Capacity Sensitivity in Additive Non-Gaussian Noise Channels

Malcolm Egan, Samir M. Perlaza, Vyacheslav Kungurtsev

HAL Id: hal-01522950

https://hal.archives-ouvertes.fr/hal-01522950

Submitted on 15 May 2017

Capacity Sensitivity in Additive Non-Gaussian Noise Channels

Malcolm Egan, Samir M. Perlaza and Vyacheslav Kungurtsev

Abstract—In this paper, a new framework based on the notion of *capacity sensitivity* is introduced to study the capacity of continuous memoryless point-to-point channels. The capacity sensitivity reflects how the capacity changes with small perturbations in any of the parameters describing the channel, even when the capacity is not available in closed-form. This includes perturbations of the cost constraints on the input distribution as well as on the channel distribution. The framework is based on continuity of the capacity, which is shown for a class of perturbations in the cost constraint and the channel distribution. The continuity then forms the foundation for obtaining bounds on the capacity sensitivity. As an illustration, the capacity sensitivity bound is applied to obtain scaling laws when the support of additive α-stable noise is truncated.

I. INTRODUCTION

In a wide class of communication systems, the channel capacity characterizes the cutoff rate beyond which the probability of error cannot be made arbitrarily close to zero. For the class of discrete memoryless channels the capacity is now well understood [1]. However, generalizing to continuous channels has proven non-trivial, with the important exception of the linear additive white Gaussian noise (AWGN) channel subject to a power constraint [1, Theorem 18].

Due to the difficulty in deriving closed-form expressions for the capacity and the optimal input distribution of continuous channels, the focus has shifted to determining structural properties of the optimal input distribution, as well as bounds and numerical methods to compute the capacity. By adopting this approach, a range of continuous channels have been considered including: non-linear or non-deterministic input-output relationships; general input constraints; and non-Gaussian noise. Early work in this direction was initiated by Smith [2] and more recently, Fahs and Abou-Faycal [3] have proven conditions for the discreteness and compactness of the optimal input distribution, which applies to a wide range of continuous channels. This provides a means to numerically compute the capacity, without resorting to the Blahut-Arimoto algorithm [4, 5].

Despite the progress in characterizing the optimal input distribution, there has been limited success in obtaining general closed-form characterizations of the capacity. Aside from the theoretical interest in such characterizations, it is also problematic for system design in the presence of non-Gaussian noise or input constraints beyond the power control—a problem for systems that experience impulsive noise [6, 7] or encode information in the timing of the signal [8].

An alternative approach to characterize the capacity of continuous channels is to focus on the *sensitivity* of the capacity, or how the capacity changes when any of the parameters describing the channel are varied. Along these lines, the effect of the input alphabet support has been studied in [9, 10]. In particular, it was shown that the gap between the capacity of the unit-variance discrete-input Gaussian memoryless channel converges exponentially fast to the capacity of the unit-variance continuous-input memoryless AWGN channel. In other work [11], the decrease of the capacity due to weak non-Gaussian contamination was studied in the context of the Gaussian noise channel. The key to these approaches is the continuity of the capacity with respect to parameters such as the input alphabet support or the value of the cost constraint.

In this paper, we introduce a general framework to study the capacity sensitivity by exploiting the theory of stability and sensitivity of optimization problems [12]. As a first step, we identify a large class of memoryless channels where the capacity is continuous with respect to parameters in the cost constraint or the channel distribution. This class includes channels with familiar cost constraints such as power and amplitude and Gaussian noise, as well as channels with many other constraints (including channels with multiple constraints) and non-Gaussian noise.

An important implication of the continuity of the capacity is that if two channels are "close", in a sense that will be clarified later, one channel can be used to approximate the capacity of the other. To this end, we derive new bounds to quantify the capacity sensitivity in two key classes of perturbations: the constraint cost; and the noise distribution when it is absolutely continuous with respect to the Lebesgue measure. We illustrate our framework by deriving a scaling law for the capacity sensitivity in truncated α-stable noise.

