Capacity Achieving Quantizer Design for Binary Channels

Thuan Nguyen, *Graduate Student Member, IEEE* and Thinh Nguyen, *Senior Member, IEEE*

***Abstract*—We consider a communication channel with a binary input X being distorted by an arbitrary continuous-valued noise which results in a continuous-valued signal Y at the receiver. A quantizer Q is used to quantize Y back to a binary output Z. Our goal is to determine the optimal quantizer Q^* and the corresponding input probability mass function $\mathbf{p}_X^*$ that achieve the capacity. We present a new lower bound and a new upper bound on the capacity in terms of quantization parameters and the structure of the associated channel matrix. Based on these theoretical results, we propose an efficient algorithm for finding the optimal quantizer.**

***Index Terms*—Quantization, mutual information, capacity.**

I. Introduction

A primary goal of a communication system is to transmit the information reliably and fast over an error-prone channel. The fastest rate with a vanishing error for a given channel is equal to the maximum mutual information $I(X;Z)$ between two random variables X and Z used to model the input and the output of channel. For a given discrete memoryless channel (DMC) specified by a channel matrix $\mathbf{A}$, it is well-known that the mutual information is a concave function in the input probability mass function $\mathbf{p}_X$ [1]. Consequently, determining the capacity achieving optimal input distribution $\mathbf{p}_X^*$ that maximizes $I(X;Z)$ for a given $\mathbf{A}$ is not difficult using existing convex optimization algorithms or other iterative algorithms [2]. Furthermore, under some special conditions on $\mathbf{A}$, it is possible to obtain closed-form expressions for the capacities of many DMCs [1], [3], [4], [5].

On the other hand, rather than using a given channel matrix $\mathbf{A}$, one assumes a given input distribution $\mathbf{p}_X$. The goal is to design an optimal quantizer Q^*, which is equivalent to selecting an optimal channel matrix $\mathbf{A}^*$ subject to a certain structure that maximizes the mutual information between the input X and the quantized output Z [6]–[10]. We note that this is not the same as designing a quantizer that achieves the capacity since the input distribution $\mathbf{p}_X$ is given. Our goal is to determine the optimal quantizer Q^* together with the optimal input distribution $\mathbf{p}_X^*$ that achieves the channel capacity. To the best of our knowledge, this problem still remains a hard problem for a general setting [11]–[14]. In [13], Singh et al. provided an algorithm for multilevel quantization, which gave near-optimal results. In [12], the author proposed a heuristic near-optimal quantization algorithm which alternatively maximizes the mutual information for a given quantizer and minimizes the probability of error for a given input distribution. However, this algorithm only works well when the signal-to-noise ratio of the channel is high. For 2-level (1-bit) quantization of general additive channels, Mathar and Dorpinghaus proved that the optimal mutual information could be achieved by using an input distribution between two support points [11]. However, it is worth noting that the result in [11] is limited for single threshold quantizers and the truly optimal quantizer may contain more than one threshold [15]. In [14], the author gave a near-optimal algorithm to find the optimal value of mutual information for binary input and an arbitrary number of the quantized output, however, this algorithm may declare a failure outcome. There are also several recent works on finding the channel capacity for Gaussian channels with quantized output. In [16], Vu et al. investigated the problem of designing the optimal signaling schemes together with capacity-achieving input distribution for Gaussian channels under the assumption of a low-resolution output quantization. In [17], Ranjbar et al. constructed the capacity region and capacity-achieving signaling schemes for 1-bit quantization with two users communicating in Rayleigh-fading channels. These works focus on finding the optimal input distribution for a pre-specified channel (Gaussian and Rayleigh) and under a given quantization scheme. In contrast, our work is more general as our focus is on obtaining both an optimal quantization scheme and optimal input distribution simultaneously. Furthermore, our results can be applied to any communication channel specified by an arbitrary conditional density of the received output given the transmitted input.

