Optimum One-Bit Quantization

Gholamreza Alirezaei and Rudolf Mathar
Institute for Theoretical Information Technology
RWTH Aachen University, D-52056 Aachen, Germany
{alirezaei, mathar}@ti.rwth-aachen.de

***Abstract*—This paper deals with discrete input one-bit output quantization. A discrete input signal is subject to additive noise and is then quantized to zero or one by comparison with a threshold q. For finitely many fixed support points and fixed threshold q we first determine the mutual information of this channel. The capacity-achieving input distribution is shown to be concentrated on merely two extreme support points. Furthermore, an elegant representations of the corresponding probabilities is found. Finally, we set out to determine the optimum threshold q, which is an extremely hard problem. By means of graphical representations a completely different behavior of the objective function is revealed, depending on the choice of parameters and the noise distribution.**

I. Introduction

One-bit quantization of a real noisy signal is of high interest from an information theoretic point of view. Because of the non-linearity the problem of determining the optimum threshold $q^\star$ is rather cumbersome. In this paper, the optimal value is characterized by the solution of an optimization problem, which involves the noise distribution and the input signaling points only. Moreover, we see certain applications in modeling information processing in biological neural systems, see [1].

But also from an engineering point of view one-bit quantization is an important element of modern digital receiver design. Instead of using high precision analog-to-digital converters (ADC) parallel one-bit quantizers could be employed. This would allow for designing high-speed systems with much lower energy consumption. For this purpose, the fundamental limits of low precision ADC must be understood.

This paper aims at contributing to this objective. The questions we ask are quite basic. Firstly, what is the capacity-achieving distribution for an additive noise channel with a one-bit output quantizer in the class of discrete input distributions with m support or signaling points? Secondly, what is the optimum threshold $q^\star$ so that the capacity of this channel is maximized?

Investigating quantizers of an input signal subject to additive white Gaussian noise (AWGN) is the most popular case, as the normal distribution has salient mathematical properties. Considering discrete channel input is quite natural from a practical point of view, since digital transmission systems usually use finite sets of signaling points. Hence, in many publications a discrete channel input in combination with AWGN is considered. For example, a related problem is treated for the AWGN channel in [2], namely maximizing mutual information subject to a power constraint. Similarly in [3], it is stated that the capacity-achieving distribution over all input distributions is discrete for AWGN noise, once an average power constraint is applied. In the work [4], the real discrete-time AWGN channel with an average power constraint is considered. The authors show that for a K-bit quantizer with a precision of $\log_2(K)$ bits the capacity-achieving input distribution is discrete with at most $K + 1$ mass points. For binary symmetric quantization this result is refined to demonstrate that antipodal signaling is optimum for arbitrary signal-to-noise ratios. The authors conjecture that symmetric quantizers are optimal, however, are not able to provide a proof. Moreover, the loss by low precision ADCs is numerically quantified.

One-bit quantization is also considered in [5] for the AWGN channel model. In the low signal-to-noise regime, as is relevant for spread-spectrum and ultra-wideband communications, it is shown that asymmetric signal constellations combined with asymmetric quantization are superior to the fully symmetric case. It is shown that with such asymmetric threshold quantizers the capacity per unit-energy of the Gaussian channel without output quantization is achieved. However, for this purpose flash-signaling input distributions are required, which are not within the class of peak power constrained input distributions as they will be considered in the present work.

Moreover, by the paper [5], the work [3] has been extended to include a peak power constraint for the Gaussian channel. This leads to the result that the capacity-achieving input distribution is concentrated on two extreme mass points, cf. [5, Prop. 1]. Capacity is written as a maximization problem over all possible thresholds, but only numerical indications are given for the optimum threshold. Moreover, for the Gaussian channel it is shown in [5] that a threshold quantizer is optimal.

Arbitrary noise distributions are considered in [6]. It is shown that the capacity-achieving distribution for the one-bit quantization channel is discrete whenever the support of the input distribution is finite.

In the work [7] one-bit quantization is interpreted as an asymmetric channel. Channel capacity and minimal error probability are investigated in parallel, and optimal threshold settings are determined numerically. In [8] the closely related problem of optimal one-bit source quantization is studied. It is shown that for symmetric and log-concave source distributions the optimal one-bit quantizer is symmetric about the origin. In the recent paper [9], for a complex-valued fading channel

This work was partially supported by the DFG grant Ma 1184/20-2 in the priority program InKomBio.

