On the Capacity of Censored Channels

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 the cascade of an additive noise channel with a clipping operator that restricts the signal to the interval $[0, 1]$. In this paper, after discussing the generality of the considered model, the mutual information of the censored channel is proven to be a continuous function of the input probability distribution whenever the probability density of the noise is bounded. Moreover, there is an input-distribution of bounded support that achieves capacity for a class of noise distributions with bounded support. For a more general class of noise distributions, the capacity can be approximated arbitrarily close using only input-distributions of bounded support. This support is determined only based on the noise distribution and the desired approximation error. These results constitute a first step for characterization of the capacity of censored channels.**

I. Introduction

Many practical communication channels involve non-linear operations. A common example is the clipping operation which in practice is the result of saturation effects, e.g., by stimulation of amplifiers with high input power. Unlike additive white Gaussian noise (AWGN) channels with power constraint, with well-known capacity characterization, these non-standard channels involve additional difficulties. It is in general unclear whether the capacity can be achieved by an input-distribution and whether it is unique. The structure of the optimal distribution is also unknown in general. It turns out that for some cases of noise distributions, the optimal input-distribution is discrete with bounded support. Despite these difficulties, the censored channel can be considered as a more realistic model for noisy communication. The output of the noisy channel can cause saturation of amplifiers in the receiver chain from above and below. Therefore there are two cutting points that restrict the value of the received signal to an interval. This can be modeled by cascading a clipping operator to a usual additive noisy channel. In this paper, we focus on these type of channels while assuming general continuous noise distributions. The censored channel was introduced in [1], where the expression of the mutual information and an upper bound for the capacity have been derived. Indeed, even if no restriction is put on the input, like power or energy constraints, the capacity is still finite. Nevertheless, the structure of an optimal input-distribution as well as its existence and uniqueness are highly non-trivial. The characterization of the capacity is indeed an optimization problem which makes the investigation of information theoretical aspects hard. The optimization is usually performed over the space of probability measures, which is an infinite dimensional metric space. A metric on such spaces can be very intricate while some useful properties, like compactness

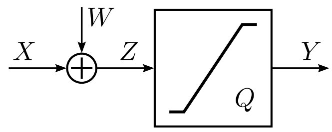

Fig. 1: The system model: some real input X is subject to additive noise W and is censored at 0 and 1 to yield Y.

of sets, are established without using a metric. The familiar relations between linearity, convexity and continuity are also lost. On the other hand, the objective function, i.e., the mutual information, is described by a complicated integral which involves the input probability measure in a specific way. This is in general a challenging problem with inheriting difficulties from calculus of variations and geometric measure theory.

Probably, the research on non-standard channels dates back to the paper by Smith [2], where he has considered the capacity of peak and average power-constrained Gaussian channels. He has discussed both the existence and uniqueness of an optimal input-distribution, and derived some features for optimizing the input. The notion of weak derivation for mutual information was also discussed therein. Shamai and Bar-David later worked on a similar problem [3]. General conditions for discreteness of optimal distribution is also discussed in [4] and [5] for average and peak power constraints. Conditionally Gaussian channels are discussed in [6], although the authors in [7] prove the discreteness of optimal distribution with finite points for Rayleigh fading channels. Another non-standard channel is the one-bit quantizer, which is investigated in [8].

Following the model in [1], the censored channel is again considered in the present paper without any restriction on the input power or energy. By removing restrictions on the input, the set of acceptable input-distributions is not compact anymore and therefore the existence of a capacity-achieving input-distribution is not guaranteed. However, subject to some constraints on the noise distribution, it is possible w.l.o.g. to limit the probability space to a compact subset of probability distributions to provide the existence of a capacity-achieving input-distribution. After presenting the model and the expression for the corresponding mutual information, the continuity of mutual information is established using similar techniques as in [2]. Next, for a specific class of noise distributions, it is shown that a search for an optimal mutual information can be limited to a compact set in which an optimal input-distribution exists. Finally, we deduce, that input-distributions with bounded support can approach capacity with a negligible error even if the noise distribution has an unbounded support.

II. Preliminaries and Mutual Information

In our previous paper [1] we have investigated the information theoretical properties of a censored channel as 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 at 0 and 1 by the function

$$Q(z) = \begin{cases} 0, & \text{if } z \leq 0, \\ z, & \text{if } 0 < z \leq 1, \\ 1, & \text{if } z > 1 \,. \end{cases} \tag{1}$$

Hence, the output signal is represented as

$$Y = Q(X + W). \tag{2}$$

In general one has to deal with the common channel

$$Q'(z') = \begin{cases} 0, & \text{if } z' \leq 0, \\ \frac{a}{b} z', & \text{if } 0 < z' \leq b, \\ a, & \text{if } z' > b \,, \end{cases}$$

where a and b are arbitrary positive real numbers. In this case the output Y' is given by $Q'(X'+W')$. Normalizing the input and output signals by $X = \frac{X'}{b}$, $W = \frac{W'}{b}$ and $Y = \frac{Y'}{a}$ leads to $Y' = Q'(X' + W') = Q'(bX + bW) = aQ(X + W) = aY$. This shows that we only need to investigate the theoretical properties of the particular case, which is described by X, W, Y and Q, for the sake of clarity.