The remainder of this paper is organized as follows. Section II consists of a formulation of the capacity sensitivity problem. Sections III and IV specialize the capacity sensitivity to perturbations of the input constraint and the noise distribution, respectively. Section V applies our main results to characterize the capacity sensitivity in truncated Gaussian

and α-stable. Section VI discusses challenges of the general capacity sensitivity problem and outlines future directions.

II. The Capacity Sensitivity Problem

We are concerned with real-valued point-to-point memoryless channels. Consider the linear additive noise channel with output Y of the form

$$Y = X + N, \tag{1}$$

where the input X has an alphabet $\mathcal{X} \subseteq \mathbb{R}$ and the noise N has a distribution function on $\mathbb{R}$, denoted by F_N. In the case the noise has a probability density function on $\mathbb{R}$, it is denoted by p_N. Note that since the channel is linear and additive, when the noise has a probability density function the channel law can be written as

$$p_{Y|X}(y|x) = p_N(y - x). \tag{2}$$

As a consequence of the noisy channel coding theorem, when the capacity of (1) exists it is obtained by optimizing the mutual information subject to any constraints on the input X. Let $\mathcal{B}(\mathbb{R})$ be the Borel σ-algebra on $\mathbb{R}$ and let $\mathcal{P}$ denote the collection of Borel probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ equipped with the topology of weak convergence, which is metrized by the Lévy-Prokhorov metric [13]. The capacity of (1) is then the solution to the optimization problem

$$\begin{aligned} \sup_{\mu \in \mathcal{P}} \quad & I(X;Y) \\ \text{subject to} \quad & \mu \in \Lambda, \end{aligned} \tag{3}$$

where Λ is a compact subset of probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Key examples of the set Λ are the p-th order constraints ($p > 0$), defined by

$$\Lambda_p = \{\mu : \mathbb{E}_\mu[|X|^p] \leq \bar{b}\}, \tag{4}$$

where $\bar{b} > 0$. Here, the second-order power constraint $\mathbb{E}_\mu[X^2] \leq \bar{b}$, arising in power-limited wired and wireless communications [1] corresponds to $p = 2$; and the first-order constraint $\mathbb{E}[|X|] \leq \bar{b}$, arising in timing channels [8] corresponds to $p = 1$.

In the case in which $N \sim \mathcal{N}(0, \sigma^2)$, it is well-known that subject to a power constraint $\bar{b}$, the capacity is given by [1, Theorem 18]

$$C = \frac{1}{2} \log \left(1 + \frac{\bar{b}}{\sigma^2} \right). \tag{5}$$

However, this result is an anomaly: in general, it is not possible to obtain a simple closed-form representation of the capacity of (1) subject to arbitrary constraints. In fact, even the capacity of the AWGN channel is not well understood with constraint sets Λ_p for $p \neq 2$.

Our focus in this paper is to characterize the *capacity sensitivity*. In the most general formulation, we can view the capacity as a map from the input alphabet $\mathcal{X}$, the output alphabet $\mathcal{Y}$, the noise distribution F_N, and the constraint set Λ to $\mathbb{R}_+$. That is, $(\mathcal{X}, \mathcal{Y}, F_N, \Lambda) \mapsto C$. We define the *capacity sensitivity* as follows.

Definition 1. *Let $\mathcal{K} = (\mathcal{X}, \mathcal{Y}, F_N, \Lambda)$ and $\hat{\mathcal{K}} = (\hat{\mathcal{X}}, \hat{\mathcal{Y}}, \hat{F}_N, \hat{\Lambda})$ be two tuples of channel parameters. The capacity sensitivity due to a perturbation from channel $\mathcal{K}$ to the channel $\hat{\mathcal{K}}$ is defined as*

$$C_{\mathcal{K} \to \hat{\mathcal{K}}} \triangleq |C(\mathcal{K}) - C(\hat{\mathcal{K}})|. \tag{6}$$

The capacity sensitivity problem is a special case of analyzing the sensitivity of nonlinear optimization problems, where we identify the capacity as the *value function*. Clearly, the problem of computing the capacity sensitivity is trivial when the capacity is available in closed-form (such as the case of Gaussian noise with a power constraint). However, the problem is significantly more challenging in the usual situation in which the only explicit characterization of the capacity is (3) under general channel perturbations. As such, we will focus on two special classes of channel perturbations: the constraint set Λ and the noise distribution F_N.