978-1-4673-7852-9/15/$31.00 ©2015 IEEE

the ergodic capacity and outage probability of one-bit output quantization for discrete I/Q modulation schemes have been determined.

In this paper, we consider a real input channel with a finite number of signaling points, arbitrary additive noise and one-bit quantization by threshold q, see Fig. 1. The same model is used in [10] where for fixed signaling points the capacity-achieving distribution is determined. What discriminates this work from others, are the following contributions.

We give an elegant direct proof of the fact that the capacity-achieving input concentrates on the extreme signaling points only. To this end, mutual information of the channel is represented explicitly as a function of the input distribution and the threshold q. Finally, we investigate some structural properties of the capacity as a function of threshold q. Determining the capacity-maximizing threshold q is an extremely hard problem. We arrive at an interesting symmetric representation of the objective function. We also demonstrate by graphical representations that completely different behavior of the objective function can be observed. In selected cases this entails well founded conjectures on the optimum threshold $q^\star$.

II. Channel Model

We assume an additive noise channel (not necessary Gaussian). Real input X with cumulative distribution function (CDF) $F(x)$ is subject to additive noise W with density function $\varphi(w)$ and corresponding CDF $\Phi(w)$. X and W are assumed to be stochastically independent. The noisy signal $X + W$ is then quantized by a binary quantizer Q with threshold q as $Q(s) = 1$, if $s > q$ and $Q(s) = 0$, otherwise. The system model is depicted in Fig. 1 and reads as

$$Y = Q(X + W). \tag{1}$$

Let $h(p)$ denote the binary entropy function, defined by

$$h(p) = -p \log_a p - (1-p)\log_a(1-p), \quad 0 \le p \le 1\,, \tag{2}$$

where $a > 1$ denotes the base of the logarithm. It is well known that $h(p)$ is a strictly concave function of $p \in [0, 1]$.

Including noise as input and maximizing mutual information $I(Z;Y)$ is an easy task. Let $\tilde{F}(z)$ denote the CDF of Z. Since $H(Y \mid Z) = 0$ we obtain

$$I(Z;Y) = H(Y) - H(Y \mid Z) = h\big(\tilde{F}(q)\big) \le \log_a 2$$

with equality if and only if $\tilde{F}(q) = 1/2$. Hence, the channel capacity of Z to Y is $\log_a 2$. It is achieved for any distribution of Z with $\tilde{F}(q) = 1/2$. In contrast, maximizing $I(X;Y)$ is much harder.

Mutual information between input X and binary output Y may be written as

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y \mid X) \qquad (3)\\ &= h\Big(\int \Phi(q-x)dF(x)\Big) - \int h\big(\Phi(q-x)\big)dF(x). \end{aligned}$$

Mutual information is hence a function of the input distribution F, the noise distribution Φ, and the quantization threshold q. This motivates the notation $I(X;Y) = I(F, \Phi, q)$. In the case that Φ is continuous (not necessary differentiable) and F corresponds to a discrete distribution with finitely many mass points m, described by the density $\boldsymbol{p} = (p_1, \dots, p_m)$ and support points $\boldsymbol{x} = (x_1, \dots, x_m)$, we also write $I(\boldsymbol{p}, \boldsymbol{x}, q)$ for (3), i.e.,

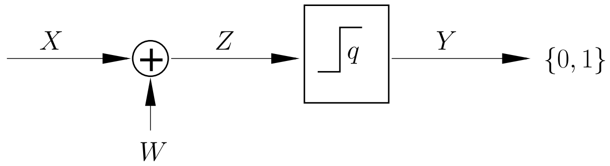


Fig. 1: The system model: some real input X is subject to additive noise W and is quantized with threshold q to yield binary output Y.

$$I(\boldsymbol{p}, \boldsymbol{x}, q) = h\Big(\sum_{i=1}^{m} p_i \Phi(q - x_i)\Big) - \sum_{i=1}^{m} p_i h\big(\Phi(q - x_i)\big). \tag{4}$$

By $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_m)$ with $\gamma_i = \Phi(q - x_i)$, $1 \le i \le m$, a concise representation of (4) is obtained as

$$I(\boldsymbol{p}, \boldsymbol{\gamma}) = h\Big(\sum_{i=1}^{m} p_i \gamma_i\Big) - \sum_{i=1}^{m} p_i h(\gamma_i). \tag{5}$$

It is well known that $I(F, \Phi, q)$ is a concave function of F and a convex function of Φ. Hence, $I(\boldsymbol{p}, \boldsymbol{\gamma})$ is concave in $\boldsymbol{p}$ and convex in $\boldsymbol{\gamma}$ as well. Obviously, γ_i is monotonically increasing and decreasing in its arguments q and x_i, respectively.

