On the Discreteness of Capacity-Achieving Distributions for the Censored Channel

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

Abstract—**The censored channel is one of the fundamental channels in information theory, which belongs to the class of non-linear channels. It is modeled by cascading an additive noise channel with a clipping operator. This paper is concerned with the information theoretic capacity of this channel. A necessary and sufficient condition for optimality of the input distribution is derived and it is shown that the capacity-achieving input distribution for the amplitude-limited censored channel has only a finite number of mass points. This result holds for a large class of noise distributions including additive Gaussian noise.**

I. Introduction

It is well-known that the information theoretic capacity of additive white Gaussian noise channels is attained by a discrete input distribution under amplitude constraints [1]. The extension of this statement to other channels is not straightforward, since fading and non-linear effects hamper some of the mathematical derivations and proof techniques. Especially in presence of censoring and truncation of information, even the determination of mutual information requires sophisticated mathematical techniques [2]. Many relevant communication channels involve non-linear operations, e.g., output quantization [3]–[5] or hinge channel [6]. One can refer for instance to clipping operation resulting from saturation effects of amplifiers. Unlike additive Gaussian channels, the received signal cannot have an arbitrary large amplitude in general and the output of the noisy channel can cause saturation of amplifiers in the receiver chain from above and below. Therefore there are two clipping thresholds, restricting the received signal to an interval. This can be modeled by cascading a clipping operator to a usual additive noisy channel. This model, called censored channel, can be considered as a more realistic model for this type of noisy communication. For the censored channel the mutual information along with an upper bound on the unknown capacity are derived in [7]. In general, no constraint is assumed on the input of censored channel. However, in [8] it is shown that for a very large class of noise distributions, it is possible to consider inputs supported on a bounded interval and at the same time to obtain sharp estimates for the channel capacity. The continuity of mutual information and existence of optimal input distribution have also been discussed in [8]. The question about the discreteness and finiteness of the optimal input-distribution has remained an open problem.

Determining the capacity of a channel is nothing but optimizing the mutual information over a set of probability measures. Smith was the first to consider the capacity of peak and average power-constrained Gaussian channels from this point of view [1]. His method consists of first discussing existence of an optimal input-distribution by proving the continuity of mutual information and compactness of the corresponding set of probability measures. He introduced the notion of weak derivation for mutual information to provide necessary and sufficient condition for optimality of an input distribution. Finally, tools from complex analysis such as analytic extension and the identity theorem have been used to study the structure of input distribution. The problem of characterizing the capacity-achieving distribution has been considered later by many authors. Shamai and Bar-David [9] proved the finiteness of support of the input distribution for peak and average-power limited quadrature Gaussian channels. General conditions for discreteness of the optimal distribution are also discussed in [10] for non-Gaussian noise. The author in [11] proved the finiteness of the support of the optimal distribution for amplitude limited channels under some general assumptions about the noise. Conditionally Gaussian channels are discussed in [12] which includes also rigorous proofs of used techniques. The authors in [13] proved the discreteness of the optimal distribution with finite points for Rayleigh fading channels.

In the present paper we show the discreteness of the optimal input-distribution and prove, that for a wide range of noise distributions the number of mass points is finite under amplitude constraint. The proof techniques are similar to [1]. Like in [11], our results are proven for a wide-class of noise distributions instead of particular ones. After introducing the channel model in Section II, some of the previous results, for instance on continuity of mutual information, are discussed in Section III. The main theorem is stated in Section IV and the sketch of the proof is given by using a series of lemmas. These lemmas are proven later in Section V.

II. System Model

The censored channel is depicted in Figure 1. The input random variable X is assumed to be real-valued and is disturbed by additive random noise W. Both X and W are assumed to be stochastically independent and time-discrete. The noisy signal $Z = X + W$ is then censored below 0 and above 1 while in between the transition is linear. The output will be the random variable Y. Note that any other censored channel, with a different slope in the linear region and different censoring levels, can be transformed to the specific one described in the present paper, cf. [8].