In this paper, we consider a special case where the channel matrix $\mathbf{A}$ is a 2×2 matrix. In particular, we consider a communication channel with a binary input X being distorted by a given arbitrary continuous-valued noise which results in a continuous-valued signal Y at the receiver. A quantizer Q is used to quantize Y back to a binary output Z. Our goal is to determine the optimal quantizer Q^*, and therefore, an induced optimal $\mathbf{A}^*$ that achieves the capacity. Importantly, we do not assume that $\mathbf{p}_X$ is given. Rather, after the optimal $\mathbf{A}^*$ is determined, the optimal $\mathbf{p}_X^*$ then can be obtained using any classic method. The main contributions of this paper include the new lower bound and upper bound of the capacity in terms of the quantization parameters, together with the structure of the associated channel matrix. Based on these, we propose an efficient algorithm for finding Q^*.

II. Problem description

We consider the setting shown in Fig. 1. The binary input modeled as a random variable $X \in \{0,1\}$, is transmitted over a channel that distorts X into a continuous valued signal modeled as a random variable Y at the receiver. The channel distortion

Thuan Nguyen and Thinh Nguyen are with the School of Electrical Engineering and Computer Science, Oregon State University, Oregon, OR, 97331 USA, e-mail: (nguyeth9@oregonstate.edu, thinhq@eecs.oregonstate.edu).

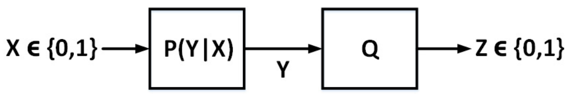

Figure 1: A binary input $X = \{0, 1\}$ is transmitted over a noisy channel which results in a continuous-valued $y \in Y$ at the receiver. The receiver attempts to recover X by quantizing Y to a discrete binary signal $Z = \{0, 1\}$.

is modeled by a conditional density of Y given X: $f_{Y|X}(y|x)$. To recover X, the receiver uses a quantizer Q that quantizes Y to a binary signal $Z \in \{0, 1\}$. Formally,

$$Q(y) = \begin{cases} z = 0 & \text{if } y \in \mathbb{H}, \\ z = 1 & \text{if } y \in \bar{\mathbb{H}}, \end{cases} \tag{1}$$

where $\mathbb{H} \cap \bar{\mathbb{H}} = \emptyset$ and $\mathbb{H} \cup \bar{\mathbb{H}} = \mathbb{R}$. For a given conditional density $f_{Y|X}(y|x)$, our goal is to design an optimal quantizer Q^* together with an optimal input distribution $\mathbf{p}_X^*$ such that the mutual information $I(X; Z)$ between X and Z is maximized:

$$Q^*, \mathbf{p}_X^* = \underset{Q, \mathbf{p}_X}{\operatorname{argmax}}\, I(X; Z). \tag{2}$$

III. Preliminaries

We consider the setting in Fig. 1. Let $\mathbf{p}_X = (p_0, p_1)$ is the input probability mass distribution and $f_{Y|X}(y|x)$ is the conditional density function of Y given X. For given $f_{Y|X}(y|x)$, let $\phi_0(y) = f_{Y|X}(y|x = 0)$ and $\phi_1(y) = f_{Y|X}(y|x = 1)$ and define:

$$u(y) = \frac{p_1 \phi_1(y)}{p_0 \phi_0(y) + p_1 \phi_1(y)}. \tag{3}$$

Definition 1. A binary quantizer Q_u is called a convex quantizer if it has the following structure:

$$Q_u(y) = \begin{cases} z = 0 & \text{if } u(y) \leq u, \\ z = 1 & \text{if } u(y) > u, \end{cases} \tag{4}$$

where $0 < u < 1$.

Burshtein et al. [18] and Kurkoski and Yagi [15] showed that the optimal binary quantizer is indeed a convex quantizer as stated in Theorem 1 below.

Theorem 1. [18], [15] For a given p_0 and p_1, the optimal binary quantizer that maximizes the mutual information $I(X; Z)$ is a convex quantizer Q_{u^*} for some optimal threshold u^*.

We should make an important remark about Theorem 1.

Remark 1. Q_{u^*} is not a capacity achieving quantizer even though it maximizes $I(X; Z)$. This is because Q_{u^*} assumes a given input distribution $\mathbf{p}_X$. On the other hand, our goal is to find the capacity achieving quantizer Q^* which maximizes $I(X; Z)$ over all the possible $\mathbf{p}_X$. A straightforward way of applying Theorem 1 to find Q^* is to search over all possible values of p_0, p_1, and u^* that maximizes $I(X; Z)$. This is however still a 2-dimensional search.

