Capacity Sensitivity in Continuous Channels

Malcolm Egan, Samir M. Perlaza, Vyacheslav Kungurtsev

Capacity Sensitivity in Continuous Channels

Malcolm Egan, Samir M. Perlaza, Vyacheslav Kungurtsev

RESEARCH REPORT
N° 9012
January 2017
Project-Team Socrate

ISSN 0249-6399 ISRN INRIA/RR--9012--FR+ENG

Capacity Sensitivity in Continuous Channels

Malcolm Egan, Samir M. Perlaza, Vyacheslav Kungurtsev

Project-Team Socrate

Research Report n° 9012 — January 2017 — 22 pages

Abstract: In this research report, 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.

Key-words: Capacity Sensitivity, Continuous Memoryless Channels, Non-Gaussian Noise

Malcolm Egan and Samir M. Perlaza are with the Institut National de Recherche en Informatique et en Automatique (INRIA), Centre de Recherche de Grenoble, 38334 Montbonnot, France. ({malcom.egan, samir.perlaza}@inria.fr). S. Perlaza is also with the Department of Electrical Engineering at Princeton University, Princeton, NJ 08544 USA.

Vyacheslav Kungurtsev is with the Faculty of Electrical Engineering, Czech Technical University in Prague, 12135 Prague 2, Czech Republic.

This research was supported in part by the European Commission under Marie Skłodowska-Curie Individual Fellowship No. 659316.

RESEARCH CENTRE
GRENOBLE – RHÔNE-ALPES

Inovallée
655 avenue de l'Europe Montbonnot
38334 Saint Ismier Cedex

Sensibilité de capacité dans les canaux continus

Résumé : Dans ce rapport de recherche, un nouveau cadre basé sur la notion de la sensibilité de capacité est présenté afin d'étudier la capacité des canaux point-à-point continus sans mémoire. La sensibilité de capacité reflète comment la capacité varie en fonctions des petites perturbations de l'un des paramètres décrivant le canal, même si l'expression explicite de la capacité n'est pas connue. Cela inclut les perturbations des contraintes de coût sur la distribution en entrée du canal ainsi que sur la distribution du canal. Ce cadre est basé sur la continuité de la capacité, qui est démontrée pour une classe de perturbations des contraintes de coût et la distribution du canal. La continuité forme ainsi la base pour obtenir les bornes sur la sensibilité de la capacité. Pour illustrer tout ça, la borne sur sensibilité de la capacité est appliquée afin d'obtenir des lois de mise à l'échelle quand le support du bruit additif α-stable est tronqué.

Mots-clés : sensibilité de capacité, canaux point-à-point continus sans mémoire, bruit non gaussien

Contents

1 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. The key to this approach 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 research report, we introduce a general framework to study the capacity sensitivity by exploiting the theory of stability and sensitivity of optimization problems [11]. 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 2 consists of a formulation of the capacity sensitivity problem. Sections 3 and 4 specialize the capacity sensitivity to perturbations of the input constraint and the noise distribution, respectively. Section 5 applies our main results to characterize the capacity sensitivity in truncated Gaussian and α-stable. Section 6 discusses challenges of the general capacity sensitivity problem and outlines future directions.

2 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 [12]. 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.

3 Constraint Perturbations

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

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

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

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

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

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

3.1 Continuity of $C(\overline{\mathbf{b}})$

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

(C1) $\Lambda(\overline{\mathbf{b}})$ in (7) is non-empty and compact.

(C2) $I(X;Y)$ is weakly continuous on $\Lambda(\overline{\mathbf{b}})$.

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

Proof. See Appendix A. □

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 [13] 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 [14, Theorem 5].

3.2 Characterization of $C_{\overline{\mathbf{b}}\to\tilde{\mathbf{b}}}$

We now bound the capacity sensitivity $C_{\overline{\mathbf{b}}\to\tilde{\mathbf{b}}}$. By virtue of 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

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

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

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

with Lagrangian

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

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*

$$\begin{aligned} D_{\mathbf{d}}C(\overline{\mathbf{b}}) &= \inf_{\boldsymbol{\lambda}\in\mathcal{L}(\overline{\mathbf{b}})} D_{\mathbf{d}}L(\mu^*, \boldsymbol{\lambda}; \overline{\mathbf{b}}) \\ &= \inf_{\boldsymbol{\lambda}\in\mathcal{L}(\overline{\mathbf{b}})} \sum_{i=1}^{m} \lambda_i d_i, \end{aligned} \tag{13}$$

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

Proof. See Appendix C. □

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*

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

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

4 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

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

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.