III. Constraint Perturbations

In this section, we consider the capacity of channels subject to constraints of the form

$$\Lambda(\bar{\mathbf{b}}) = \{\mu : \mathbb{E}_\mu[f_i(|X|)] \leq \bar{b}_i, \ i = 1, 2, \ldots, m\}, \tag{7}$$

where μ is an input probability measure, $f_i : \mathbb{R} \to \mathbb{R}$, $i = 1, 2, \ldots, m$ are positive, non-decreasing functions with $f_i(|\cdot|)$ lower semicontinuous, and $\bar{b}_i \in \mathbb{R}_+$, $i = 1, 2, \ldots, m$. Moreover, we assume that $\mathbb{E}_\mu[f_i(|X|)]$ is weakly continuous for each $i = 1, 2, \ldots, m$. Note that this class of constraints includes the constraint sets Λ_p as special cases when X is also restricted to have compact support.

In the case of constraints of the form in (7), define the capacity function as

$$C(\bar{\mathbf{b}}) = \sup_{\mu \in \Lambda(\bar{\mathbf{b}})} I(X;Y), \tag{8}$$

where $\bar{\mathbf{b}} = [\bar{b}_1, \ldots, \bar{b}_m]^T$. We seek to characterize the capacity sensitivity for perturbations of $\bar{\mathbf{b}}$. More precisely, let $\tilde{\mathbf{b}} \in \mathbb{R}^m$. Then, the capacity sensitivity for perturbations of $\bar{\mathbf{b}}$ is given by

$$C_{\bar{\mathbf{b}} \to \tilde{\mathbf{b}}} = |C(\bar{\mathbf{b}}) - C(\tilde{\mathbf{b}})|. \tag{9}$$

A. Continuity of $C(\bar{\mathbf{b}})$

The first step to characterizing the capacity sensitivity $C_{\bar{\mathbf{b}} \to \tilde{\mathbf{b}}}$ is to establish continuity of $C(\bar{\mathbf{b}})$. Consider the following conditions.

(C1) $\Lambda(\bar{\mathbf{b}})$ in (7) is non-empty and compact.
(C2) $I(X;Y)$ is weakly continuous on $\Lambda(\bar{\mathbf{b}})$.

Theorem 1. *Suppose that conditions (C1) and (C2) hold. Then, $C(\bar{\mathbf{b}})$ in (8) is continuous at $\bar{\mathbf{b}}$.*

Proof. See Appendix A in [14]. $\square$

Observe that Theorem 1 relies on the weak continuity of the mutual information. The mutual information is weakly continuous for the case of discrete probability measures [15] but not in general for the continuous case. Despite this, a

range of continuous channels have been shown to satisfy weak continuity; including the Gaussian channel with a power constraint [9]. In particular, this implies that the capacity $C(\overline{\mathbf{b}})$ of the Gaussian channel with a power constraint is continuous, which is clearly consistent with (5). General conditions for the weak continuity of the mutual information have recently been provided in [16, Theorem 5].

B. Characterization of $C^-_{\overline{\mathbf{b}} \to \tilde{\mathbf{b}}}$

We now bound the capacity sensitivity $C^-_{\overline{\mathbf{b}} \to \tilde{\mathbf{b}}}$. Under the conditions in Theorem 1, $C(\overline{\mathbf{b}})$ in (8) is continuous. It then follows that if the directional derivative exists, we can apply the multivariate form of Taylor's theorem to quantify the effect of perturbing $\overline{\mathbf{b}}$ to $\tilde{\mathbf{b}}$. More precisely, Taylor's theorem yields

$$C(\tilde{\mathbf{b}}) = C(\overline{\mathbf{b}}) + D_{\mathbf{d}} C(\overline{\mathbf{b}}) + o(\|\tilde{\mathbf{b}} - \overline{\mathbf{b}}\|), \qquad (10)$$

where the direction $\mathbf{d}$ is given by $\mathbf{d} = \tilde{\mathbf{b}} - \overline{\mathbf{b}}$ and $D_{\mathbf{d}} C(\overline{\mathbf{b}})$ is the derivative of the capacity C in (8) in the direction $\mathbf{d}$ evaluated at the point $\overline{\mathbf{b}}$.