III. Capacity-achieving Input Distribution

From now on we will only consider discrete input distributions with finitely many support points $x_1, \dots, x_m$. From (5) it follows that degenerate distributions cannot be capacity-achieving, since in the case that $p_i = 1$ for some $i \in \{1, \dots, m\}$ the identity $I(\boldsymbol{p}, \boldsymbol{\gamma}) = h(\gamma_i) - h(\gamma_i) = 0$ holds, which cannot be the maximum of $I(\boldsymbol{p}, \boldsymbol{\gamma})$.

We first observe that the maximum of $I(\boldsymbol{p}, \boldsymbol{x}, q)$ over threshold q with $\boldsymbol{p}$ fixed does not change if each support point is shifted by the same amount.

Proposition 1: The maximum $\max_q I(\boldsymbol{p}, \boldsymbol{x}, q)$ is invariant to a constant additive shift α of the support points for any fixed input distribution $\boldsymbol{p}$.

Proof. For any $\alpha \in \mathbb{R}$ it follows from (4) that $I(\boldsymbol{p}, \boldsymbol{x}, q) = I(\boldsymbol{p}, \boldsymbol{x} + \alpha \mathbf{1}_m, q + \alpha)$, where $\mathbf{1}_m$ denotes the all-ones vector of size m. Hence, the maximum with shifted support points is attained at $q^\star + \alpha$, if $q^\star$ is a solution of the original problem. ∎

For any $\boldsymbol{x}$ and q, the capacity achieving distribution is concentrated on the extreme support points, as is shown in the following, see also [11, pp. 91–96].

Proposition 2: Consider the set $\mathcal{D}$ of discrete distributions with support points $x_1 < \cdots < x_m$ and corresponding probabilities $\boldsymbol{p} = (p_1, \dots, p_m)$. Then the maximum $\max_{\boldsymbol{p}} I(\boldsymbol{p}, \boldsymbol{x}, q)$ is attained at a distribution concentrated on the extreme points x_1 and x_m. Hence, the capacity-achieving distribution over $\mathcal{D}$ has at most two signaling points x_1 and x_m.

Proof. Suppose $x_1 < \cdots < x_m$ and consider the first derivative of $I(\boldsymbol{p},\boldsymbol{\gamma})$ w.r.t. γ_j for some $j \in \{1,\dots,m\}$, i.e.,

$$\frac{\partial I(\boldsymbol{p},\boldsymbol{\gamma})}{\partial \gamma_j} = p_j \log_a\left(1 + \frac{(\gamma_j - \bar{\gamma}_j)(1-p_j)}{(1-\gamma_j)\sum_{i=1}^m p_i\gamma_i}\right),$$

where $\bar{\gamma}_j = \sum_{\substack{i=1\\i\neq j}}^m p_i\gamma_i / \sum_{\substack{i=1\\i\neq j}}^m p_i$.

For any $\gamma_j \geq \bar{\gamma}_j$ the derivative $\frac{\partial I(\boldsymbol{p},\boldsymbol{\gamma})}{\partial \gamma_j}$ is positive. Decreasing x_j towards x_1 increases γ_j towards γ_1 such that for fixed $p_1,\dots,p_m$ the inequality $I(\boldsymbol{p},\tilde{\boldsymbol{\gamma}}) \geq I(\boldsymbol{p},\boldsymbol{\gamma})$ results. The sequence $\tilde{\boldsymbol{\gamma}}$ is identical to $\boldsymbol{\gamma}$ with the exception of $\gamma_j = \gamma_1$.

Vice versa, if $\gamma_j \leq \bar{\gamma}_j$, the derivative $\frac{\partial I(\boldsymbol{p},\boldsymbol{\gamma})}{\partial \gamma_j}$ is negative. Increasing x_j towards x_m decreases γ_j towards γ_m such that $I(\boldsymbol{p},\tilde{\boldsymbol{\gamma}}) \geq I(\boldsymbol{p},\boldsymbol{\gamma})$ results. Analogously, the sequence $\tilde{\boldsymbol{\gamma}}$ is identical to $\boldsymbol{\gamma}$ with the exception of $\gamma_j = \gamma_m$.