In [1] 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 $\varphi(w)$, by deducing the output entropy

$$\begin{aligned} h_Y(F) = {} & \rho\Big(\int \int_{-\infty}^{0} \varphi(u-x)\,\mathrm{d}u\,\mathrm{d}F(x)\Big) \\ & + \rho\Big(\int \int_{1}^{\infty} \varphi(u-x)\,\mathrm{d}u\,\mathrm{d}F(x)\Big) \\ & + \int_0^1 \rho\Big(\int \varphi(u-x)\,\mathrm{d}F(x)\Big)\,\mathrm{d}u \end{aligned} \tag{3}$$

and the conditional entropy

$$\begin{aligned} h_{Y|X}(F) = {} & \int \rho\Big(\int_{-\infty}^{0} \varphi(u-x)\,\mathrm{d}u\Big)\,\mathrm{d}F(x) \\ & + \int \rho\Big(\int_{1}^{\infty} \varphi(u-x)\,\mathrm{d}u\Big)\,\mathrm{d}F(x) \\ & + \int_0^1 \int \rho\big(\varphi(u-x)\big)\,\mathrm{d}F(x)\,\mathrm{d}u \,. \end{aligned} \tag{4}$$

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

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

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

and

$$\alpha(u;F) = \int \varphi(u-x)\,\mathrm{d}F(x) \,,$$

to enable a slightly different representation of the entropies $h_Y(F)$ and $h_{Y|X}(F)$.

In addition, the following statements are utilized in the present work.

Proposition 1: For any density f and arbitrary real numbers a and b with $a < b$, the identity

$$\begin{aligned} & \rho\Big(\int_a^b f(u)\,\mathrm{d}u\Big) - \int_a^b \rho\big(f(u)\big)\,\mathrm{d}u \\ & \qquad = -\int_a^b f(\tilde{u})\,\mathrm{d}\tilde{u} \int_a^b \rho\bigg(\frac{f(u)}{\int_a^b f(\tilde{u})\,\mathrm{d}\tilde{u}}\bigg)\,\mathrm{d}u \end{aligned}$$

holds.

Proof: By applying the simple identity $\rho(pq) = p\rho(q) + q\rho(p)$, for any non-negative real numbers p and q, on the expression $\rho\big(\frac{f(u)}{q} q\big)$ with $q = \int_a^b f(\tilde{u})\,\mathrm{d}\tilde{u}$ we obtain the above identity. ∎

Proposition 2: Let f and g be non-negative real functions with bounded integrals $\int_a^b f(u)\,\mathrm{d}u$ and $\int_a^b g(u)\,\mathrm{d}u$ for arbitrary real numbers a and b with $a < b$. Then the inequality

$$\int_a^b g(u)\,\rho\Big(\frac{f(u)}{g(u)}\Big)\,\mathrm{d}u \leq \int_a^b g(\tilde{u})\,\mathrm{d}\tilde{u}\;\rho\bigg(\frac{\int_a^b f(u)\,\mathrm{d}u}{\int_a^b g(\tilde{u})\,\mathrm{d}\tilde{u}}\bigg)$$

holds.

Proof: Since ρ is concave w.r.t. its argument, the above inequality results from the well-known Jensen's inequality. ∎

III. Comparing Mutual Information

By using the identity in Prop. 1 and the inequality in Prop. 2, we can upper bound the mutual information $I_{X;Y}$ by

$$\begin{aligned} & + \rho\Big(\int \int_{-\infty}^{0} \varphi(u-x)\,\mathrm{d}u\,\mathrm{d}F(x)\Big) \\ & - \int \rho\Big(\int_{-\infty}^{0} \varphi(u-x)\,\mathrm{d}u\Big)\,\mathrm{d}F(x) \\ & + \int_0^{\infty} \rho\Big(\int \varphi(u-x)\,\mathrm{d}F(x)\Big)\,\mathrm{d}u \\ & - \int_0^{\infty} \int \rho\big(\varphi(u-x)\big)\,\mathrm{d}F(x)\,\mathrm{d}u \,, \end{aligned} \tag{5}$$

which is the mutual information of the hinge function, cf. [9]. A further utilization of Prop. 1 and Prop. 2 yields the upper bound

$$\begin{aligned} & + \int_{-\infty}^{\infty} \rho\Big(\int \varphi(u-x)\,\mathrm{d}F(x)\Big)\,\mathrm{d}u \\ & - \int_{-\infty}^{\infty} \int \rho\big(\varphi(u-x)\big)\,\mathrm{d}F(x)\,\mathrm{d}u \,, \end{aligned} \tag{6}$$

which is the mutual information of common additive noise channels. In this way, the mutual information of the censored channel is less than the mutual information of the hinge channel. In turn, the mutual information of the hinge channel is less than the mutual information of additive channels. This is not surprising, since due to censoring, signal information is lost and cannot be recovered.

