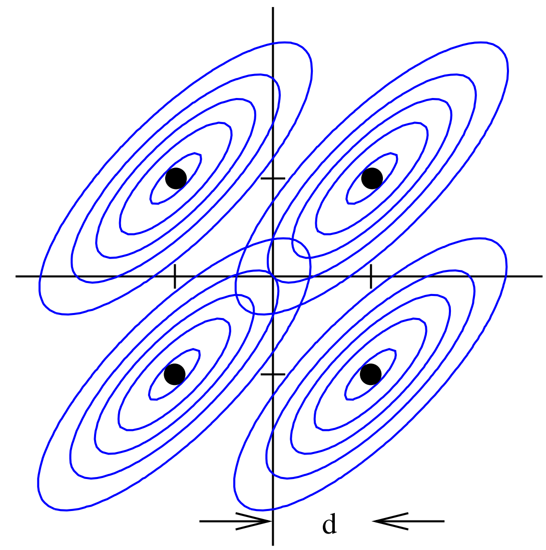

Fig. 4. 4-QAM, signaling points (black circles) and contour lines of the two-dimensional Gaussian noise distribution with unit variances and correlation $\rho = 0.8$.

holds. If $f(y) = \lambda e^{-\lambda y}$, $y \geq 0$, the density of the exponential distribution, the optimal input distribution (p_1, p_2) is explicitly determined in [8] through

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

as

$$p_1 = \frac{1}{1+\alpha} \quad \text{and} \quad p_2 = \frac{1}{1+\alpha}.$$

The probability $p_1(d)$ of using signaling point $-d$ is depicted as a function of d in Fig. 2 (solid line). It can be seen that the further the signaling points are apart the closer the optimum distribution approaches the uniform, i.e., $p_1 = 0.5$. For signaling points nearby there is a significant deviation from the uniform distribution in that the $x_2 = +d$ is selected with higher probability, in the limit $\lim_{d \to 0} p_1(d) = 1/e$, as can be easily seen.

Explicit results for other noise distributions seem to be hard to achieve. Numerically, however, equation (8) can be solved by using according Matlab subroutines. Two classes of distributions were investigated that way, gamma and Rayleigh distributions. The according densities are

$$f_{\Gamma}(y) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)} y^{\alpha-1} e^{-\lambda y}, \quad y > 0 \quad (\alpha, \lambda > 0),$$

$$f_{\text{Ray}}(y) = \frac{y}{\sigma^2} e^{-\frac{y^2}{2\sigma^2}}, \quad y > 0 \quad (\sigma^2 > 0).$$

The values $p_1(d)$ for parameters $\lambda = 1$ and $\alpha = 1, 2, 4$ are shown in Fig. 2. Parameter $\alpha = 1$ represents the above mentioned exponential distribution. Interestingly, the behavior is non-monotonic for the cases $\alpha = 2$ and $\alpha = 4$.

A non-monotonic behavior of $p_1(d)$ is also observed in the case of Rayleigh distributions, as is depicted in Fig. 3. Again, with signaling points far apart, the uniform distribution is optimum, while signaling points close by show a remarkable deviation from uniformity.

VI. 4-QAM with Correlated Gaussian Noise

Four signaling points $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4$ are given as $(d, d)', (-d, d)', (-d, -d)', (d, -d)'$, $d > 0$, see Fig. 4. Contour lines of the two-dimensional Gaussian noise distribution with unit variances and correlation $\rho = 0.8$ are also shown in this graph.

To allow for explicit numerical solutions we only consider the class $\big(p/2, (1-p)/2, p/2, (1-p)/2\big)$, $0 \leq p \leq 1$, of diagonally symmetric input distributions. The problem is hence reduced to one dimension. Channel densities $f_i(\boldsymbol{y})$ are assumed to correspond to Gaussian $\mathrm{N}(\boldsymbol{x}_i, \Sigma)$-distributions with some fixed covariance matrix $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. In this case, optimality condition (6) reads as

$$\int \big((f_1+f_3)-(f_2+f_4)\big) \ln\Big(\frac{p}{2}(f_1+f_3)-\frac{1-p}{2}(f_2+f_4)\Big) = 0$$

for $0 < p < 1$. It is satisfied whenever

$$\int (f_i - f_j) \ln\Big(\frac{p}{2}(f_1 + f_3) - \frac{1-p}{2}(f_2 + f_4)\Big) = 0$$

holds for $(i, j) \in \{(1, 2), (3, 4)\}$ and the same $p \in (0, 1)$.

The last equation is numerically evaluated and the according values of p are represented as a function of d in Fig. 5. The case $\rho = 0$ corresponds to stochastically independent symmetric noise which yields a uniform distribution independent of d, as demonstrated in Section IV. As d increases, the uniform distribution is asymptotically optimum for any correlation ρ. Interestingly, for small values of d the deviation from uniformity reverts if ρ passes approximately the value $|\rho| \approx 0.5$.