Observe that (10) provides a means of obtaining first-order estimates of the capacity at a point $\tilde{\mathbf{b}}$ given that the capacity is known at $\overline{\mathbf{b}}$. Aside from providing a general characterization of the capacity sensitivity, our approach can also be used to simplify numerical approximations of the capacity. In particular, suppose that it is challenging to obtain a large number of capacity points corresponding to different choices of $\overline{\mathbf{b}}$, then (10) forms a basis for the computation of piecewise linear approximations.

The problem that remains is to ensure the existence of $D_{\mathbf{d}} C(\overline{\mathbf{b}})$. To this end, recall that the capacity problem in (8) is convex and consider the dual of the problem in (8), given by

$$\inf_{\boldsymbol{\lambda} \geq 0} \sup_{\mu \in \mathcal{P}} I(X;Y) - \sum_{i=1}^m \lambda_i \left(\mathbb{E}_\mu[f_i(|X|)] - \overline{b}_i \right), \qquad (11)$$

with Lagrangian

$$L(\mu, \boldsymbol{\lambda}; \overline{\mathbf{b}}) = I(X;Y) - \sum_{i=1}^m \lambda_i \left(\mathbb{E}_\mu[f_i(|X|)] - \overline{b}_i \right). \qquad (12)$$

Consider the following condition.

(C3) There exists a unique optimal input probability measure μ^* for the problem (8).

We then have the following characterization of the directional derivative.

Lemma 1. *Let $\mathcal{L}(\overline{\mathbf{b}})$ be the set of Lagrange multipliers $\boldsymbol{\lambda}$ that optimize (11). Suppose that conditions (C1)-(C3) hold. Then, the directional derivative $D_{\mathbf{d}} C(\overline{\mathbf{b}})$ exists and is given by*

$$D_{\mathbf{d}} C(\overline{\mathbf{b}}) = \inf_{\boldsymbol{\lambda} \in \mathcal{L}(\overline{\mathbf{b}})} D_{\mathbf{d}} L(\mu^*, \boldsymbol{\lambda}; \overline{\mathbf{b}})$$

$$= \inf_{\boldsymbol{\lambda} \in \mathcal{L}(\overline{\mathbf{b}})} \sum_{i=1}^m \lambda_i d_i. \qquad (13)$$

where $\mathbf{d} = [d_1, \ldots, d_m]$ and $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_m]$.

Proof. See Appendix C in [14]. □

A bound for the capacity sensitivity with respect to $\overline{\mathbf{b}}$ in (7) then follows immediately by applying the triangle inequality to (10) and using Lemma 1.

Theorem 2. *Suppose the conditions (C1)-(C3) hold. Then, the capacity sensitivity $C^-_{\overline{\mathbf{b}} \to \tilde{\mathbf{b}}}$ is upper bounded by*

$$|C(\overline{\mathbf{b}}) - C(\tilde{\mathbf{b}})| \leq \|\boldsymbol{\lambda}^*\| \, \|\tilde{\mathbf{b}} - \overline{\mathbf{b}}\| + |o(\|\tilde{\mathbf{b}} - \overline{\mathbf{b}}\|)|, \qquad (14)$$

where $\boldsymbol{\lambda}^*$ is the Lagrange multiplier that optimizes (13).