IV. Continuity of the Mutual Information

The general formula for the capacity is given by $C_{X;Y} = \sup_{F\in\mathcal{F}} I_{X;Y}(F)$. The set $\mathcal{F}$ is defined as the set of all probability distribution functions over the input space and is equipped with weak* topology[1]. As first step toward the solution of this problem, we discuss existence and uniqueness of the solution. A sufficient condition for the existence of the solution is the compactness of the set F and the continuity of $I_{X;Y}(F)$ in F. In this section, we discuss the continuity of the mutual information.

We prove the continuity for a class of noise distributions, namely the continuous distributions with a bounded probability density function. Note that from Lebesgue's decomposition theorem, the noise distribution in general can be decomposed into the sum of discrete, absolutely continuous and singular measures. We assume that the noise distribution does not have any singular and discrete parts. The following proposition settles the continuity of the mutual information for this class.

Proposition 3: Assuming an absolutely continuous noise distribution with bounded density, the mutual information $I_{X;Y}(F)$ is continuous in $F \in \mathcal{F}$.

Proof: To prove the continuity in F, it is enough to show that for each sequence of probability distributions $F_n \xrightarrow{w^*} F$, the function $I_{X;Y}(F_n)$ converges to $I_{X;Y}(F)$. We decompose the mutual information $I_{X;Y}(F)$ to $h_Y(F) - h_{Y|X}(F)$ and argue for continuity of each term. Fix the sequence of probability distributions $F_n \xrightarrow{w^*} F$. The conditional entropy as a function of F can be written

$$\begin{aligned} h_{Y|X}(F) = &- \int \ell(x)\log\ell(x)\,\mathrm{d}F(x) && (7)\\ &- \int r(x)\log r(x)\,\mathrm{d}F(x) && (8)\\ &- \int \left(\int_{-x}^{1-x} \varphi(w)\log\varphi(w)\,\mathrm{d}w\right)\mathrm{d}F(x). && (9)\end{aligned}$$

To prove the continuity, it is enough to prove boundedness and continuity of each term inside the integrals in (7), (8) and (9). If they are bounded and continuous, then the limit of integrals with respect to the measure F_n is equal to the integrals with respect to the measure F and the continuity follows.

First, since $r(x), \ell(x) \in [0,1]$, both terms $|r(x)\log r(x)|$ and $|\ell(x)\log\ell(x)|$ are bounded by $\frac{\log \mathrm{e}}{\mathrm{e}}$. Moreover both are continuous function since $r(x)$ and $\ell(x)$ are continuous. Therefore the continuity of (7) and (8) follows.

[1]Hereinafter, we use the short forms compactness and continuity instead of weak* compactness and weak* continuity, respectively.

We use the boundedness assumption of noise probability density for the third term. If the noise density is bounded, then $|\varphi(w)\log\varphi(w)|$ is also bounded by some M, and we have

$$\begin{aligned}\left|\int_{-x}^{1-x}\varphi(w)\log\varphi(w)\,\mathrm{d}w\right| &\le \int_{-x}^{1-x}|\varphi(w)\log\varphi(w)|\,\mathrm{d}w\\ &\le \int_{-x}^{1-x} M\,\mathrm{d}w = M. \qquad (10)\end{aligned}$$

The inequality (10) establishes the boundedness of the term inside the integral in (9). Morever the integral is continuous function of x and therefore the continuity of the last term follows. Therefore $h_{Y|X}(F)$ is continuous in F.

Now consider the expression of the output entropy $h_Y(F)$. Let Y_n denote the channel output random variable induced by the choice of input-distribution F_n. The continuity of $h_Y(F)$ amounts to the equality

$$\lim_{n\to\infty}\int_0^1 f_{Y_n}(y)\log f_{Y_n}(y)\,\mathrm{d}y = \int_0^1 \lim_{n\to\infty} f_{Y_n}(y)\log f_{Y_n}(y)\,\mathrm{d}y.$$

Note that the output entropy is equal to

$$\begin{aligned} h_Y(F) = &- L(F)\log L(F) - R(F)\log R(F)\\ &- \int_0^1 \alpha(y;F)\log\alpha(y;F)\,\mathrm{d}y. \qquad (11)\end{aligned}$$

To prove the continuity of $L(F)\log L(F)$ and $R(F)\log R(F)$, it is enough to prove the continuity of the linear functions $L(F)$ and $R(F)$. This is straightforward since the functions $\ell(x)$ and $r(x)$ are both bounded by one and hence uniformly integrable which implies the continuity of $L(F)$ and $R(F)$.