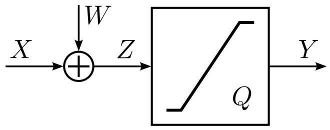

Fig. 1. The system model: a real input X is subject to additive noise W and then is censored at 0 and 1 to yield Y.

In [7] we have presented the mutual information $I_{X;Y}(F) = h_Y(F) - h_{Y|X}(F)$ of the censored channel, as a function of the input-distribution $F(x)$, for a given continuous noise distribution $\Phi(w)$ with density $\phi(w)$ (not necessarily Gaussian), by deducing the output entropy

$$\begin{aligned} h_Y(F) = \rho\Big(\int\int_{-\infty}^{0} \phi(u-x)\,\mathrm{d}u\,\mathrm{d}F(x)\Big) \\ + \rho\Big(\int\int_{1}^{\infty} \phi(u-x)\,\mathrm{d}u\,\mathrm{d}F(x)\Big) \\ + \int_0^1 \rho\Big(\int \phi(u-x)\,\mathrm{d}F(x)\Big)\,\mathrm{d}u \end{aligned} \tag{1}$$

and the conditional entropy

$$\begin{aligned} h_{Y|X}(F) = \int \rho\Big(\int_{-\infty}^{0} \phi(u-x)\,\mathrm{d}u\Big)\,\mathrm{d}F(x) \\ + \int \rho\Big(\int_{1}^{\infty} \phi(u-x)\,\mathrm{d}u\Big)\,\mathrm{d}F(x) \\ + \int_0^1 \int \rho\big(\phi(u-x)\big)\,\mathrm{d}F(x)\,\mathrm{d}u\,. \end{aligned} \tag{2}$$

Note that the weighted self-information $\rho(q) = -q\log q$, $q \geq 0$, is a strictly concave function of its argument q and the logarithm is of general base. Because of the concavity of ρ, mutual information is a concave function w.r.t. the input-distribution F.

For the sake of compactness, we use the short forms

$$\ell(x) = \int_{-\infty}^{-x} \phi(w)\,\mathrm{d}w\,, \qquad L(F) = \int \ell(x)\,\mathrm{d}F(x)\,,$$
$$r(x) = \int_{1-x}^{\infty} \phi(w)\,\mathrm{d}w\,, \qquad R(F) = \int r(x)\,\mathrm{d}F(x)\,,$$

and $\alpha(u;F) = \int \phi(u-x)\,\mathrm{d}F(x)$ to enable a slightly different representation of the entropies $h_Y(F)$ and $h_{Y|X}(F)$ as

$$\begin{aligned} h_Y(F) = &- L(F)\log L(F) - R(F)\log R(F) \\ &- \int_0^1 \alpha(y;F)\log\alpha(y;F)\,\mathrm{d}y \end{aligned} \tag{3}$$

and

$$\begin{aligned} h_{Y|X}(F) = &- \int \ell(x)\log\ell(x)\,\mathrm{d}F(x) \\ &- \int r(x)\log r(x)\,\mathrm{d}F(x) \\ &- \int \Big(\int_{-x}^{1-x} \phi(w)\log\phi(w)\,\mathrm{d}w\Big)\,\mathrm{d}F(x), \end{aligned} \tag{4}$$

respectively.

III. Continuity and Existence of the Solution

In this section, we review some of the relevant results in the context of the censored channel. The following proposition from [8] establishes the continuity of mutual information.

Proposition 1 ([8, Prop. 3])**.** Assuming an absolutely continuous noise distribution with bounded density, the mutual information $I_{X;Y}(F)$ is continuous in $F \in \Omega$.

The proof of continuity in F is based on showing that for each sequence of probability distributions $F_n \overset{w^*}{\to} F$, convergence in weak* topology, the function $I_{X;Y}(F_n)$ converges to $I_{X;Y}(F)$. The proof utilizes dominated convergence theorem and properties of convergence of measures.