In both cases we end up with an increased $I(\boldsymbol{p},\tilde{\boldsymbol{\gamma}})$ having support points $x_1,\dots,x_{j-1},x_{j+1},\dots,x_m$ with probabilities $(p_1+p_j,\dots,p_{j-1},p_{j+1},\dots,p_m)$ or $(p_1,\dots,p_{j-1},p_{j+1},\dots,p_j+p_m)$, respectively.

Iterating this process for all $1 < j < m$ yields a distribution with only two points x_1 and x_m of positive probability. ∎

The explicit form of the capacity-achieving distribution has been derived in [10]. We here present a different form by a different proof, which finally allows for determining the optimal threshold q.

Proposition 3: For any fixed support points $x_1 < \cdots < x_m$ and any threshold q the capacity-achieving distribution $\boldsymbol{p}^\star$ is concentrated on the extreme support points x_1 and x_m with probabilities

$$p_1^\star = \frac{1-(1+a^s)\gamma_m}{(1+a^s)(\gamma_1-\gamma_m)} \quad \text{and} \quad p_m^\star = \frac{(1+a^s)\gamma_1 - 1}{(1+a^s)(\gamma_1-\gamma_m)}. \tag{6}$$

The corresponding channel capacity is given as

$$\max_{\boldsymbol{p}} I(\boldsymbol{p},\boldsymbol{\gamma}) = \log_a\big(1+a^s\big) - (s+t) \tag{7}$$

with constants

$$s = \frac{h(\gamma_1)-h(\gamma_m)}{\gamma_1-\gamma_m} \quad \text{and} \quad t = \frac{\gamma_1 h(\gamma_m) - \gamma_m h(\gamma_1)}{\gamma_1-\gamma_m}. \tag{8}$$

Proof. Since $I(\boldsymbol{p},\boldsymbol{\gamma})$ is a concave function of $\boldsymbol{p}$, the maximization problem $\max_{\boldsymbol{p}} I(\boldsymbol{p},\boldsymbol{\gamma})$ is a convex optimization program such that the Karush-Kuhn-Tucker conditions characterize the global optimum. The partial derivatives w.r.t. p_j of the corresponding Lagrangian

$$L(\boldsymbol{p},\boldsymbol{\gamma};t,\boldsymbol{\mu}) = -I(\boldsymbol{p},\boldsymbol{\gamma}) + \left(1-\sum_{i=1}^m p_i\right)t - \sum_{i=1}^m p_i\mu_i$$

must vanish for all $j = 1,\dots,m$. After some algebra we find that $\partial L(\boldsymbol{p},\boldsymbol{\gamma};t,\boldsymbol{\mu})/\partial p_j = 0$ holds if and only if $h(\gamma_j) = \gamma_j s + t + \mu_j$, for all $j = 1,\dots,m$ with

$$s = \log_a \frac{1-\sum_{i=1}^m p_i\gamma_i}{\sum_{i=1}^m p_i\gamma_i}. \tag{9}$$

Since $h(\gamma)$ is a strictly concave function of γ there are at most two intersecting points with the affine function $\gamma s + t + \mu_j$. Hence, the capacity-achieving distribution has only two mass points with $p_i + p_j = 1$. From Proposition 2 we know that these correspond to the extreme support points x_1 and x_m. The complementary slackness condition states that $\mu_j = 0$ whenever $p_j > 0$. Hence, $t = h(\gamma_1) - \gamma_1 s$ and $t = h(\gamma_m) - \gamma_m s$ follows. These equations allow for determining t and s explicitly, as is presented in (8).

From (9) the probabilities in (6) are deduced. Finally, the optimum value $I(\boldsymbol{p}^\star,\boldsymbol{\gamma})$ in (7) is derived by incorporating $\boldsymbol{p}^\star$, s and t into $I(\boldsymbol{p},\boldsymbol{\gamma})$. All step-by-step computations above are rather lengthy and tedious to accomplish and are thus omitted for reasons of brevity. ∎

Concentrating on the extreme support points allows for an interpretation as a binary asymmetric channel where the error probabilities are given by $\Phi(q-x_m)$ and $\Phi(q-x_1)$, respectively. Related early works on determining the capacity are [12]–[15].

We now set out to demonstrate that capacity increases as the distance between x_1 and x_m does.