To prove the continuity of $\int_0^1 \alpha(y;F)\log\alpha(y;F)\,\mathrm{d}y$, we use dominated convergence theorem for an uniform distribution on the interval $(0,1)$. The boundedness assumption on noise density turns out to be needed here, as well. Bounded density implies that $\alpha(y;F) = \int \varphi(y-x)\mathrm{d}F$ is bounded and therefore, similar to what discussed previously, $\alpha(y;F)$ is continuous in F and $|\alpha(y;F)\log\alpha(y;F)|$ is bounded, say by M. The constant function is integrable with respect to uniform distribution and therefore it constitutes an integrable upper bound for $|\alpha(y;F)\log\alpha(y;F)|$. Hence, the dominated convergence theorem implies that

$$\begin{aligned}&\lim_{n\to\infty}\int_0^1 \alpha(y;F_n)\log\alpha(y;F_n)\,\mathrm{d}y\\ &\qquad = \int_0^1 \lim_{n\to\infty}\alpha(y;F_n)\log\alpha(y;F_n)\,\mathrm{d}y\\ &\qquad = \int_0^1 \alpha(y;F)\log\alpha(y;F)\,\mathrm{d}y,\end{aligned}$$

where the last step is justified by the continuity of $\alpha(y;F)$. This establishes the continuity of $h_Y(F)$ and hence the continuity of $I_{X;Y}(F)$. ∎

Note that the assumption on the boundness of the noise density is essential for the proof of Prop. 3.

V. On Optimal Input-Distribution

Existence of an optimal distribution follows if the set $\mathcal{F}$ is weak* compact and the mutual information $I_{X;Y}(F)$ is weak* continuous in $\mathcal{F}$. The latter we have proved above. If we would have a constraint on one of the moments of the input-distribution, the simple usage of Markov inequality would imply that the set $\mathcal{F}$ is tight. Tightness is important since it is a necessary condition for the relative compactness[2] of $\mathcal{F}$ using Prokhorov's theorem [10]. The compactness would follow from the sequential compactness of $\mathcal{F}$ and the metrizability of weak* topology by Lévy metric. Unfortunately, the set $\mathcal{F}$ is not compact in general and therefore one cannot argue for the existence of any solution. However, for some classes of noise distributions, it is possible to argue that the optimal distribution lies in a compact subset of $\mathcal{F}$ and therefore one can limit the search to this subset instead. More precisely, we show that for some classes, we can limit the search to a tight set of probability distributions. Therefore the existence follows consequently. Note that within the course of the argument, it can be seen that the uniqueness of the solution is automatically ruled out. Many distributions can be optimal but nevertheless are excluded to guarantee the tightness of the set.

The first result in this direction focuses on the class of noise distributions with bounded support. We assume that the noise has a bounded support, say in the set $[-K, K]$ for a fixed $K > 0$. The following proposition suggest that one can look only at input-distributions with bounded support to achieve capacity.

Proposition 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 $\delta > 0$.

Proof: Recall that $\ell(x) = \mathbb{P}(W \leq -x)$ and $r(x) = \mathbb{P}(W \geq 1-x)$. If $x > K$ then $\ell(x) = 0$ while for $x < -K$, $\ell(x) = 1$ holds. In both cases, $\ell(x)\log\ell(x)$ is zero. Similarly, if $x < -K+1$ then $r(x) = 0$ while if $x > 1+K$, then $r(x) = 1$. Again, $r(x)\log r(x) = 0$. This implies that the values of (7) and (8) do not depend on the value of the input-distribution outside $[-K, K+1]$. Now consider the term $\int_{-x}^{1-x}\varphi(w)\log\varphi(w)\mathrm{d}w$. For $x > K+1$ and $x < -K$, the integral is zero and therefore the value of (9) depends only on the value of the input-distribution inside $[-K, K+1]$. This implies that $h_{Y|X}(F)$ depends only on the value of the input-distribution inside $[-K, K+1]$.

Now consider $L(F)$ and $R(F)$. Since $\ell(x)$ is zero for $x \in (K, \infty)$, $L(F)$ does not depend on the value of F on (K, ∞). Moreover since $\ell(x)$ is one for $x \in (-\infty, K)$, $L(F)$ can be written as

$$L(F) = \int_{[-K,K]} \ell(x)\mathrm{d}F + \mathbb{P}(X < -K).$$

Using a similar argument for $R(F)$, we can see that:

$$R(F) = \int_{[1-K,1+K]} r(x)\mathrm{d}F + \mathbb{P}(X > K+1).$$

Note that for $\int\varphi(y-x)\mathrm{d}F$, $\varphi(y-x)$ is zero for $x \notin [-K+y, K+y]$. Since $y \in (0,1)$, this means that the integral depends on the value of F inside $[-K, K+1]$. Therefore $h_Y(F)$ depends only on the value of F on $[-K, K+1]$ and on both tail probabilities $\mathbb{P}(X < -K)$ and $\mathbb{P}(X > K+1)$. Therefore an arbitrary input-distribution F can be replaced by another distribution $\tilde{F}$ such that, first $I_{X;Y}(F) = I_{X;Y}(\tilde{F})$ and second the distribution of $\tilde{F}$ is equal to F on $[-K, K+1]$ with two mass points on $-K-\delta$ and $K+1+\delta$ with probabilities $\mathbb{P}(X < -K)$ and $\mathbb{P}(X > K+1)$, respectively. With this choice the mutual information remains the same. ∎