The existence of optimal distribution is guaranteed if it can be shown that Ω is weak* compact. If one of the moments of the input-distribution is bounded, the Markov inequality can be used to show that the set Ω is tight. Tightness is a necessary condition for the relative compactness[1] of Ω using Prokhorov's theorem [14]. In that case, the sequential compactness of Ω can be shown by proving that the limits of sequences belong to Ω. Finally, compactness would follow by the metrizability of the weak* topology by Lévy metric. In the above model, no constraint was assumed on the input and therefore tightness will not hold in general for Ω.

One possibility to circumvent this problem is to show that it is enough to consider only a compact subset of $\mathbb{R}$ to find the capacity or at least approximate it. This has been discussed in [8]. If the noise has a density with bounded support, a compact set can indeed be found for this purpose.

Proposition 2 ([8, Prop. 4])**.** If the support of the noise W lies in the set $[-K, K]$ for some $K \in \mathbb{R}$, then for every choice of the input-distribution F, there is an input-distribution $\tilde{F}$ such that $I_{X;Y}(F) = I_{X;Y}(\tilde{F})$ and $\tilde{F}$ is supported on $[-K-\delta, K+1+\delta]$ for an arbitrary $\delta > 0$ independent of F.

Consequently one can limit the search for optimal input distribution to a compact set. In this case, the optimal input distribution exists but evidently it will not be unique. In general, common noises like Gaussian noise do not have density with bounded support. But it is known that for each probability measure $\mathbb{P}$ on $\mathbb{R}$ and for each $\epsilon > 0$, there exists a number $K > 0$ such that $\mathbb{P}([-K, K]) > 1 - \epsilon$. In other words, each probability measure on $\mathbb{R}$ is tight. This is a simple conclusion of [14, Thm. 1.3] by using the fact that $\mathbb{R}$ is complete and separable under standard topology. The following proposition shows that considering bounded support can still be useful to approximate the capacity.

Proposition 3 ([8, Cor. 8])**.** For the noise density $\phi(x)$ suppose that $\int_{-K}^{K} \phi(x)\mathrm{d}x \geq 1-\epsilon$. If the noise has a probability density function with bounded derivative on $\mathbb{R}$, there exists an input-distribution $\tilde{F}$, which maximizes the mutual information on the bounded support $[-K-\delta, K+1+\delta]$ for $\delta > 0$ and approaches the capacity of the censored channel within the $O(\epsilon^{1/4})$ gap.

[1] Relative compactness means that every sequence of probability distributions F_n has a sub-sequence converging to a probability distribution F.

The bound on the absolute value of the derivative of ϕ is used to show that the noise probability density function $\phi(w)$ over $(-\infty, -K) \cup (K, \infty)$ is bounded by a function of ϵ. Note that the capacity gap is only a function of ϵ, which is in turn a function of K. In other words, the gap can be fixed a priori regardless of optimal input distributions by choosing K from the noise probability density function. These results justify considering an amplitude limited channel since it can closely approximate the capacity of censored channel without input constraints. For the rest of this paper, the censored channel is assumed to be with amplitude constraint. Therefore the set Ω of probability distributions with amplitude constraint is compact and the optimal distribution exists [8].

IV. On Finiteness of Input Distribution

As a result of the previous section, suppose that input distributions are supported on $[-A, A]$ for a choice of $A > 0$. The main result of this paper is that an optimal input distribution has finite number of mass points supported on $[-A, A]$.

Theorem 1. Let the noise distribution be an absolutely continuous measure with a continuous bounded and strictly positive density $\phi(x)$ with the following additional properties:

1) $\phi(x)$ has a nowhere zero analytic extension on the set $\mathcal{D}_\delta \triangleq \{z \in \mathbb{C} : |\mathrm{Im}(z)| < \delta\}$.
2) There is a nonincreasing function $U : [T, \infty) \to \mathbb{R}^+$ such that $|\phi(z)| \le U(|\mathrm{Re}(z)|)$.
3) The function U is integrable over $[T, \infty)$.