Proposition 4: Set $\gamma_1 = \gamma$ and $\gamma_m = \gamma + \delta$, and consider the capacity $I(\gamma_1,\gamma_m) = \max_{\boldsymbol{p}} I(\boldsymbol{p},\boldsymbol{\gamma})$ derived in (7). Then for any $0 \leq \gamma \leq 1$ the capacity $I(\gamma,\gamma+\delta)$ is an increasing function of δ, $0 \leq \delta \leq 1-\gamma$.

Proof. The first derivative of $I(\gamma,\gamma+\delta)$ w.r.t. δ is given by

$$I(\gamma,\gamma+\delta)' = -\Big(\frac{1}{1+a^s} - \gamma\Big) \tag{10}$$

$$\cdot \frac{(\gamma+\delta)h(\gamma) - \gamma h(\gamma+\delta) + \delta\log_a(1-\gamma-\delta)}{\delta^2(\gamma+\delta)}. \tag{11}$$

We first show that the first factor (10) is non-negative, and then prove that the quotient in (11) is less than or equal to zero.

By Taylor's theorem [16, p. 14, eq. 3.6.1-3.6.5] with the Lagrange form of the remainder and a proper number ξ, $0 \leq \xi \leq \delta$, the following representation is obtained:

$$h(\gamma+\delta) - h(\gamma) = \delta \underbrace{\log_a\Big(\frac{1-\gamma}{\gamma}\Big)}_{=h'(\gamma)} + \frac{\delta^2}{2}\underbrace{\frac{-1/\ln(a)}{(1-\xi)\xi}}_{=h''(\xi)}.$$

Since the second derivative $h''(\xi)$ is always negative, it holds that $h(\gamma+\delta) - h(\gamma) \geq \delta\log_a\big(\frac{1-\gamma}{\gamma}\big)$ which is equivalent to $\frac{1}{1+a^s} \geq \gamma$.

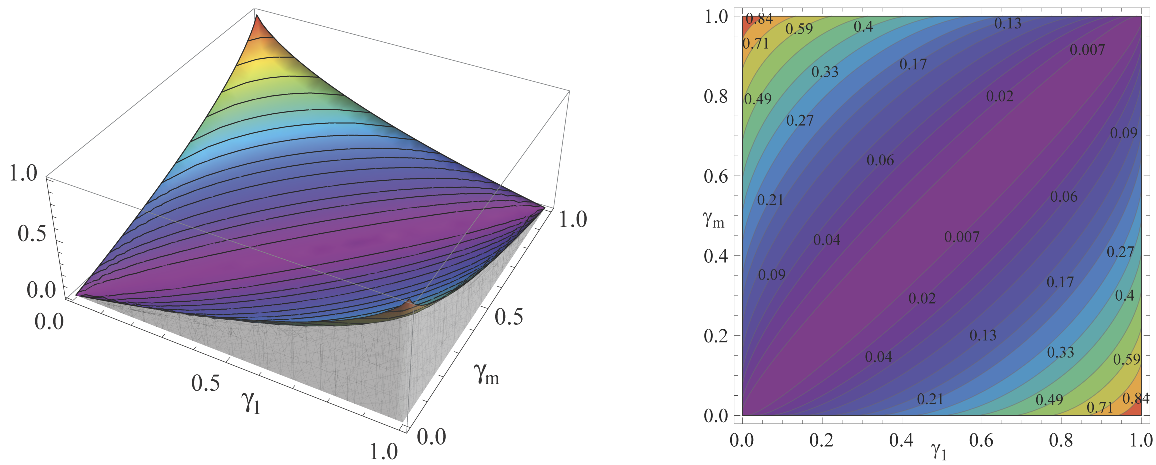
From the well-known inequality $\log_a p \leq \frac{p-1}{\ln a}$ it follows that

$$\begin{aligned} h(\gamma) &- \frac{\gamma}{\gamma+\delta}h(\gamma+\delta) + \frac{\delta}{\gamma+\delta}\log_a(1-\gamma-\delta) \\ &= \gamma\log_a\Big(\frac{\gamma+\delta}{\gamma}\Big) + (1-\gamma)\log_a\Big(\frac{1-\gamma-\delta}{1-\gamma}\Big) \\ &\leq \frac{\gamma}{\ln(a)}\Big(\frac{\gamma+\delta}{\gamma} - 1\Big) + \frac{1-\gamma}{\ln(a)}\Big(\frac{1-\gamma-\delta}{1-\gamma} - 1\Big) = 0 \end{aligned}$$

which shows that the ratio (11) is non-positive.