Corollary 5: For the censored channel, if the noise has a bounded probability density function with bounded support then there exist an input distribution that maximizes the mutual information.

Proof: From Prop. 4, one can limit the search on the set of input-distributions supported on $[-K-\delta, K+1+\delta]$. The set of all these probability distributions are trivially tight and hence compact. The continuity of I implies the existence of a solution. ∎

The advantage of the last corollary is that a search for an optimal input-distribution can be limited to a tight set of distributions. It is more difficult to carry on the same approach for noise distributions with an unbounded support. One reason is the difficulty of determining the level-sets of the entropy function. In other words, it is not clear which probability distributions in general yield the same entropy. However the following theorem proves that we can approach the optimal mutual information arbitrarily close using distributions with bounded support.

Theorem 6: Suppose that the noise has a probability density function with bounded derivative on $\mathbb{R}$. Let $\delta > 0$ be sufficiently small. Then there is an interval $A_\delta \subset \mathbb{R}$ such that for each choice of the input-distribution F, there exists another input-distribution $\tilde{F}$ supported on A_δ that approximates[3] the mutual information in the sense

$$\left| I_{X;Y}(\tilde{F}) - I_{X;Y}(F) \right| \doteq O(\delta).$$

Remark 7: In the previous theorem, we find an interval, e.g., $[-K, K]$, that contains most of the noise probability. In other words, the probability that the noise belongs to the set $(-\infty, -K) \cup (K, \infty)$ is at most $\varepsilon \ll 1$. The bound on the absolute value of the derivative of φ is then used to bound the noise probability density function $\varphi(w)$ itself over $(-\infty, -K) \cup (K, \infty)$. To see this, suppose that $|\varphi'(w)|$ is strictly bounded by α. Now consider the line $y = \alpha(w - w_0) + \varphi(w_0)$ passing through the point $(w_0, \varphi(w_0))$ for $w_0 \in (-\infty, K)$ with slope α. The line will not cross $\varphi(w)$ in another point $w \in (-\infty, K)$. If it would cross the

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

[3]We use the big O notation to describe the order of the error term in the approximation.

arbitrary point $(w_1, \varphi(w_1))$, then according to the mean value theorem, there must be a point w_2 between w_0 and w_1, such that $\varphi'(w_2) = \frac{\varphi(w_1)-\varphi(w_0)}{w_1-w_0} = \alpha$ which is contrary to α being the upper bound. Therefore the line through $(w_0, \varphi(w_0))$ will totally be under $\varphi(w)$ and so is a triangle created by this line, the x-axis, and the line $w = w_0$. The area of this triangle is $\frac{\varphi(w_0)^2}{2\alpha}$ and is less than ε. Using a similar argument for $w \in (K, \infty)$ by using a line with slope $-\alpha$, we can see that φ is bounded by $\sqrt{2\alpha\varepsilon}$ on all points outside $[-K, K]$.

Proof of Thm. 6: For the sake of compactness of the proof, w.l.o.g. we make use of the natural logarithm $\ln x = \frac{\log x}{\log \mathrm{e}}$ hereinafter.

To prove Thm. 6 we need to use the fact that for each probability measure $\mathbb{P}$ on $\mathbb{R}$ and for each $\varepsilon > 0$, there exists a number $K > 0$ such that $\mathbb{P}([-K, K]) > 1 - \varepsilon$. This means that each probability measure on $\mathbb{R}$ is tight and it is a simple conclusion of [10, Theorem 1.3] by using the fact that $\mathbb{R}$ is complete and separable. Now this fact can be applied to the noise distribution Φ to find $K > 0$ such that $\mathbb{P}(-K \leq W \leq K) > 1-\varepsilon$. Suppose that the input-distribution is F. $h_{Y|X}(F)$ can be expressed as in (7), (8) and (9). See that $\ell(x) = \mathbb{P}(W \leq -x)$. If $x > K$, we have $\ell(x) \leq \varepsilon$. This bound along with the simple inequality $-x \ln x \leq \frac{2\sqrt{x}}{\mathrm{e}}$, can be used to show that

$$-\ell(x) \ln \ell(x) \leq \frac{2}{\mathrm{e}}\sqrt{\ell(x)} \leq \frac{2}{\mathrm{e}}\sqrt{\varepsilon}\,, \tag{12}$$

and therefore

$$-\int_{(K,\infty)} \ell(x) \ln \ell(x)\, \mathrm{d}F \leq \frac{2}{\mathrm{e}}\sqrt{\varepsilon}\,. \tag{13}$$