Then the optimal input distribution over the interval $[-A, A]$ has a finite support.

Proof:

The proof is presented through a series of lemmas. The proof sketches of the lemmas are presented in the next section. The proof consists of first finding the necessary and sufficient conditions for optimal input distributions. Next, similar to [1], [11], [12], the optimality condition provides zeros of a certain function, which is proven to have an analytic extension on an open set. These zeros correspond to the points of increase of the optimal distribution. Subsequently it is shown that if the number of zeros, i.e., the number of points of increase of the optimal distribution, over a set, say $[-A, A]$, is infinite and hence the zeros have an accumulation point, the identity theorem [15, Thm. 4.8] shows that the function should be zero over the whole open set. Finally it is shown that if this function is totally zero over the open set, a contradiction is found and therefore the input distribution must have a finite number of points of increase over $[-A, A]$, due to Bolzano-Weierstrass theorem.

Lemma 1. Let $a(x)$ be defined as

$$a(x) = \ell(x) \log \tfrac{\ell(x)}{L(F_0)} + r(x) \log \tfrac{r(x)}{R(F_0)} + \int_0^1 \phi(y - x) \log \tfrac{\phi(y-x)}{\alpha(y; F_0)} \, \mathrm{d}y \,. \tag{5}$$

For the censored channel, if F_0 is a capacity achieving input distribution with $I_{X;Y}(F_0) = C$, then $a(x) \le C$ for $x \in [-A, A]$. Equality holds at all points of increase of F_0.

The above lemma is obtained by applying necessary and sufficient conditions for optimality of an input distribution based on weak derivative of mutual information and will be discussed in the next section. Note that the function $a(x)$ is equal to C for all points of increase of F_0.

Lemma 2. Suppose that the noise density $\phi(x)$ satisfies the conditions (1), (2) and (3) of Theorem 1. Then $a(z)$, $z \in \mathbb{C}$ is analytic over $\mathcal{D}_\delta$.

The reason behind choosing $\mathcal{D}_\delta$, similar to [11], is to assure that $\phi(z)$ remains analytic under translation along the real line.

Lemma 3. If the points of increase of F_0 are infinite over $[-A, A]$, then $a(x) = C$ for all $x \in \mathbb{R}$.

In the above lemma, the points of increase, if infinite, have an accumulation point in $[-A, A]$. But the points of increase are zeros of $a(z) - C$ and if the zero set of $a(z) - C$ on a compact set admits an accumulation point, then according to the identity theorem [16, Theorem 10.18] and [15, Theorem 4.8], $a(z) - C$ should be constant zero over the whole real line, i.e., $a(x) = C$ for $x \in \mathbb{R}$. We show that this leads to a contradiction and therefore the input distribution must have a finite number of points of increase over $[-A, A]$.

Suppose that $a(x) = C$ for $x \in \mathbb{R}$. Therefore $\lim_{x\to\infty} a(x) = C$. We have

$$\begin{aligned} &\lim_{x\to\infty} \Bigg[\ell(x) \log \tfrac{\ell(x)}{L(F_0)} + r(x) \log \tfrac{r(x)}{R(F_0)} \\ &\qquad + \int_0^1 \phi(y - x) \log \tfrac{\phi(y-x)}{\alpha(y; F_0)} \, \mathrm{d}y \Bigg] \\ &\overset{(a)}{=} \log \tfrac{1}{R(F_0)} + \lim_{x\to\infty} \int_0^1 \phi(y - x) \log \tfrac{\phi(y-x)}{\alpha(y; F_0)} \, \mathrm{d}y \\ &\overset{(b)}{=} \log \tfrac{1}{R(F_0)}, \end{aligned}$$