IV. NOISE DISTRIBUTION PERTURBATIONS

In this section, we turn to the capacity sensitivity to perturbations in the noise distribution F_N. Throughout this section, we assume that F_N corresponds to an absolutely continuous probability measure with respect to the Lebesgue measure. Therefore there exists a noise probability density function p_N and the capacity sensitivity to perturbations of p_N is denoted by $C_{p_N^0 \to p_N^1}$, given by

$$C_{p_N^0 \to p_N^1} = |C(p_N^0) - C(p_N^1)|. \qquad (15)$$

Consider a sequence $\{p_N^i\}_{i=1}^\infty$ with $\|p_N^i - p_N^0\|_{TV} \to 0$. We first establish conditions on the sequence $\{p_N^i\}_{i=1}^\infty$ such that $\lim_{i \to \infty} C(p_N^i) = C(p_N^0)$. Using this result, we then derive an upper bound on $C_{p_N^0 \to p_N^1}$ in terms of $\|p_N^0 - p_N^1\|_{TV}$.

Note that the mutual information functional is completely determined by the input probability measure μ and the noise probability density function p_N. As such, we adopt the notation $I(X;Y) = I(\mu, p_N)$ to make the dependence explicit.

A. Convergence of $C(p_N^i)$

Consider a convergent sequence of probability density functions $\{p^i\}_{i=1}^\infty$ in an appropriate sense (i.e., pointwise, in total variation, in Kullback-Leibler divergence, or weakly) with $p^i \to p$. It is not true in general that the differential entropy converges [16, 17]; i.e., $\lim_{i \to \infty} h(p^i) \neq h(p)$. As a consequence, in order to ensure convergence of the mutual information and the capacity in (3) it is necessary to place restrictions on the sequence of probability density functions $\{p^i\}_{i=1}^\infty$.

In order to prove convergence of $C(p_N^i)$, it is therefore also necessary to place restrictions on the sequence of noise probability density functions $\{p_N^i\}_{i=1}^\infty$. The following convergence theorem is obtained by using the fact that the constraint set Λ is independent of the choice of p_N and applying a variation of Berge's maximum theorem [18].

Theorem 3. *Let $\{p_N^i\}_{i=1}^\infty$ be a pointwise convergent sequence with limit p_N^0. Let Λ be a compact set of probability measures not dependent on p_N, and $\{\mu_i\}_{i=1}^\infty$ be a weakly convergent sequence of probability measures in Λ with limit μ_0. Suppose the following conditions hold:*

(C4) *The mutual information $I(\mu, p_N)$ is weakly continuous on Λ.*

(C5) *For the convergent sequence $\{p_N^i\}_{i=1}^\infty$ and all weakly convergent sequences $\{\mu_i\}_{i=1}^\infty$ in Λ,*

$$\lim_{i \to \infty} I(\mu_i, p_N^i) = I(\mu_0, p_N^0). \qquad (16)$$

(C6) *There exists an optimal input probability measure μ_i^* for each noise probability density p_N^i.*

Then, $\lim_{i \to \infty} C(p_N^i) = C(p_N^0)$.

Proof. See Appendix D in [14]. $\qquad\qquad\qquad\square$

B. Characterization of $C_{p_N^0 \to p_N^1}$

Theorem 3 provides conditions on the sequence of probability density functions $\{p_N^i\}_{i=1}^{\infty}$ to ensure that the capacity $C(p_N^i)$ converges; however, it does not provide an explicit characterization of the capacity sensitivity $C_{p_N^0 \to p_N^1}$. We address the capacity sensitivity in the following theorem.

Theorem 4. *Let $\{p_N^i\}_{i=1}^{\infty}$ be a convergent sequence in total variation distance of noise probability density functions with limit p_N^0. Suppose that the conditions (C4)-(C6) in Theorem 3 hold. Further, suppose that the following condition holds:*

(C7) *Let $0 \le \theta \le 1$ and for all p_N^i define*

$$q_N^i(\theta) = (1 - \theta)p_N^0 + \theta p_N^i. \qquad (17)$$

For each i, suppose there exists $M_i < \infty$ and $N_i < \infty$ such that

$$\left| \lim_{\theta \to 0^+} \frac{I(\mu_0^*, q_N^i) - I(\mu_0^*, p_N^0)}{\theta \|p_N^0 - p_N^i\|_{TV}} \right| = M_i,$$

$$\left| \lim_{\theta \to 0^+} \frac{I(\mu_1^*, p_N^0) - I(\mu_1^*, q_N^i)}{\theta \|p_N^0 - p_N^i\|_{TV}} \right| = N_i, \qquad (18)$$

$$M = \sup_i M_i < \infty \text{ and } N = \sup_i N_i < \infty.$$

Then for any $i \ge 1$,

$$C_{p_N^0 \to q_N^i(\theta)} \le \max\{M, N\}\theta \|p_N^0 - p_N^i\|_{TV} + o(\theta). \quad (19)$$