Next, instead of given $\mathbf{p}_X$, suppose a quantizer Q is given, we want to determine the capacity $C = \max_{\mathbf{p}_X} I(X; Z)$. The given Q induces a channel matrix:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

where $a_{12} = 1 - a_{11}$ and $a_{21} = 1 - a_{22}$. We will show how $\mathbf{A}$ is related to Q shortly. The capacity of this binary DMC is given in Theorem 2 below [3].

Theorem 2. [3] The capacity of a binary DMC for a given channel matrix $\mathbf{A}$ is:

$$\begin{aligned} C &= \log_2 \Big(2^{-\dfrac{a_{22} H(a_{11}) + (a_{11} - 1) H(a_{22})}{a_{11} + a_{22} - 1}} \\ &+ 2^{-\dfrac{(a_{22} - 1) H(a_{11}) + a_{11} H(a_{22})}{a_{11} + a_{22} - 1}} \Big), \end{aligned} \tag{5}$$

where $H(w) = -w \log_2(w) - (1 - w) \log_2(1 - w)$.

We will use Theorems 1 and 2 to describe a more efficient procedure for finding the capacity achieving Q^*.

IV. Design of Capacity Achieving Quantizer

Theorem 3 below is a variant of Theorem 1 which will be used to design a capacity achieving quantizer.

Theorem 3. (Structure of optimal quantizer)
For given $\phi_0(y)$ and $\phi_1(y)$, define:

$$r(y) = \frac{\phi_0(y)}{\phi_1(y)}. \tag{6}$$

Let Q_r be a convex quantizer with the following structure:

$$Q_r(y) = \begin{cases} 0 & \text{if } r(y) \geq r, \\ 1 & \text{if } r(y) < r, \end{cases} \tag{7}$$

for some $0 < r < \infty$, then there exists a capacity achieving convex quantizer Q_{r^*} for some optimal threshold r^*.

Proof. For any $\mathbf{p}_X$, we have:

$$u(y) = \frac{p_1 \phi_1(y)}{p_0 \phi_0(y) + p_1 \phi_1(y)} = \frac{1}{\dfrac{p_0}{p_1} \dfrac{\phi_0(y)}{\phi_1(y)} + 1} = \frac{1}{\dfrac{p_0}{p_1} r(y) + 1}. \tag{8}$$

Thus,

$$r(y) = \frac{p_1 (1 - u(y))}{p_0 u(y)}. \tag{9}$$

Now using Theorem 1, and writing $u(y)$ in terms of $r(y)$, we obtain:

$$r^* = \frac{p_1 (1 - u^*)}{p_0 u^*}. \tag{10}$$

Furthermore, for any valid $p_0 > 0$, it is straightforward to show that $0 < r^* < \infty$. It is important to note that p_0 and p_1 need not to be given, even though they are related to $r(y)$ through (9) for some p_0 and p_1. Instead, $r(y)$ is defined as $r(y) = \dfrac{\phi_0(y)}{\phi_1(y)}$. If there is a method to find the optimal r^* directly, then the corresponding p_0^* and p_1^* can be found based on r^*. Importantly, since Q_{r^*} maximizes $I(X; Z)$ (by Theorem 1) without given p_0 and p_1, Q_{r^*} is a capacity achieving quantizer. □

Remark 2. The use of $r(y)$ in Theorem 3 rather than $u(y)$ in Theorem 1 is an important step in designing a capacity achieving quantizer. $r(y)$ as defined in (6), does not depend on p_0 and p_1. Therefore, to find a capacity achieving quantizer, one can employ an exhaustive search to find the optimal threshold

r^*. Specifically, for each value of the threshold r, a quantizer Q can be constructed based on Theorem 3 in which one compares $r(y)$ with r. This comparison does not need p_0 and p_1. On the other hand, using $u(y)$ and search for the optimal u^* in Theorem 1, one is required to know p_0 and p_1 since $u(y)$ is defined in terms of p_0 and p_1.