where the step (a) follows from $\lim_{x\to\infty} \ell(x) = 0$ and $\lim_{x\to\infty} r(x) = 1$. The step (b) follows from the dominated convergence theorem by showing that the expression inside the integral has bounded absolute value. This can be done using the fact that $\phi(y - x)$ and $\alpha(y; F_0)$, which is the convolution of F_0 and $\phi(x)$, are nonzero and bounded over $[0, 1]$. After taking the limit inside the integration, it can be seen that $\lim_{x\to\infty} \phi(y - x) = 0$ from (2) and (3) and therefore the whole limit is zero. The above limit shows that $\log \frac{1}{R(F_0)} = C$. Using the same argument for the case $x \to -\infty$, we have $\lim_{x\to-\infty} a(x) = \log \frac{1}{L(F_0)} = C$. Therefore $L(F_0) = R(F_0)$. Using this result, it can be seen that

$$\begin{aligned} C = a(x) &= \ell(x) \log \ell(x) + r(x) \log r(x) \\ &\quad + (\ell(x) + r(x)) \log \tfrac{1}{L(F_0)} + \int_0^1 \phi(y - x) \log \tfrac{\phi(y-x)}{\alpha(y; F_0)} \, \mathrm{d}y \\ &< C + \int_0^1 \phi(y - x) \log \tfrac{\phi(y-x)}{\alpha(y; F_0)} \, \mathrm{d}y \end{aligned}$$

where the inequality follows from $r(x), \ell(x), \ell(x) + r(x) \in (0, 1)$. Therefore the integral on the right hand side should be always positive . However since $\alpha(y;F_0)$ has a non-zero minimum over $[0,1]$, for sufficiently large x, $\phi(y-x)<\alpha(y;F_0)$ for $y\in[0,1]$ and hence the integral would be negative and hence we have a contradiction. ■

Note that the conditions of Theorem 1 are satisfied by variety of distributions including Gaussian distribution. Moreover the amplitude limited assumption is only needed in two places, first for proving compactness of the set of input distributions and second for showing that there exists an accumulation point in $[-A,A]$.

V. Sketch of Proofs

A. Proof of Lemma 1

The notion of weak derivative was introduced in [1] which can be used to derive both the necessary and sufficient conditions for optimality of an input distribution.

Definition 1 (Weak Derivative)**.** Let $f(\mu):\Omega\to\mathbb{R}$ be a real valued function on the space of probability measures. The function f is weakly differentiable at the point μ_0 along the line between μ_0 and μ_1, if the limit

$$D_{\mu_1}f(\mu_0)=\lim_{\theta\to 0}\frac{f((1-\theta)\mu_0+\theta\mu_1)-f(\mu_0)}{\theta}.$$

exists. The weak derivative of f is denoted by $D_{\mu_1}f(\mu_0)$.

The following proposition provides some useful properties of the weak derivative. The proofs are standard.

Proposition 4. The following properties hold for the weak derivative $D_{\mu_1}f(\mu_0)$ defined above.

1) $D_{\mu_1}(g\cdot f)(\mu_0)=D_{\mu_1}g(\mu_0)\cdot f(\mu_0)+g(\mu_0)\cdot D_{\mu_1}f(\mu_0)$
2) $D_{\mu_1}(g\circ f)(\mu_0)=D_{\mu_1}f(\mu_0)\cdot g'(f(\mu_0))$

Finally the relation between weak derivative and optimality is given in the following proposition.

Proposition 5. Let $f(\mu)$ be a concave function over a subset Ω of the space of probability measures. The probability measure μ maximizes $f(\mu)$, if and only if, for any probability measure $\nu\in\Omega$, the inequality $D_\nu f(\mu)\le 0$ holds.

The proof is omitted here and can be found in [12, p. 2076]. Note that the mutual information is a concave function of the input-distribution and therefore satisfies the condition of the above proposition.