In total, the derivative of $I(\gamma,\gamma+\delta)$ w.r.t. δ is non-negative and $I(\gamma,\gamma+\delta)$ is an increasing function of δ. ∎

Corollary 5: A conclusion of the previous proposition is that due to the monotonicity of Φ, capacity is increasing with respect to the difference $|x_m - x_1|$.



(a) 3d-visualization with contour lines

(b) Representation only by contour lines (top-view)

Fig. 2: Mutual information $I(\gamma_1, \gamma_m)$ from (12) as a function of (γ_1, γ_m).

Capacity (7) as a function of (γ_1, γ_m) has the following representation which reveals inherent symmetry properties. The derivation of (12) needs straightforward tedious algebra.

$$\begin{aligned} I(\gamma_1, \gamma_m) &= \log_a\left(a^{\frac{(1-\gamma_1)h(\gamma_m)-(1-\gamma_m)h(\gamma_1)}{\gamma_1-\gamma_m}} + a^{\frac{\gamma_m h(\gamma_1)-\gamma_1 h(\gamma_m)}{\gamma_1-\gamma_m}}\right) \quad (12) \\ &= \log_a\left(a^{\frac{(1-\gamma_1)(1-\gamma_m)}{\gamma_1-\gamma_m}\int_{\gamma_m}^{\gamma_1}\frac{\log_a(\gamma)d\gamma}{(1-\gamma)^2}} + a^{\frac{\gamma_m\gamma_1}{\gamma_1-\gamma_m}\int_{\gamma_m}^{\gamma_1}\frac{\log_a(1-\gamma)d\gamma}{\gamma^2}}\right) \end{aligned}$$

Since the capacity-achieving distribution has only two mass points x_1 and x_m, and the maximum capacity w.r.t. a threshold q is invariant to an additive shift of the support points, cf. Proposition 1, the support points of the input distribution may be chosen symmetrically to zero as $x_1' = (x_1 - x_m)/2$ and $x_m' = (x_m - x_1)/2$ without changing capacity.

Determining the capacity-maximizing threshold $q^\star$ in general is an extremely hard problem, the reason for which becomes clear from the following graphical representations.

IV. Selected Numerical Results

First, mutual information $I(\gamma_1, \gamma_m)$ from (12) is depicted as a function of two variables, neglecting that γ_1 and γ_m are subject to

$$\gamma_1 = \Phi(q - x_1) \quad \text{and} \quad \gamma_m = \Phi(q - x_m)\,, \ q \in \mathbb{R}. \qquad (13)$$

The corresponding three-dimensional plot and contour plot are shown in Fig. 2a and Fig. 2b, respectively. The shape of the surface $I(\gamma_1, \gamma_m)$ indicates convexity, which is actually true and can also be proven analytically.

Further, three types of noise distributions are considered, the Gaussian, uniform and mixture of two Gaussians. In the first row of Fig. 3 the corresponding densities are shown. In the second row the shifted cumulative distribution functions $\Phi(q-x_i)$ and their differences $\Phi(q-x_1)-\Phi(q-x_m)$ are plotted as a function of q; solid curves for the case $x_m - x_1 = 1$ and dashed ones for $x_m - x_1 = 5/2$. The third row contains the contour lines of $I(\gamma_1, \gamma_m)$ overlaid by the trajectories $(\gamma_1, \gamma_m)(q) = (\Phi(q - x_1), \Phi(q - x_m))$, $q \in \mathbb{R}$.

Some interesting phenomena may be observed which make the problem $\max_q I(\gamma_1, \gamma_m)$, such that (13) holds, quite complicated. In the case of Gaussian noise the optimum seems to be always attained at some point $\gamma_1^\star = 1 - \gamma_m^\star$, $\gamma_m^\star \in (0, 1)$, leading to a threshold in the middle between x_1 and x_m. The same holds true for certain cases of two Gaussian mixtures, however if the spacing between x_1 and x_m is too large, two optima occur close to the boundaries, which are asymmetric in the sense that $\gamma_1^\star \neq 1 - \gamma_m^\star$. For the uniform distribution the optimum is always attained at the boundary values either with $\gamma_m^\star = 0$ or/and with $\gamma_1^\star = 1$. The solutions probably do not have a general explicit form for arbitrary noise distributions. Derivatives of $I(\gamma_1, \gamma_m)$ w.r.t. q and a Lagrangian approach seem to be technically too demanding to deal with for further investigations.