Proof. See Appendix E in [14]. $\qquad\qquad\qquad\square$

Observe that Theorem 4 is bounded in terms of the total variation distance $\|p_N^0 - p_N^i\|_{TV}$. In particular, the theorem implies that $C_{p_N^0 \to p_N^i} = O(\|p_N^0 - p_N^i\|_{TV})$. In contrast with other metrics on spaces of probability density functions, the total variation distance can often be computed or bounded. In the following section, we apply Theorem 4 to investigate the effect of truncating symmetric α-stable noise probability density functions via the capacity sensitivity $C_{p_N^0 \to p_N^i}$.

V. CAPACITY SENSITIVITY IN α-STABLE NOISE

In this section, we characterize the capacity sensitivity $C_{p_N^0 \to p_N^1}$ in the case of truncated α-stable noise using the results in the previous section. The class of α-stable noise includes Gaussian noise ($\alpha = 2$) and Cauchy noise ($\alpha = 1$) as special cases. More generally, α-stable noise ($0 < \alpha < 2$) is often used as a model for impulsive noise and arises in wireless [6] and molecular [19] communication systems. We focus on the subclass of symmetric α-stable noise with $0 < \alpha < 2$, which has a characteristic function

$$\Phi(t) = e^{-\sigma^\alpha |t|^\alpha}, \qquad (20)$$

where $\sigma > 0$ is the scale parameter.

To proceed, let p_N^0 be a symmetric α-stable probability density function and p_N^T be a truncation of level $T > 0$ of p_N^0 defined by

$$p_N^T(x) = \begin{cases} \frac{p_N^0(x)}{\kappa_T}, & |x| \le T \\ 0, & \text{otherwise} \end{cases} \qquad (21)$$

where the normalization constant is given by

$$\kappa_T = \int_{|y| \le T} p_N^0(y) \mathrm{d}y. \qquad (22)$$

We assume that the constraint set $\Lambda = \{\mu : \mathbb{E}_\mu[|X|^r] \le c\}$ with $0 < r < \alpha$.

Our goal is to show that the conditions **(C4)**-**(C7)** in Theorem 3 and Theorem 4 hold, which implies that the capacity converges as $\|p_N^0 - p_N^1\|_{TV} \to 0$ and the sensitivity is bounded by (19).

Verification of (C4): Observe that the sequence $\{p_N^n\}_{n=1}^{\infty}$ converges pointwise and in total variation distance by the definition in (21). Moreover, the constraint set $\Lambda = \{\mu : \mathbb{E}_\mu[|X|^r] \le c\}$ for $0 < r < \alpha$ is compact in the topology of weak convergence. For a fixed p_N^n, by [16, Theorem 2] it follows that $I(\mu, p_N^n)$ is weakly continuous on Λ.

Verification of (C5): We need to show that for the sequence $\{p_N^n\}$ and all weakly convergent sequences $\{\mu_n\}$, we have $\lim_{n \to \infty} I(\mu_n, p_N^n) = I(\mu_0, p_N^0)$. By [16, Theorem 1], since $\{p_N^n\}$ converges pointwise to p_N^0 and has finite fractional moments it follows that the differential entropy

$$-\int_{-\infty}^{\infty} p_N^n(x) \log p_N^n(x) \mathrm{d}x \overset{n \to \infty}{\to} -\int_{-\infty}^{\infty} p_N^0(x) \log p_N^0(x) \mathrm{d}x. \qquad (23)$$