We now derive the channel matrix $\mathbf{A}$ for a given quantizer Q_r in Theorem 3. Define:

$$\mathbb{H}_r = \{y : r(y) = \frac{\phi_0(y)}{\phi_1(y)} \geq r\}, \tag{11}$$

$$\bar{\mathbb{H}}_r = \{y : r(y) = \frac{\phi_0(y)}{\phi_1(y)} < r\}. \tag{12}$$

Thus,

$$Q_r(y) = \begin{cases} z = 0 & \text{if } y \in \mathbb{H}_r, \\ z = 1 & \text{if } y \in \bar{\mathbb{H}}_r. \end{cases} \tag{13}$$

The channel matrix $\mathbf{A}$ that corresponds to quantizer Q_r is:

$$\mathbf{A} = \begin{bmatrix} a_{11}(r) & a_{12}(r) \\ a_{21}(r) & a_{22}(r) \end{bmatrix},$$

where

$$a_{11}(r) = \int_{y \in \mathbb{H}_r} \phi_0(y) dy, \tag{14}$$

$$a_{22}(r) = \int_{y \in \bar{\mathbb{H}}_r} \phi_1(y) dy, \tag{15}$$

and $a_{12}(r) = 1 - a_{11}(r)$, $a_{21}(r) = 1 - a_{22}(r)$.

Using Theorem 2 and Theorem 3, the capacity in (5) is a function of r:

$$C(r) = \log_2\Big(2^{-\frac{a_{22}(r)H(a_{11}(r)) + (a_{11}(r)-1)H(a_{22}(r))}{a_{11}(r)+a_{22}(r)-1}} + 2^{-\frac{(a_{22}(r)-1)H(a_{11}(r)) + a_{11}(r)H(a_{22}(r))}{a_{11}(r)+a_{22}(r)-1}}\Big). \tag{16}$$

We note that each value of r corresponds to a different channel matrix $\mathbf{A}$ associated with a different Q_r. Therefore, based on (16), an exhaustive search can be used to find r^* that maximizes $C(r)$. This is a one-dimensional search on r which is more efficient than searching for u, p_0, and p_1 as discussed earlier. Furthermore, we will derive an upper and lower bound on r to increase the search efficiency. Lemma 1 below describes the structure of the channel matrix that corresponds to a convex quantizer Q_r.

Lemma 1. (Structure of the channel matrix induced by Q_r)
For $\forall\ r \in (0, +\infty)$,
(1) $a_{11}(r) \in (0, 1)$ and is a monotonic decreasing function.
(2) $a_{22}(r) \in (0, 1)$ and is a monotonic increasing function.
(3) $1 < a_{11}(r) + a_{22}(r) \leq a_{11}(1) + a_{22}(1)$.

Proof. Please see the proof in Appendix A. □

Theorem 4. (Capacity bounds)
Define $\delta = a_{11}(1) + a_{22}(1)$, then the maximum capacity $C(r^*)$ over all possible channel matrices induced by all convex quantizers Q_r is bounded by:

$$1 - H(\frac{2-\delta}{2}) \leq C(r^*) \leq \log_2 \delta. \tag{17}$$

Proof. From Lemma 1, $\forall r$, we have:

$$a_{11}(r) + a_{22}(r) > 1 = a_{11}(r) + a_{12}(r), \tag{18}$$

$$a_{11}(r) + a_{22}(r) > 1 = a_{21}(r) + a_{22}(r). \tag{19}$$

Thus,

$$a_{22}(r) > a_{12}(r), \tag{20}$$

$$a_{11}(r) > a_{21}(r). \tag{21}$$

Upper bound: The Boyd-Chiang's upper bound [19] of the channel capacity associated with a given channel matrix $\mathbf{A}$ is:

$$C_{\mathbf{A}} \leq \log_2(\sum_j \max_i a_{ij}). \tag{22}$$

For a binary channel associated with a convex quantizer Q_r, using (22), we have:

$$C(r) \leq \log_2\big(\sum_{j=1} \max_i a_{ij}(r)\big) = \log_2\big(a_{11}(r) + a_{22}(r)\big) \tag{23}$$

$$\leq \log_2\big(a_{11}(1) + a_{22}(1)\big) = \log_2 \delta, \tag{24}$$

where (23) is due to (20) and (21), (24) is due to (3) in Lemma 1. Since the upper bound in (24) holds for every r, it must hold for r^*.