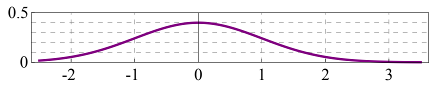
V. Conclusions

We have considered a one-bit quantization channel, where a discrete real signal undergoes additive noise and is then quantized according to a threshold q. We first have argued that it is sufficient to deal with discrete input, since for input with finite support the capacity-achieving distribution is discrete anyway. We have shown that the capacity-achieving input distribution puts only probability mass on the extreme signaling points. Determining the optimum threshold $q^\star$ is extremely demanding. We have clarified the structure of the problem and demonstrated by graphical representations that depending on the actual parameters and noise distribution completely different behavior may be observed.

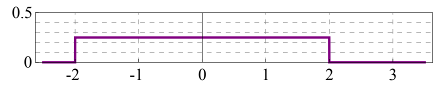
References

[1] R. Mathar and A. Schmeink, "A bio-inspired approach to condensing information," in *IEEE International Symposium on Information Theory (ISIT)*, Saint-Petersburg, Aug. 2011, pp. 2524–2528.

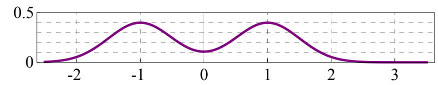
[2] A. Mezghani and J. Nossek, "Optimal 1-bit quantizer for the awgn channel," Munich University of Technology, Institute for Circuit Theory and Signal Processing, Tech. Rep. TUM-LNS-TR-01-02, 2013.



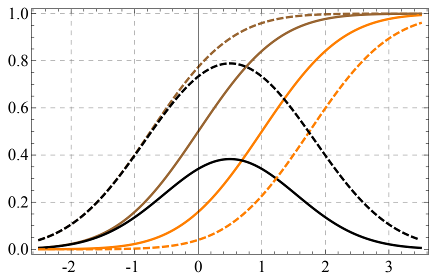
(a) PDF of a Gaussian noise with unit variance.



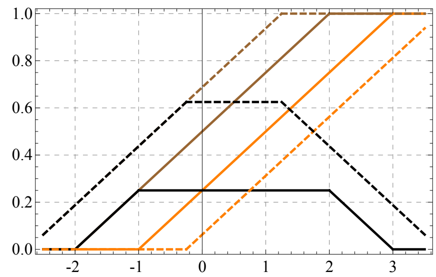
(b) PDF of a uniformly distributed noise with variance of $4/3$.



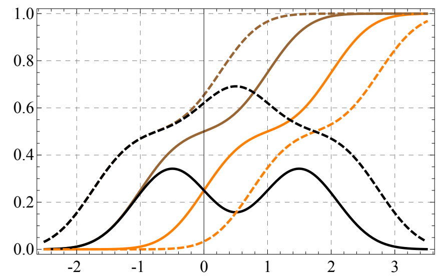
(c) PDF of a mixture of two Gaussians – each of which with variance of $1/4$ and a distance of 2.



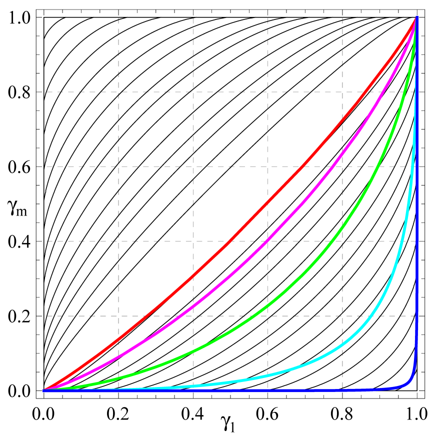
(d) CDFs and difference of CDFs for a Gaussian distribution.



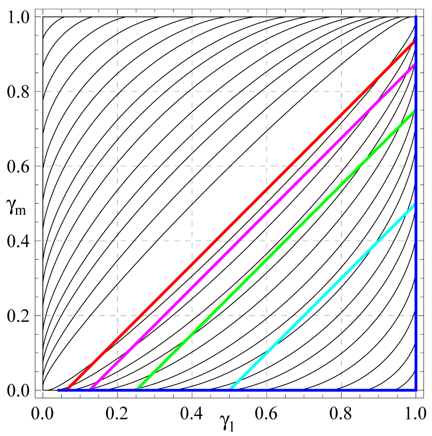
(e) CDFs and difference of CDFs for a uniform distribution.