It remains to find the weak derivative of the mutual information of censored channels. First of all, since $h_{Y|X}(F)$ is a linear function of F, it can be seen that

$$\begin{aligned}D_{F_1}h_{Y|X}(F_0)&=\lim_{\theta\to 0}\frac{h_{Y|X}((1-\theta)F_0+\theta F_1)-h_{Y|X}(F_0)}{\theta}\\&=h_{Y|X}(F_1)-h_{Y|X}(F_0).\end{aligned}$$

Now consider $h_Y(F)$ which consists of three terms. Consider the first and the second term, namely $-L(F)\log L(F)$ and $-R(F)\log R(F)$. $L(F)$ and $R(F)$ are linear functions of F. Using Proposition 4, we get

$$D_{F_1}(-L(F_0)\log L(F_0))=-(L(F_1)-L(F_0))\log eL(F_0)$$

and

$$D_{F_1}(-R(F_0)\log R(F_0))=-(R(F_1)-R(F_0))\log eR(F_0).$$

Let us consider the last term $-\int_0^1\alpha(y;F)\log\alpha(y;F)\,\mathrm{d}y$ with its weak derivative

$$\begin{aligned}D_{F_1}&\int_0^1\alpha(y;F_0)\log\alpha(y;F_0)\,\mathrm{d}y\\&=-\int_0^1\alpha(y;F_0)\log\alpha(y;F_0)\,\mathrm{d}y\\&+\lim_{\theta\to 0}\int_0^1\alpha(y;F_1)\log\alpha(y;F_\theta)\,\mathrm{d}y\\&+\lim_{\theta\to 0}\frac{1-\theta}{\theta}\int_0^1\alpha(y;F_0)\log\frac{\alpha(y;F_\theta)}{\alpha(y;F_0)}\,\mathrm{d}y.\end{aligned}$$

Note that given the conditions of Theorem 1, the function $\alpha(y;F)$ is continuous for each F and bounded by $0<m_F\le\alpha(y;F)\le M_F$ over the compact set $[0,1]$ for some positive numbers m_F and M_F. Therefore $m_{F_1}\log\min(m_{F_1},m_{F_0})\le\alpha(y;F_1)\log\alpha(y;F_\theta)\le M_{F_1}\log\max(M_{F_1},M_{F_0})$. Using dominated convergence theorem, we deduce

$$\lim_{\theta\to 0}\int_0^1\alpha(y;F_1)\log\alpha(y;F_\theta)\,\mathrm{d}y=\int_0^1\alpha(y;F_1)\log\alpha(y;F_0)\,\mathrm{d}y.$$

Using a similar argument, the term inside the last integral is bounded by finite values and hence the limit can be taken inside the integral in the last term. Using l'Hôpital's rule, the last term can be simplified as

$$\begin{aligned}\lim_{\theta\to 0}&\frac{1-\theta}{\theta}\int_0^1\alpha(y;F_0)\log\frac{\alpha(y;F_\theta)}{\alpha(y;F_0)}\,\mathrm{d}y\\&=\int_0^1\alpha(y;F_1)\,\mathrm{d}y-\int_0^1\alpha(y;F_0)\,\mathrm{d}y.\end{aligned}\tag{6}$$

Therefore the weak derivative of $h_Y(F)$ writes as

$$\begin{aligned}D_{F_1}h_Y(F_0)=&-(L(F_1)-L(F_0))\log eL(F_0)\\&-(R(F_1)-R(F_0))\log eR(F_0)\\&+\int_0^1\alpha(y;F_0)\,\mathrm{d}y-\int_0^1\alpha(y;F_1)\,\mathrm{d}y\\&+\int_0^1\alpha(y;F_0)\log\alpha(y;F_0)\,\mathrm{d}y\\&-\int_0^1\alpha(y;F_1)\log\alpha(y;F_0)\,\mathrm{d}y\end{aligned}$$

Using Proposition 5, if $I_{X;Y}(F_0)=C$, then $D_{F_1}I_{X;Y}(F_0)\le 0$ for all F_1. Pick a step function for F_1 at x and the expression of Lemma 1 follow using standard manipulations. Finally the fact that $D_\nu f(\mu)=0$ for points of increase of F_0, follows from [1, Cor. 1].