This phenomenon is closer investigated in Fig. 6, where the behavior of the input distribution is depicted as a function of ρ for fixed $d \in \{0.7, 1, 1.5, 3\}$ and unit variances of the noise

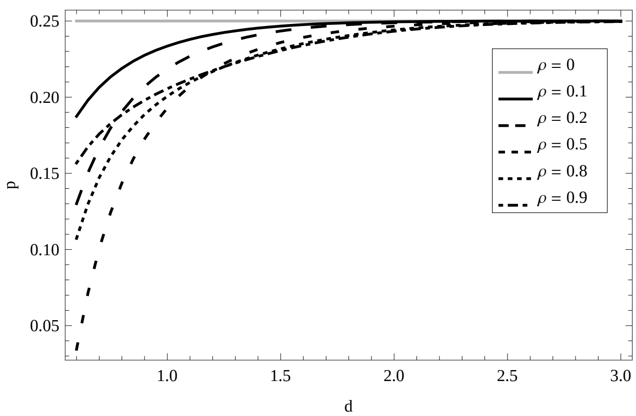

Fig. 5. The probability p of selecting signaling point $\boldsymbol{x}_1$ as a function of d for two-dimensional Gaussian noise with unit variances and correlation $\rho \in \{0, 0.1, 0.2, 0.5, 0.8, 0.9\}$. Uncorrelatedness ($\rho = 0$) yields a uniform distribution independent of d.

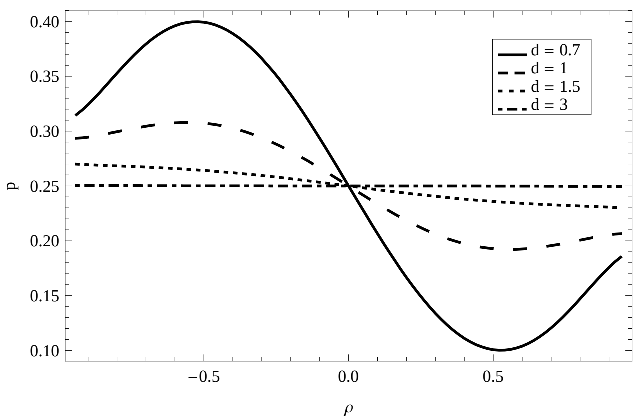

Fig. 6. The probability p of selecting signaling point $\boldsymbol{x}_1$ as a function of ρ for different values of $d = 0.7, 1.0, 1.5, 3.0$.

distribution. Deviation from uniformity attains its maximum and minimum, respectively, at $|\rho| = 0.526$ for $d = 0.7$ and $|\rho| = 0.554$ for $d = 1$. Monotone behavior is observed for $d \in \{1.5, 3.0\}$. The value $d = 3$ corresponds to the case that the signaling points are far apart. The nearly constant value $p = 0.25$ points to the fact that despite correlations in the noise the uniform distribution is capacity-achieving. Significant deviations from the uniform distribution may be observed if the signaling points are close-by, e.g., the case $d = 0.7$. The behavior of the optimum distribution is extremely interesting. For negative correlations the signaling points $\boldsymbol{x}_1$ and $\boldsymbol{x}_3$ (first and third quadrant) are selected with higher probability, positive correlations induce more emphasis on $\boldsymbol{x}_2$ and $\boldsymbol{x}_4$.

VII. Conclusions and Outlook

We have investigated the optimal input distribution for finite given sets of signaling points when the channel is subject to arbitrary noise distributions. Mutual information serves as the objective function. Convexity of the basic problem has been shown, and based on this, a general optimality condition in terms of the Kullback-Leibler divergence has been given. It has been demonstrated that for asymmetric noise distributions the optimal input is significantly different from being uniform, which should be exploited by the bitmapper in real systems. Important issues for subsequent research are numeric stability of algorithms to determine the optimum and, furthermore, including the position of the signaling points into the objective function.

Acknowledgement. This work was partially supported by the UMIC Research Center at RWTH Aachen University.

References

[1] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *European Transactions on Telecommuncations*, vol. 10, no. 6, pp. 585–595, 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] J. Sommerfeld, I. Bjelakovic, and H. Boche, "On the boundedness of the support of optimal input measures for Rayleigh fading channels," in *Proceedings ISIT'08*, Toronto, July 2008.

[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] I. Csiszar and J. Körner, *Information Theory, Coding theorems for discrete memoryless sources*. London: Academic Press, 1981.

[7] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York: Cambridge University Press, 2004.

[8] A. Feiten and R. Mathar, "Capacity-achieving discrete signaling over additive noise channels," in *Proceedings ICC 2007*, Glasgow, June 2007.