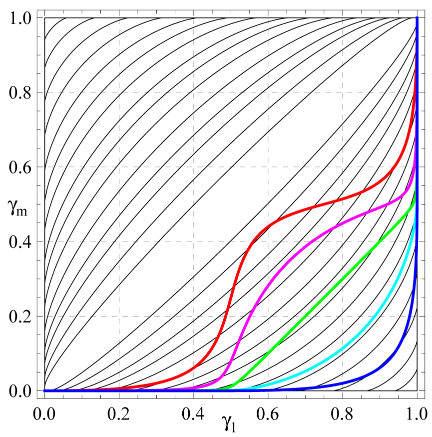
(f) CDFs and difference of CDFs for a mixture of two Gaussian distributions.



(g) Contours of $I(\gamma_1, \gamma_m)$ together with trajectories of a Gaussian noise for $x_m - x_1 \in \{1/4, 1/2, 1, 2, 4\}$.



(h) Contours of $I(\gamma_1, \gamma_m)$ together with trajectories of a uniformly distributed noise for $x_m - x_1 \in \{1/4, 1/2, 1, 2, 4\}$.



(i) Contours of $I(\gamma_1, \gamma_m)$ together with trajectories of a mixture of two Gaussians for $x_m - x_1 \in \{1, 3/2, 2, 5/2, 3\}$.

Fig. 3: Graphical representation of PDF, CDF, difference of CDFs, and contour plots of $I(\gamma_1, \gamma_m)$ together with overlaid trajectories $(\gamma_1, \gamma_m) = (\Phi(q - x_1), \Phi(q - x_m)), q \in \mathbb{R}$, for selected noise distributions. The PDFs, CDFs, and difference of CDFs are plotted as functions of $q \in \mathbb{R}$. For CDFs and their differences the solid curves correspond to $x_m - x_1 = 1$ and dashed curves to $x_m - x_1 = 5/2$. All trajectories in the contour plots start at $(\gamma_1, \gamma_m) = (0, 0)$ for $q \mapsto -\infty$ and they end at $(1, 1)$ for $q \mapsto \infty$.

[3] T. Koch and A. Lapidoth, "Asymmetric quantizers are better at low SNR," in *International Symposium on Information Theory, ISIT 2011*, St Petersburg, Russia, 2011, pp. 2592–2596.

[4] J. Singh, O. Dabeer, and U. Madhow, "On the limits of communication with low-precision analog-to-digital conversion at the receiver," *IEEE Transactions on Communications*, vol. 57, no. 12, pp. 3629–3639, December 2009.

[5] T. Koch and A. Lapidoth, "At low snr, asymmetric quantizers are better," *Information Theory, IEEE Transactions on*, vol. 59, no. 9, pp. 5421–5445, Sep. 2013.

[6] G. Alirezaei and R. Mathar, "On the discreteness of capacity-achieving input distributions with finite support," Institute for Theoretical Information Technology, RWTH Aachen University, Tech. Rep., 2015.

[7] P. Amblard, O. Michel, and S. Morfu, "Revisiting the asymmetric binary channel: joint noise enhanced detection and information transmission through threshold devices," in *Noise in Complex Systems and Stochastic Dynamics III*, 2005, pp. 50–60.

[8] A. Magnani, A. Ghosh, and R. M. Gray, "Optimal one-bit quantization," in *Proceedings of the Data Compression Conference, DCC'05*, Snowbird, Utah, USA, 2005.

[9] S. Krone and G. Fettweis, "Fading channels with 1-bit output quantization: optimal modulation, ergodic capacity and outage probability," in *Proceedings of the IEEE Information Theory Workshop (ITW'10)*, Dublin, August 2010, pp. 50–60.

[10] R. Mathar and M. Dörpinghaus, "Threshold optimization for capacity-achieving discrete input one-bit output quantization," in *IEEE International Symposium on Information Theory (ISIT 2013)*, Istanbul, Turkey, Jul. 2013, pp. 1999–2003.

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

[12] S. Muroga, "On the capacity of a discrete channel I," *Journal of the Physical Society of Japan*, vol. 8, no. 4, pp. 484–494, 1953.

[13] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *Information Theory, IEEE Transactions on*, vol. 18, no. 1, pp. 14–20, Jan 1972.

[14] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *Information Theory, IEEE Transactions on*, vol. 18, no. 4, pp. 460–473, Jul 1972.

[15] S. Takano, "On a method of calculating the capacity of a discrete memoryless channel," *Information and Control*, vol. 29, pp. 327–336, 1975.

[16] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 10th ed. Washington D.C.: National Bureau of Standards, 1972.