B. Proof of Lemma 2

In [11], the analyticity of the terms in mutual information is proved using Morera's theorem. The following lemma can also be proven using a similar argument.

Lemma 4 ([15, Thm. 5.4])**.** Let $F(z,y)$ be defined over $\Omega\times[0,1]$ where Ω is an open set in $\mathbb{C}$. If F is analytic in z for each

y and is continuous on $\Omega \times [0,1]$, then the function $f(z) = \int_0^1 F(z,y)\mathrm{d}y$ is analytic.

An important issue is the choice of $\mathcal{D}_\delta$. For this choice, if $\phi(z)$ is analytic on $\mathcal{D}_\delta$, then any shifted version of the function along the real axis, $\phi(y-z)$, is also analytic. To show $a(z)$ is analytic, we show that each term is analytic. Since $\phi(z)$ is analytic and nowhere zero, $\phi(y-z)\log\phi(y-z)$ is also analytic in z and continuous on $y \in [0,1]$. Therefore according to Lemma 4, the integral $\int_0^1 \phi(y-z)\log\phi(y-z)\mathrm{d}y$ is also analytic. Next consider that, since $\phi(z)$ is nowhere zero and analytic, the logarithm of the output density $\log\alpha(y;F_0)$ is also continuous. Therefore $\phi(y-z)\log\alpha(y;F_0)$ is continuous in y and analytic in z. Hence $\int_0^1 \phi(y-z)\log\alpha(y;F_0)\,\mathrm{d}y$ is analytic according to Lemma 4. It remains to show that $\ell(z)$ and $r(z)$ are analytic and nowhere zero. It is enough to prove this only for

$$\ell(z) = \int_{-\infty}^{0} \phi(y-z)\mathrm{d}y.$$

Since $\phi(z)$ is nowhere zero, so is $\ell(z)$. We show that $\ell(z)$ is analytic using Morera's theorem. Suppose that $z_k \to z$ and $\overline{\mathcal{B}(z,\epsilon)}$ is a closed ball such that for a $N > 0$, it contains z_k's for $k > N$ and lies inside $\mathcal{D}_\delta$. It yields

$$\begin{aligned}
&\int_{-\infty}^{0} |\phi(y-z_k)|\mathrm{d}y \le \int_{-\infty}^{0} \max_{\xi\in\overline{\mathcal{B}(z,\epsilon)}} |\phi(y-\xi)|\mathrm{d}y \\
&= \int_{-(T+|z|+\epsilon)<y<0} \max_{\xi\in\overline{\mathcal{B}(z,\epsilon)}} |\phi(y-\xi)|\mathrm{d}y \\
&\qquad + \int_{-\infty<y\le-(T+|z|+\epsilon)} \max_{\xi\in\overline{\mathcal{B}(z,\epsilon)}} |\phi(y-\xi)|\mathrm{d}y \\
&\le M_0 - (T+|z|+\epsilon) \\
&\qquad + \int_{-\infty<y\le-(T+|z|+\epsilon)} \max_{\xi\in\overline{\mathcal{B}(z,\epsilon)}} U(|y-\mathrm{Re}(\xi)|)\mathrm{d}y \\
&\le M_0 - (T+|z|+\epsilon) \\
&\qquad + \int_{-\infty<y\le-(T+|z|+\epsilon)} U(|y|-|z|-\epsilon)\mathrm{d}y < \infty.
\end{aligned}$$

The last step follows from the assumption of integrability of U. Using dominated convergence theorem, we have

$$\lim_{z_k\to z} \ell(z_k) = \ell(z),$$

which implies the continuity of $\ell(z)$. To use Morera's theorem, it should be shown that for all triangles Δ with the perimeter length $|\Delta|$, we have:

$$\int_{\partial\Delta} \ell(z)\mathrm{d}z = 0.$$

Note that similar to the argument above, $\int_{\partial\Delta} |\ell(z)|\mathrm{d}z < \infty$. Now we have:

$$\int_{\partial\Delta}\int_{-\infty}^{0} \phi(y-z)\,\mathrm{d}y\mathrm{d}z = \int_{-\infty}^{0}\int_{\partial\Delta} \phi(y-z)\,\mathrm{d}z\mathrm{d}y = 0,$$

where the order of integrals are changed using Fubini's theorem and analyticity of $\phi(y-z)$ implies $\int_{\partial\Delta}\phi(y-z)\,\mathrm{d}z = 0$. Therefore $\ell(z)$ is analytic using Morera's theorem.