Lower bound: Recall that the Fano's inequality [1] with alphabet size of $|X| = 2$ is:

$$H(X|Z) \leq H(p_e) + p_e \log(|X| - 1) = H(p_e), \tag{25}$$

where p_e is the probability of error when transmitting a signal over the channel and using a quantizer Q_r for recovering the signal. Next, using the uniform input distribution $\mathbf{p}_X$ i.e., $p_0 = p_1 = 1/2$ and the convex quantizer Q_1 ($r = 1$), we have:

$$p_e = p_0 a_{12}(1) + p_1 a_{21}(1) = \frac{1}{2}(a_{12}(1) + a_{21}(1)) \tag{26}$$

$$= \frac{1}{2}(2 - a_{11}(1) - a_{22}(1)) = \frac{2-\delta}{2}, \tag{27}$$

and $H(X) = 1$.

Now, since the maximum capacity $C(r^*)$ is at least as large as the mutual information using $p_0 = p_1 = 1/2$, and $r = 1$, from (25) and (27), we have:

$$C(r^*) \geq I(X;Z) = H(X) - H(X|Z) \tag{28}$$

$$\geq H(X) - H(p_e) = 1 - H(\frac{2-\delta}{2}). \tag{29}$$

□

Remark 3. We note that the lower bound reaches to the upper bound when $\delta \to 2$ or $\delta \to 1$ as illustrated in Fig. 2. Moreover, a larger value of δ implies a smaller overlapped area between the noise density $\phi_0(y)$ and $\phi_1(y)$ or a higher probability of correct decoding. In an additive channel with an identical noise for transmitting symbols 0 and 1, $\phi_0(y)$ and $\phi_1(y)$ are shifted versions of each other. The larger shift results in a larger value of δ and a higher probability of correct decoding. Thus, our bounds are tighter for low noise regimes. The upper and lower bounds as functions of δ are visualized in Fig. 2.

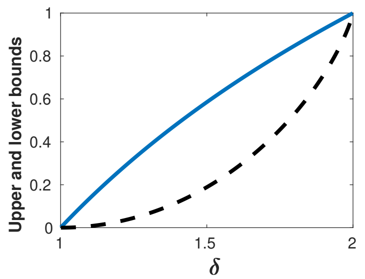

Figure 2: Upper bound and lower bound of channel capacity as functions of δ.

Also, as an extension, if the input size is more than two, then using the identical proof, the δ in the upper bound in Theorem 4 is:

$$\delta = \int_{y \in Y} \max_i \phi_i(y) dy, \tag{30}$$

where $\phi_i(y) = f_{Y|X}(y|x_i), i = 1, 2, \ldots, N$ with N being the size of the input alphabet.

Next, we will use Theorem 4 to narrow down the range to search for the optimal r^*. We have the following theorem.

Theorem 5. (Bound on optimal r^*)

Let $0 < v \leq \frac{1}{2}$ be a positive number such that:

$$H(v) = H(1-v) = 1 - H(\frac{2-\delta}{2}). \tag{31}$$

If Q_{r^*} is optimal, then:

$$a_{11}(r^*) \geq v, \tag{32}$$

$$a_{22}(r^*) \geq v. \tag{33}$$

Furthermore, $r^* \in [r_2, r_1]$ where $a_{11}(r_1) = a_{22}(r_2) = v$.

Proof. Suppose that a quantizer Q_r produces $H(Z)$, and

$$H(Z) \leq 1 - H(\frac{2-\delta}{2}) = H(v) = H(1-v) \tag{34}$$

for some $v \in (0, 0.5]$. Since $1 - H(\frac{2-\delta}{2}) \geq H(Z) \geq I(X;Z)$, based on the lower bound of Theorem 4, Q_r cannot be an optimal quantizer.

We will show that if:

$$a_{11}(r) < v, \tag{35}$$

$$a_{22}(r) < v, \tag{36}$$

then Q_r is suboptimal.