4.1 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 [14, 15]; 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 [16].

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,*

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

(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. □

4.2 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* $0 \leq \theta \leq 1$ *and for all* p_N^i *define*

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

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

$$\begin{aligned} \left| \lim_{\theta \to 0^+} \frac{I(\mu_0^*, q_N^i) - I(\mu_0^*, p_N^0)}{\theta \|p_N^0 - p_N^i\|_{TV}} \right| &= M_i, \\ \left| \lim_{\theta \to 0^+} \frac{I(\mu_1^*, p_N^0) - I(\mu_1^*, q_N^i)}{\theta \|p_N^0 - p_N^i\|_{TV}} \right| &= N_i, \end{aligned} \tag{18}$$

$M = \sup_i M_i < \infty$ *and* $N = \sup_i N_i < \infty$*.*

Then for any $i \geq 1$*,*

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

Proof. See Appendix E. □

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 $C_{p_N^0 \to p_N^i}$.

5 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 [17] communication systems. We focus on the subclass of symmetric α-stable noise with $0 < \alpha < 2$, which has a characteristic function

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

where $\sigma > 0$ is the scale parameter. In general, symmetric α-stable noise does not have a closed-form probability density function. As such, the characteristic function plays an important role.

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

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

where the normalization constant is given by

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

We assume that the constraint set $\Lambda = \{\mu : \mathbb{E}_\mu[|X|^r] \leq c\}$ 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 $\Lambda = \{\mu : \mathbb{E}_\mu[|X|^r] \leq c\}$ for $0 < r < \alpha$ is compact in the topology of weak convergence. For a fixed p_N^n, by [14, 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 [14, Theorem 1], since $\{p_N^n\}$ converges pointwise to p_N^0 and has finite fractional moments it follows that the differential entropy

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

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

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

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

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

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

$$\begin{aligned}\lim_{n\to\infty} \Phi_{Y_n}(t) &= \lim_{n\to\infty} \Phi_{X_n}(t)\Phi_{N_n}(t)\\ &= \Phi_{X_0}(t)\Phi_{N_0}(t) = \Phi_{Y_0}(t).\end{aligned} \tag{26}$$

This implies that Y_n converges weakly and hence $p_{Y_n}(y)$ converges pointwise. Again applying [14, 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 [14, 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 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

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

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 [18, 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

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

6 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? More generally, is it possible to characterize the effect of perturbing noise distributions of mixed type?

Appendices

A Proof of Theorem 1

Throughout this appendix, we use the following notation. Let $\mathbf{b}, \mathbf{b}' \in \mathbb{R}^m$, $m \in \mathbb{N}$, then $\mathbf{b} > \mathbf{b}'$ implies that $b_i > b_i'$ for all $i = 1, 2, \ldots, m$. We now turn to the proof of Theorem 1, which relies on the following lemma, proven in Appendix B.

Lemma 2. *Suppose that $\Lambda(\overline{\mathbf{b}})$ is non-empty and denote the strict interior as $\mathcal{I}_{\overline{\mathbf{b}}} = \{\mu : \mathbb{E}_\mu[f_i(|X|)] < \bar{b}_i,\ i = 1, 2, \ldots, m\}$. Further, suppose that $I(X;Y)$ is weakly continuous, there exists a $\tilde{\mathbf{b}} > \overline{\mathbf{b}}$ such that $\Lambda(\tilde{\mathbf{b}})$ is compact, and the closure of the strict interior $\mathrm{cl}(\mathcal{I}_{\overline{\mathbf{b}}}) = \Lambda(\overline{\mathbf{b}})$. Then, the capacity $C(\overline{\mathbf{b}})$ is continuous at $\overline{\mathbf{b}}$.*

To use Lemma 2, the following are proven: (a) there exists $\tilde{\mathbf{b}} > \overline{\mathbf{b}}$ such that $\Lambda(\tilde{\mathbf{b}})$ and (b) the closure condition $\mathrm{cl}(\mathcal{I}_{\overline{\mathbf{b}}}) = \Lambda(\overline{\mathbf{b}})$ holds.

We begin by proving (a). Let $\tilde{\mathbf{b}} \in \mathbb{R}_+^m$. To proceed, we first show that $\Lambda(\tilde{\mathbf{b}})$ is compact by applying Prokhorov's theorem [12]. In particular, we need to show that $\Lambda(\tilde{\mathbf{b}})$ is tight and closed.