VI. Conclusions and Future Works

In this work, it has been proven that the optimal input distribution for the censored channel with amplitude constraint has finite support. The proof is based on using weak derivative to derive necessary and sufficient condition for optimality. This condition provides the zero set of a function $a(x)-C$ at points of increase of the optimal distribution. After extending the function $a(x)-C$ to an analytic function, the identity theorem implies that $a(x) = C$ for all $x \in \mathbb{R}$, if the zero set is infinite. It is shown that it leads to a contradiction and hence the zero set should be finite. Future works will involve numerical evaluations of different input distributions of finite support and optimizations of their parameters to approach the capacity.

References

[1] J. G. Smith, "The information capacity of amplitude- and variance-constrained sclar gaussian channels," *Information and Control*, vol. 18, no. 3, pp. 203 – 219, 1971.

[2] M. Pinsker, *Information and Information Stability of Random Variables and Processes*. San Francisco, California, USA: Holden-Day, 1964.

[3] G. Alirezaei and R. Mathar, "Optimum one-bit quantization," in *The IEEE Information Theory Workshop - Fall (ITW'15)*, Jeju Island, Korea, Oct. 2015, pp. 357–361.

[4] ——, "Optimal one-bit quantizers are asymmetric for additive uniform noise," in *Information Theory and Applications Workshop (ITA'17)*, San Diego, California, USA, Feb. 2017.

[5] Y. Wu, L. M. Davis, and R. Calderbank, "On the capacity of the discrete-time channel with uniform output quantization," in *2009 IEEE International Symposium on Information Theory*, June 2009, pp. 2194–2198.

[6] G. Alirezaei and R. Mathar, "On the information capacity of hinge functions," in *2016 International Symposium on Information Theory and its Applications (ISITA'16)*, Monterey, California, USA, Oct. 2016, pp. 438–442.

[7] ——, "An upper bound on the capacity of censored channels," in *The 9th International Conference on Signal Processing and Communication Systems (ICSPCS'15)*, Cairns (Barrier Reef), Australia, Dec. 2015.

[8] A. Behboodi, G. Alirezaei, and R. Mathar, "On the capacity of censored channels," in *11th International ITG Conference on Systems, Communications and Coding (SCC'17)*, Hamburg, Germany, Feb. 2017.

[9] S. Shamai and I. Bar-David, "The capacity of average and peak-power-limited quadrature gaussian channels," *Information Theory, IEEE Transactions on*, vol. 41, no. 4, pp. 1060–1071, Jul 1995.

[10] A. Das, "Capacity-achieving distributions for non-gaussian additive noise channels," in *Information Theory, 2000. Proceedings. IEEE International Symposium on*, 2000, pp. 432–.

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

[12] 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, Jun. 2005.

[13] I. Abou-Faycal, M. Trott, and S. Shamai, "The capacity of discrete-time memoryless rayleigh-fading channels," *Information Theory, IEEE Transactions on*, vol. 47, no. 4, pp. 1290–1301, May 2001.

[14] P. Billingsley, *Convergence of Probability Measures*. Wiley, 1999.

[15] E. M. Stein and R. Shakarchi, *Complex analysis*, ser. Princeton lectures in analysis. Princeton, N.J: Princeton University Press, 2003, no. 2.

[16] W. Rudin, *Real and complex analysis*, 3rd ed. New York: McGraw-Hill, 1987.