Since the binary entropy is symmetric i.e., $H(v) = H(1-v)$ and $v \leq 1/2$, then $v \leq 1/2 \leq 1 - v$. From (21),

$$\begin{aligned} p(Z{=}0) &= p_0 a_{11}(r) + p_1 a_{21}(r) \geq p_0 a_{21}(r) + p_1 a_{21}(r) \\ &= a_{21}(r) = 1 - a_{22}(r). \end{aligned} \tag{37}$$

Since the binary entropy is monotonically increased over $[0, 0.5]$ and monotonically decreased over $[0.5, 1]$, if $1 - a_{22}(r) > 1 - v$ or $a_{22}(r) < v$, then:

$$H(Z) = H(p(Z=0)) < H(1-v) = 1 - H(\frac{2-\delta}{2}). \tag{38}$$

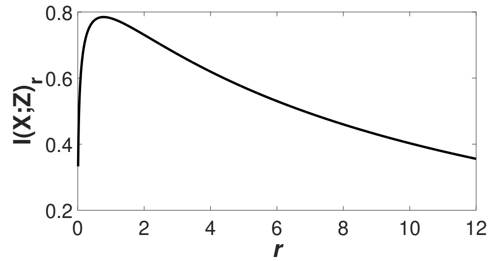

Figure 3: Mutual information $I(X;Z)_r$ as a function of r.

From (34) and (38), a quantizer that produces $a_{22}(r) < v$ is not the optimal one. Therefore, $a_{22}(r^*) \geq v$. A similar proof can be constructed to show that $a_{11}(r^*) \geq v$.

Next, due to $\delta > 1$ (Lemma 1-(3)), we have $0 < 1 - H(\frac{2-\delta}{2}) \leq 1$. Therefore, there exists $v \in (0, 1/2]$ that satisfies (31). From Lemma 1, there exists two positive numbers r_1 and r_2 such that $a_{11}(r_1) = v$ and $a_{22}(r_2) = v$. Moreover, $a_{11}(r)$ and $a_{22}(r)$ are monotonic decreasing and increasing functions, respectively (Lemma 1), thus $r^* \in [r_2, r_1]$. □

Exhaustive search. The proposed algorithm is to search for r in the range of $[r_2, r_1]$. Since $a_{11}(r)$ and $a_{22}(r)$ are monotonic decreasing and increasing functions, finding r_1 and r_2 such that $a_{11}(r_1) = v$ and $a_{22}(r_2) = v$ can be performed efficiently using existing root-finding algorithms, for example, the bisection method. For each value of r in the range $[r_2, r_1]$, we determine the channel matrix $\mathbf{A}$ then use (16) to compute the corresponding capacity. The algorithm outputs the largest mutual information in this range together with r^*. From r^*, Q_{r^*} can be found. Next, based on [3], p_0^* can be obtained as:

$$\begin{aligned} p_0^* &= a_{21}(r^*)[a_{21}(r^*) - a_{11}(r^*)]^{-1} \\ &- [a_{21}(r^*) - a_{11}(r^*)]^{-1} \Big[1 + 2^{\frac{H(a_{21}(r^*)) - H(a_{11}(r^*))}{a_{21}(r^*) - a_{11}(r^*)}}\Big]^{-1}, \end{aligned}$$

where $H(x) = -[x \log_2(x) + (1-x)\log_2(1-x)]$ is the binary entropy function.

V. Numerical Results

Consider a channel having $\phi_0(y) = N(\mu_0 = -1, \sigma_0 = 0.5)$ and $\phi_1(y) = N(\mu_1 = 1, \sigma_1 = 0.6)$. One wants to find the optimal quantizer together with the input distribution such that the mutual information is maximized.

Now, for $r = 1$, $\delta = 1.9299$, $v = 0.2316$, and $r_2 = 0.11$, $r_1 = 11.08$. By performing an exhaustive search with the resolution $\epsilon = 0.01$ over $r \in [r_2, r_1]$, the optimal of mutual information is $I(X;Z)^* = 0.7847$ at $r^* = 0.78$. The corresponding channel capacity upper and lower bounds using Theorem 4 are 0.9479 and 0.7787, respectively. Fig. 3 illustrates $I(X;Z)_r$ as a function of r.

VI. Conclusion

We presented both a new lower bound and a new upper bound on the capacity in terms of quantization parameters and the structure of the associated channel matrix for binary quantization channel. Based on these theoretical results, we propose an efficient algorithm for finding the optimal quantizer.

APPENDIX

A. Proof of Lemma 1

Proof for (1) and (2). From the definition in (14), $a_{11}(r)$ represents the quantized bit "0" which is the integral of $\phi_0(y)$ over the set of y such that $r(y) \geq r$. Thus, if r increases, the set of y reduces. Since $\phi_0(y) \geq 0$ and the set of y reduces, $a_{11}(r)$ must decrease.