If $x < -K$, then $\ell(x) = \mathbb{P}(W \leq -x) \geq 1 - \varepsilon$. Analogously, using $-x \ln x \leq (1 - x)$, we can see $-\ell(x) \ln \ell(x) \leq 1 - \ell(x) \leq \varepsilon$, which results in

$$-\int_{(-\infty,-K)} \ell(x) \ln \ell(x)\, \mathrm{d}F \leq \varepsilon. \tag{14}$$

This basically shows that

$$\int_{(-\infty,\infty)} \ell(x) \ln \frac{1}{\ell(x)}\, \mathrm{d}F - \int_{[-K,K]} \ell(x) \ln \frac{1}{\ell(x)}\, \mathrm{d}F \leq \frac{2+\mathrm{e}}{\mathrm{e}}\sqrt{\varepsilon}\,. \tag{15}$$

On the other hand, $r(x) = \mathbb{P}(W \geq 1-x)$. Similarly if $1-x > K$, $r(x) \leq \varepsilon$ and if $1 - x < -K$, $r(x) \geq 1 - \varepsilon$. Equivalent bounding techniques will show that

$$\int_{(-\infty,\infty)} r(x) \ln \frac{1}{r(x)}\, \mathrm{d}F - \int_{[1-K,1+K]} r(x) \ln \frac{1}{r(x)}\, \mathrm{d}F \leq \frac{2+\mathrm{e}}{\mathrm{e}}\sqrt{\varepsilon}\,. \tag{16}$$

Finally consider the last term (9) in $h_{Y|X}(F)$. For $1-x < -K$ or $-x > K$, we would like to bound $\left|\int_{-x}^{1-x} \varphi(w) \ln \varphi(w)\, \mathrm{d}w\right|$. The starting point is the triangle inequality

$$\left|\int_{-x}^{1-x} \varphi(w) \ln \varphi(w)\, \mathrm{d}w\right| \leq \int_{-x}^{1-x} \varphi(w)\, |\ln \varphi(w)|\, \mathrm{d}w.$$

The integral is decomposed into two parts. For the first part, it is assumed that $\varphi(w) \leq 1$. For this assumption we use the indicator function $\mathbb{1}$ to obtain

$$\begin{aligned}
&\int_{-x}^{1-x} \varphi(w)\, |\ln \varphi(w)|\, \mathbb{1}(\varphi(w) \leq 1)\, \mathrm{d}w \\
&\quad = -\int_{-x}^{1-x} \varphi(w) \ln \varphi(w) \mathbb{1}(\varphi(w) \leq 1)\, \mathrm{d}w \\
&\quad \overset{(a)}{\leq} \int_{-x}^{1-x} \frac{2}{\mathrm{e}}\sqrt{\varphi(w)}\, \mathbb{1}(\varphi(w) \leq 1)\, \mathrm{d}w \\
&\quad \overset{(b)}{\leq} \frac{2}{\mathrm{e}}\sqrt{\int_{-x}^{1-x} \varphi(w)\, \mathbb{1}(\varphi(w) \leq 1)\, \mathrm{d}w}\ \sqrt{\int_{-x}^{1-x} \mathrm{d}w} \\
&\quad \leq \frac{2}{\mathrm{e}}\sqrt{\varepsilon}\,,
\end{aligned}$$

where (a) is due to $-x \ln x \leq \frac{2}{\mathrm{e}}\sqrt{x}$ and (b) is due to Cauchy-Schwartz inequality. Using Rem. 7, $\varphi(w)$ is bounded by $\sqrt{2\alpha\varepsilon}$. If $\varphi(W)$ is sometimes larger than 1, then $\sqrt{2\alpha\varepsilon} > 1$. Now the assumption $\varphi(w) > 1$ implies that

$$\begin{aligned}
&\int_{-x}^{1-x} \varphi(w)\, |\ln \varphi(w)|\ \mathbb{1}(\varphi(w) > 1)\, \mathrm{d}w \\
&\quad = \int_{-x}^{1-x} \varphi(w) \ln \varphi(w)\, \mathbb{1}(\varphi(w) > 1)\, \mathrm{d}w \\
&\quad \overset{(c)}{\leq} \int_{-x}^{1-x} \sqrt{2\alpha\varepsilon}\, \ln \sqrt{2\alpha\varepsilon}\, \mathbb{1}(\varphi(w) > 1)\, \mathrm{d}w \\
&\quad \leq \sqrt{2\alpha\varepsilon}\, \ln \sqrt{2\alpha\varepsilon} \leq 2\alpha\varepsilon\,,
\end{aligned}$$

where (c) is because $x \ln x$ is increasing for $x \geq 1$. Using this decomposition, for $1 - x < -K$ or $-x > K$, the following holds for a constant $c_1 > 0$:

$$\left|\int_{-x}^{1-x} \varphi(w) \ln \varphi(w)\, \mathrm{d}w\right| \leq \frac{2}{\mathrm{e}}\sqrt{\varepsilon} + 2\alpha\varepsilon \leq c_1\sqrt{\varepsilon}\,.$$