Let $Y_n = X_n + N_n$, where X_n is a random variable corresponding to the input probability measure μ_n and N_n is the noise random variable with probability density function p_N^n. We now show that the differential entropy $h(Y_n) \to h(Y_0)$. Since p_N^n is absolutely continuous for each n, it follows that

$$p_{Y_n}(y) = \int_{-\infty}^{\infty} p_N^n(y - x) \mathrm{d}\mu_n(x). \qquad (24)$$

The characteristic function of Y_n, denoted by $\Phi(Y_n)$, is then given by

$$\Phi_{Y_n}(t) = \Phi_{X_n}(t)\Phi_{N_n}(t), \qquad (25)$$

where Φ_{X_n} and Φ_{N_n} are the characteristic functions of X_n and N_n, respectively. As μ_n converges weakly and p_N^n converges pointwise, we then have

$$\lim_{n \to \infty} \Phi_{Y_n}(t) = \lim_{n \to \infty} \Phi_{X_n}(t)\Phi_{N_n}(t)$$
$$= \Phi_{X_0}(t)\Phi_{N_0}(t) = \Phi_{Y_0}(t). \qquad (26)$$

This implies that Y_n converges weakly and hence $p_{Y_n}(y)$ converges pointwise. Again applying [16, Theorem 1], it follows that $h(Y_n) \to h(Y_0)$. This completes the proof that $\lim_{n \to \infty} I(\mu_n, p_N^n) = I(\mu_0, p_N^0)$.

Verification of (C6): By [16, Theorem 2], there exists a unique optimal input probability measure μ_n^*.

As the conditions in Theorem 3 are satisfied, it follows that $\lim_{n\to\infty} C(p_N^n) = C(p_N^0)$. In other words, the capacity converges as the truncation level $T \to \infty$. One implication of this result is that numerical approximations of the capacity based on truncations of the support of symmetric α-stable noise converge as $T \to \infty$.

In order to obtain an estimate of the capacity sensitivity $C_{p_{N^0} \to p_{N,T}}$, we seek to use Theorem 4. As conditions (**C4**)-(**C6**) are satisfied, all that remains is to show that condition (**C7**) also holds, which is verified in [14, Appendix F].

Having shown that conditions (**C4**)-(**C7**) hold, we now evaluate the bound in Theorem 4 for the cases of truncated symmetric α-stable noise. In general, the capacity of symmetric α-stable noise channels under constraints of the form $\mathbb{E}_\mu[|X|^r] \leq c$ are not known. To understand the effect of the truncation on the capacity sensitivity, we investigate the asymptotic scaling law $|C(p_N^0) - C(p_N^n)| = O(\|p_N^0 - p_N^n\|_{TV})$, which is a consequence of Theorem 4. Observe that

$$
\int_{|x| \leq n} |p_N^0(x) - p_N^n(x)| \mathrm{d}x
$$
$$
= \left| 1 - \frac{1}{\kappa_n} \right| \left(1 - \int_{|x| > n} p_N^0(x) \mathrm{d}x \right) = 1 - \kappa_n. \tag{27}
$$

Similarly,

$$
\int_{|x| > n} |p_N^0(x) - p_N^n(x)| \mathrm{d}x = \int_{|x| > n} p_N^0(x) \mathrm{d}x = 1 - \kappa_n, \tag{28}
$$

from which it follows that $\|p_N^0 - p_N^n\|_{TV} = \frac{1}{2}(1 - \kappa_n)$ with κ_n as defined in (22).

Now, the asymptotic probability density function tail representation for the symmetric α-stable random variable N_0 corresponding to p_N^0, given by [20, Eq. (1.2.10)]

$$
\mathbb{P}(N_0 > \lambda) = \sigma^\alpha C_\alpha \lambda^{-\alpha}, \tag{29}
$$

where C_α is a constant only depending on α. As such, $1 - \kappa_n = O(n^{-\alpha})$. Applying this result to Theorem 4, then implies that the capacity sensitivity for a truncation level $T = n$ is given by

$$
|C(p_N^0) - C(p_N^n)| = O(n^{-\alpha}). \tag{30}
$$

VI. Conclusions

With the important exception of Gaussian point-to-point channels subject to an average power constraint, there has been limited success in characterizing the capacity of continuous channels. In this paper, we have approached this problem using a framework based on the new notion of capacity sensitivity. In particular, we provided general conditions to guarantee continuity of the capacity with respect to parameters describing the channel. The continuity then formed the foundations to obtain bounds on the capacity sensitivity. The sensitivity bound was applied to obtain scaling laws for the capacity when the support is truncated for Gaussian and α-stable noise distributions.