Moreover, if $r \to 0$, $r(y) \geq r \ \forall y$ then $a_{11}(r) \to 1$. On the other hand, if $r \to +\infty$, then $r(y) < r \ \forall y$ and $a_{11}(r) \to 0$. Thus, $a_{11}(r) \in (0, 1)$.

A similar proof can be constructed for $a_{22}(r)$.

Proof for (3). From the definition of $r(y)$, $\mathbb{H}_r$ and $\bar{\mathbb{H}}_r$, we have $\frac{\phi_0(y)}{\phi_1(y)} \geq r, \forall y \in \mathbb{H}_r$ and $\frac{\phi_0(y)}{\phi_1(y)} < r, \forall y \in \bar{\mathbb{H}}_r$. Next, we consider two possible cases: $r > 1$ and $r \leq 1$. In both cases, we show that $a_{11}(r) + a_{22}(r) > 1$.

- If $r > 1$, $\phi_0(y) > \phi_1(y)$ for $\forall\ y \in \mathbb{H}_r$. Therefore,

$$\begin{aligned} a_{11}(r) + a_{22}(r) &= \int_{y \in \mathbb{H}_r} \phi_0(y)dy + \int_{y \in \bar{\mathbb{H}}_r} \phi_1(y)dy \\ &> \int_{y \in \mathbb{H}_r} \phi_1(y)dy + \int_{y \in \bar{\mathbb{H}}_r} \phi_1(y)dy \\ &= 1. \end{aligned} \tag{39}$$

- If $r \leq 1$, $\phi_1(y) > \phi_0(y)$ for $\forall\ y \in \bar{\mathbb{H}}_r$. Therefore,

$$\begin{aligned} a_{11}(r) + a_{22}(r) &= \int_{y \in \mathbb{H}_r} \phi_0(y)dy + \int_{y \in \bar{\mathbb{H}}_r} \phi_1(y)dy \\ &> \int_{y \in \mathbb{H}_r} \phi_0(y)dy + \int_{y \in \bar{\mathbb{H}}_r} \phi_0(y)dy \\ &= 1. \end{aligned} \tag{40}$$

Combining (39) and (40), $a_{11}(r) + a_{22}(r) > 1$, $\forall r$.

Next, we show that $a_{11}(1) + a_{22}(1) \geq a_{11}(r) + a_{22}(r)$, $\forall r$. Indeed, from the definition of $\mathbb{H}_1$ and $\bar{\mathbb{H}}_1$, $\frac{\phi_0(y)}{\phi_1(y)} \geq 1, \forall y \in \mathbb{H}_1$ and $\frac{\phi_0(y)}{\phi_1(y)} < 1, \forall y \in \bar{\mathbb{H}}_1$. Thus,

$$\phi_0(y) \geq \phi_1(y), \forall y \in \mathbb{H}_1, \tag{41}$$

$$\phi_0(y) < \phi_1(y), \forall y \in \bar{\mathbb{H}}_1. \tag{42}$$

From the definition of $a_{11}(r)$ and $a_{22}(r)$ in (14) and (15),

$$\begin{aligned} a_{11}(r) + a_{22}(r) &= \int_{y \in \mathbb{H}_r} \phi_0(y)dy + \int_{y \in \bar{\mathbb{H}}_r} \phi_1(y)dy \\ &\leq \int_{y \in \mathbb{H}_r} \max\big(\phi_0(y), \phi_1(y)\big)dy \\ &\quad + \int_{y \in \bar{\mathbb{H}}_r} \max\big(\phi_0(y), \phi_1(y)\big)dy \\ &= \int_{y \in \mathbb{H}_r \cup \bar{\mathbb{H}}_r = \mathbb{R}} \max\big(\phi_0(y), \phi_1(y)\big)dy \\ &= \int_{y \in \mathbb{H}_1} \max\big(\phi_0(y), \phi_1(y)\big)dy \\ &\quad + \int_{y \in \bar{\mathbb{H}}_1} \max\big(\phi_0(y), \phi_1(y)\big)dy \\ &= \int_{y \in \mathbb{H}_1} \phi_0(y)dy + \int_{y \in \bar{\mathbb{H}}_1} \phi_1(y)dy \quad (43) \\ &= a_{11}(1) + a_{22}(1) \quad (44) \\ &= \delta, \quad (45) \end{aligned}$$

where (43) due to (41) and (42), (44) and (45) due to the definitions of $a_{11}(r)$, $a_{22}(r)$ and δ. The equality happens if $r = 1$.