Finally the last term is bounded by

$$\begin{aligned}
&\left|-\int \left(\int_{-x}^{1-x} \varphi(w) \ln \varphi(w)\, \mathrm{d}w\right) \mathrm{d}F \right. \\
&\quad \left. + \int_{[-K,K+1]} \left(\int_{-x}^{1-x} \varphi(w) \ln \varphi(w)\, \mathrm{d}w\right) \mathrm{d}F\right| = \\
&\left|-\int_{(-\infty,-K)\cup(K+1,\infty)} \left(\int_{-x}^{1-x} \varphi(w) \ln \varphi(w)\, \mathrm{d}w\right) \mathrm{d}F\right| \\
&\leq c_1\sqrt{\varepsilon}.
\end{aligned} \tag{17}$$

Using (15), (16) and (17), it can be seen that the conditional entropy $h_{Y|X}(F)$ is mostly resulted from the input-distribution on $[-K, K+1]$. The rest only perturbs it on the order of $\sqrt{\varepsilon}$. Therefore if another input-distribution $\tilde{F}_1$ agrees completely with F on $[-K, K+1]$, then it satisfies all above inequalities and $h_{Y|X}(\tilde{F}_1)$ agrees with $h_{Y|X}(F)$ for the most part. In this way, for some constant $c_2 > 0$ we have

$$\left|h_{Y|X}(F) - h_{Y|X}(\tilde{F}_1)\right| \leq c_2\sqrt{\varepsilon}\,. \tag{18}$$

Now we consider $h_Y(F)$. Similar to above arguments, we have $\int_{x\in[K,\infty)} \ell(x)\mathrm{d}F \le \varepsilon$ and $\int_{x\in(-\infty,-K]} \ell(x)\mathrm{d}F \ge (1-\varepsilon)\mathbb{P}(X<-K)$. Using them one can bound $L(F)$ by

$$\tilde{L}(F)-\varepsilon \le \tilde{L}(F)-\varepsilon\,\mathbb{P}(X<-K) \le L(F) \le \tilde{L}(F)+\varepsilon,$$

where $\tilde{L}(F)=\int_{x\in[-K,K]} \ell(x)\mathrm{d}F+\mathbb{P}(X<-K)$. The goal is to show that $\tilde{L}(F)\ln\frac{1}{\tilde{L}(F)}$ is at most $O(\sqrt{\varepsilon})$ far from $L(F)\ln\frac{1}{L(F)}$. Two cases can happen. First if $L(F)<\varepsilon$, then $0\le\tilde{L}(F)<2\varepsilon$. Moreover $L(F)\ln\frac{1}{L(F)}\le\frac{2}{\mathrm{e}}\sqrt{\varepsilon}$ and also $\tilde{L}(F)\ln\frac{1}{\tilde{L}(F)}\le\frac{2}{\mathrm{e}}\sqrt{2\varepsilon}$. This yields the intended result. The second case is when $L(F)\ge\varepsilon$. In this case we use the mean value theorem on ρ to achieve the inequality

$$-1=-\ln\mathrm{e}\le\frac{\rho(L)-\rho(\tilde{L})}{L-\tilde{L}}\le-\ln(\mathrm{e}\,\varepsilon)$$

or equivalently $|\rho(L)-\rho(\tilde{L})|\le|L-\tilde{L}|\max\{1,|\ln(\mathrm{e}\,\varepsilon)|\}$. With $|L-\tilde{L}|\le\varepsilon$ from above and $|\ln(\mathrm{e}\,\varepsilon)|\le\frac{2}{\sqrt{\mathrm{e}\,\varepsilon}}$ it follows $|\rho(L)-\rho(\tilde{L})|\le\sqrt{\varepsilon}\max\{1,\frac{2}{\sqrt{\mathrm{e}}}\}=\frac{2}{\sqrt{\mathrm{e}}}\sqrt{\varepsilon}$. Using this fact shows that in both cases, $|L(F)-\tilde{L}(F)|\le\varepsilon$ implies

$$\left|\tilde{L}(F)\ln\frac{1}{\tilde{L}(F)}-L(F)\ln\frac{1}{L(F)}\right|\doteq O(\sqrt{\varepsilon}). \tag{19}$$