From a more general perspective, the capacity sensitivity framework provides a new means of understanding how channel parameters affect the capacity. Beyond the perturbations we have considered, there are many other parameters of the channel that are of interest. Some of the open questions beyond the scope of this paper include what is the influence of more general perturbations of the constraint set on the capacity? More concretely, how is the capacity influenced by changes from a power constraint to low order fractional moment constraints? Another open question is whether or not it is possible to obtain closed-form bounds on the capacity sensitivity for truncated α-stable noise distributions?

References

[1] C. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423,623–656, Jul., Oct. 1948.

[2] J. Smith, "The information capacity of amplitude- and variance-constrained scalar Gaussian channels," *Information and Control*, vol. 18, no. 3, pp. 203–219, April 1971.

[3] J. Fahs and I. Abou-Faycal, "Input constraints and noise density functions: a simple relation for bounded-support and discrete-capacity achieving inputs," *arXiv:1602.00878*, 2016.

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

[5] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 18, no. 4, pp. 14–20, Jan. 1972.

[6] P. Pinto and M. Win, "Communication in a Poisson field of interferers-part II: Channel capacity and interference spectrum," *IEEE Transactions on Wireless Communications*, vol. 9, no. 7, pp. 2187–2195, Jul. 2010.

[7] M. Egan, M. de Freitas, L. Clavier, A. Goupil, G. Peters, and N. Azzaoui, "Achievable rates for additive isotropic alpha-stable noise channels," in *IEEE International Symposium on Information Theory*, Barcelona, Spain, Jul. 2016.

[8] K. Srinivas, A. Eckford, and R. Adve, "Molecular communication in fluid media: The additive inverse Gaussian noise channel," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4678–4692, Jul. 2012.

[9] Y. Wu and S. Verdú, "Functional properties of minimum mean-square error and mutual information," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1289–1301, Mar. 2012.

[10] Y. Wu and S. Verdú, "The impact of constellation cardinality on Gaussian channel capacity," in *Proc. of the 48th Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, Sept 2010.

[11] M. Pinsker, V. Prelov, and S. Verdú, "Sensitivity of channel capacity," *IEEE Transactions on Information Theory*, vol. 41, no. 6, pp. 1877–1888, 1995.

[12] J. Bonnans and A. Shapiro, "Optimization problems with perturbations: A guided tour," *SIAM Rev.*, vol. 40, no. 2, pp. 228–264, Jun. 1998.

[13] P. Billingsley, *Convergence of Probability Measures*, 2nd ed. New York, NY: John Wiley and Sons, 1999.

[14] M. Egan, S. M. Perlaza, and V. Kungurtsev, "Capacity sensitivity of continuous channels," INRIA, Lyon, France, Tech. Rep. 9012, Jan. 2017.

[15] R. Rockafellar, *Convex Analysis*. Princeton, N.J.: Princeton University Press, 1970.

[16] J. Fahs and I. Abou-Faycal, "On the finiteness of the capacity of continuous channels," *IEEE Transactions on Communications*, vol. 54, no. 1, pp. 166–173, Jan. 2016.

[17] J. Silva and P. Parada, "On convergence properties of Shannon entropy," *Problems of Information Transmission*, vol. 45, no. 2, pp. 75–94, Jun. 2009.

[18] E. Ok, *Real Analysis with Economic Applications*. Princeton, N.J.: Princeton University Press, 2007.

[19] N. Farsad, W. Guo, C.-B. Chae, and A. Eckford, "Stable distributions as noise models for molecular communication," in *Proc. IEEE Global Communications Conference*, San Diego, CA, 2015.

[20] G. Samorodnitsky and M. Taqqu, *Stable Non-Gaussian Random Processes*. New York, NY: Chapman and Hall, 1994.