REFERENCES

[1] Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[2] Richard Blahut. Computation of channel capacity and rate-distortion functions. *IEEE transactions on Information Theory*, 18(4):460–473, 1972.

[3] R Silverman. On binary channels and their cascades. *IRE Transactions on Information Theory*, 1(3):19–27, 1955.

[4] Thuan Nguyen and Thinh Nguyen. On closed form capacities of discrete memoryless channels. In *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, pages 1–5. IEEE, 2018.

[5] Thuan Nguyen and Thinh Nguyen. On bounds and closed-form expressions for capacities of discrete memoryless channels with invertible positive matrices. *IEEE Transactions on Vehicular Technology*, 69(9):9910–9920, 2020.

[6] Brian M Kurkoski and Hideki Yagi. Quantization of binary-input discrete memoryless channels. *IEEE Transactions on Information Theory*, 60(8):4544–4552, 2014.

[7] Thuan Nguyen and Thinh Nguyen. On thresholding quantizer design for mutual information maximization: Optimal structures and algorithms. In *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, pages 1–5. IEEE, 2020.

[8] Thuan Nguyen and Thinh Nguyen. Structure of optimal quantizer for binary-input continuous-output channels with output constraints. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pages 1450–1455. IEEE, 2020.

[9] Thuan Nguyen, Yu-Jung Chu, and Thinh Nguyen. A new fast algorithm for finding capacity of discrete memoryless thresholding channels. In *2020 International Conference on Computing, Networking and Communications (ICNC)*, pages 56–60. IEEE, 2020.

[10] Thuan Nguyen and Thinh Nguyen. On binary quantizer for maximizing mutual information. *IEEE Transactions on Communications*, 68(9):5435–5445, 2020.

[11] Rudolf Mathar and Meik Dörpinghaus. Threshold optimization for capacity-achieving discrete input one-bit output quantization. In *2013 IEEE International Symposium on Information Theory Proceedings (ISIT)*, pages 1999–2003. IEEE, 2013.

[12] Thuan Nguyen, Yu-Jung Chu, and Thinh Nguyen. On the capacities of discrete memoryless thresholding channels. In *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, pages 1–5. IEEE, 2018.

[13] Jaspreet Singh, Onkar Dabeer, and Upamanyu Madhow. On the limits of communication with low-precision analog-to-digital conversion at the receiver. *IEEE Transactions on Communications*, 57(12):3629–3639, 2009.

[14] Brian M Kurkoski and Hideki Yagi. Finding the capacity of a quantized binary-input dmc. In *2012 IEEE International Symposium on Information Theory Proceedings*, pages 686–690. IEEE, 2012.

[15] Brian M Kurkoski and Hideki Yagi. Single-bit quantization of binary-input, continuous-output channels. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 2088–2092. IEEE, 2017.

[16] Minh N Vu, Nghi H Tran, Dissanayakage G Wijeratne, Khanh Pham, Kye-Shin Lee, and Duy HN Nguyen. Optimal signaling schemes and capacity of non-coherent rician fading channels with low-resolution output quantization. *IEEE Transactions on Wireless Communications*, 18(6):2989–3004, 2019.

[17] M. Ranjbar, N. H. Tran, M. N. Vu, T. V. Nguyen, and M. Cenk Gursoy. Capacity region and capacity-achieving signaling schemes for 1-bit adc multiple access channels in rayleigh fading. *IEEE Transactions on Wireless Communications*, 19(9):6162–6178, 2020.

[18] David Burshtein, Vincent Della Pietra, Dimitri Kanevsky, and Arthur Nadas. Minimum impurity partitions. *The Annals of Statistics*, pages 1637–1646, 1992.

[19] Mung Chiang and Stephen Boyd. Geometric programming duals of channel capacity and rate distortion. *IEEE Transactions on Information Theory*, 50(2):245–258, 2004.