The very same steps can be used to bound $R(F)=\int r(x)\mathrm{d}F$. Basically $\int_{x\in(K+1,\infty)} r(x)\mathrm{d}F\ge(1-\varepsilon)\,\mathbb{P}(X>K+1)$ and $\int_{x\in(\infty,1-K)} r(x)\mathrm{d}F\le\varepsilon$ and this implies that

$$\tilde{R}(F)-\varepsilon\le R(F)\le\tilde{R}(F)+\varepsilon,$$

where $\tilde{R}(F)=\int_{x\in[1-K,1+K]} r(x)\mathrm{d}F+\mathbb{P}(X>K+1)$. Similarly $|R(F)-\tilde{R}(F)|<\varepsilon$ implies

$$\left|\tilde{R}(F)\ln\frac{1}{\tilde{R}(F)}-R(F)\ln\frac{1}{R(F)}\right|\doteq O(\sqrt{\varepsilon}). \tag{20}$$

Now consider $\alpha(y;F)$ for $y\in(0,1)$. Remember that using Rem. 7, $\varphi(w)$ is bounded by $\sqrt{2\alpha\varepsilon}$ on the points outside $[-K,K]$. Hence $\varphi(y-x)$ is bounded by $\sqrt{2\alpha\varepsilon}$ for $y-x\notin[-K,K]$. Given that $y\in(0,1)$, we have $\int_{x\notin[-K,K+1]}\varphi(y-x)\mathrm{d}F\le\sqrt{2\alpha\varepsilon}$. Therefore $\alpha(y;F)\le\tilde{\alpha}(y;F)+\sqrt{2\alpha\varepsilon}$ where $\tilde{\alpha}(y;F)=\int_{x\in[-K,K+1]}\varphi(y-x)\mathrm{d}F$. Since $|\alpha(y;F)-\tilde{\alpha}(y;F)|<\sqrt{2\alpha\varepsilon}$, using the previous techniques leads to

$$\left|\alpha(y;F)\ln\frac{1}{\alpha(y;F)}-\tilde{\alpha}(y;F)\ln\frac{1}{\tilde{\alpha}(y;F)}\right|\doteq O(\varepsilon^{1/4}),$$

which in turn yields

$$\left|\int_0^1\alpha(y;F)\ln\frac{1}{\alpha(y;F)}\mathrm{d}y-\int_0^1\tilde{\alpha}(y;F)\ln\frac{1}{\tilde{\alpha}(y;F)}\mathrm{d}y\right| \doteq O(\varepsilon^{1/4})\,. \tag{21}$$

The equations (19), (20) and (21) show that if a measure $\tilde{F}_2$ agrees with F on $[-K,K+1]$ and also has the same probability on the sets $(K+1,\infty)$ and $(-\infty,-K)$ as $\mathbb{P}(X>K+1)$ and $\mathbb{P}(X<-K)$, respectively, then it approximates $h_Y(F)$ with an error of order $\varepsilon^{1/4}$. It means that for some constant $c_3>0$, $\left|h_Y(F)-h_Y(\tilde{F}_2)\right|\le c_3\varepsilon^{1/4}$. Note that $\tilde{F}_2$ also satisfies all conditions needed for (18). Therefore we finally have

$$\left|I_{X;Y}(F)-I_{X;Y}(\tilde{F}_2)\right|\doteq O(\delta) \tag{22}$$

with $\delta=\varepsilon^{1/4}$. ∎

Thm. 6 tells us that if $1-\varepsilon$ of the noise mass is concentrated on a set A, the capacity of censored channel can be approximated with the error $O(\varepsilon^{1/4})$ using only input-distributions supported on A, or in other words on a slightly larger and yet bounded set. More precisely we get the following.

Corollary 8: Let the noise have a probability density function with bounded derivative on $\mathbb{R}$. There exists an input-distribution $\tilde{F}$, which maximizes the mutual information on a bounded support and approaches the capacity of the censored channel within $O(\varepsilon^{1/4})$ gap.

VI. Conclusion and Future Work

In the present paper, we have investigated the censored channel with respect to its information theoretical properties. First we have shown that the considered model can completely describe the transfer function of arbitrary censored channels. Second we have presented two upper bounds on the mutual information of the censored channels to make further comparisons with common channels easier. Third, we have discussed the continuity of the mutual information and the existence of an optimal input-distribution to achieve capacity. We have derived, that capacity-achieving input has a bounded support whenever the noise distribution is continuous and has a bounded support, as well. Finally, input-distributions with bounded support can approach capacity with a negligible error even if the noise distribution has an unbounded support. Furture works focus on showing finite support of the capacity-achieving input-distribution.

References

[1] G. Alirezaei and R. Mathar, "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.

[2] 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.

[3] 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.

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

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

[6] 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.

[7] 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.

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

[9] ——, "On the information capacity of hinge functions," in *The IEEE International Symposium on Information Theory and Its Applications (ISITA'16)*, Monterey, California, USA, 2016.

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