Observe that for any $\epsilon > 0$ and $i \in \{1, 2, \ldots, m\}$, there exists an $a_{i,\epsilon} > 0$ such that for all $\mu \in \Lambda(\tilde{\mathbf{b}})$,

$$\mathbb{P}(|X| \geq a_{i,\epsilon}) \leq \frac{\mathbb{E}_\mu[f_i(|X|)]}{a_{i,\epsilon}} \leq \frac{\bar{b}_i}{a_{i,\epsilon}} < \epsilon, \tag{31}$$

which follows from the Markov inequality and the fact that each f_i is a positive, non-decreasing function. Choose $\mathcal{K}_\epsilon = [-a_{*,\epsilon}, a_{*,\epsilon}]$, where $a_{*,\epsilon} = \max_i a_{i,\epsilon}$. Hence, $\mathcal{K}_\epsilon$ is compact on $\mathbb{R}$ and $\mu(\mathcal{K}_\epsilon) \geq 1 - \epsilon$ for all $\mu \in \Lambda(\tilde{\mathbf{b}})$. Recall that $\Lambda(\tilde{\mathbf{b}})$ is tight if for all probability measures $\mu \in \Lambda(\tilde{\mathbf{b}})$ there exists a compact subset $\mathcal{K}_\epsilon$ of $\mathbb{R}^m$ such that $\mu(\mathcal{K}_\epsilon) > 1 - \epsilon$. We have therefore shown that $\Lambda(\tilde{\mathbf{b}})$ is tight.

To show that $\Lambda(\tilde{\mathbf{b}})$ is closed, let $\{\mu_n\}_{n=1}^\infty$ be a convergent sequence in $\Lambda(\tilde{\mathbf{b}})$ with limit μ_0. Recall that $f_i(|x|)$ is lower semicontinuous and bounded from below. By a consequence of the Portmanteau theorem [12], for each $i = 1, 2, \ldots, m$ we have

$$\begin{aligned} \mathbb{E}_{\mu_0}[f_i(|X|)] &= \int_{-\infty}^{\infty} f_i(|x|)\mathrm{d}\mu_0 \\ &\leq \liminf_{n\to\infty} \int_{-\infty}^{\infty} f_i(|x|)\mathrm{d}\mu_n(x) \leq \bar{b}_i. \end{aligned} \tag{32}$$

The inequality in (32) means that the convergent sequence $\{\mu_n\}_{n=1}^\infty$ converges to a probability measure in $\Lambda(\tilde{\mathbf{b}})$. Since the inequality in (32) holds for all $i = 1, 2, \ldots, m$, it follows that $\mu_0 \in \Lambda(\tilde{\mathbf{b}})$. As our choice of convergent sequence was arbitrary, it also follows that $\Lambda(\tilde{\mathbf{b}})$ is closed and hence $\Lambda(\tilde{\mathbf{b}})$ is compact.

Note that since our choice of $\tilde{\mathbf{b}}$ was arbitrary, for any given $\overline{\mathbf{b}} \in \mathbb{R}_+^m$ we can find a choice of $\tilde{\mathbf{b}}$ such that $\tilde{\mathbf{b}} > \overline{\mathbf{b}}$ and $\Lambda(\tilde{\mathbf{b}})$ is compact, which completes the proof of step (a).

We now prove step (b) to complete the proof. To show that $\mathrm{cl}(\mathcal{I}_{\overline{\mathbf{b}}}) = \Lambda(\overline{\mathbf{b}})$, we require that the limit point of any convergent sequence in $\mathcal{I}_{\overline{\mathbf{b}}}$ lies in $\Lambda(\overline{\mathbf{b}})$. This follows immediately from same argument as in (32).

B Proof of Lemma 2

To prove Lemma 2, we extend the results of Evans and Gould [19] to the case of probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ equipped with the topology of weak convergence. The key to the extension is that fact that the topology of weak convergence is metrized by the Lévy-Prokhorov metric, denoted by ρ, which means that a similar argument can be applied. To make this report self-contained, we provide the details of the proof.

Define $\mathcal{B}$ as the set of points $\overline{\mathbf{b}}$ such that $\Lambda(\overline{\mathbf{b}})$ is non-empty. Denote $\frac{1}{\mathbf{n}} = [\frac{1}{n}, \ldots, \frac{1}{n}]^T \in \mathbb{R}^m$, $n \geq 1$. Further, define $N_\epsilon(\mu)$ as an ϵ-ball centered at μ and let $\rho(\mu, \Lambda(\overline{\mathbf{b}})) = \inf_{\hat{\mu} \in \Lambda(\overline{\mathbf{b}})} \rho(\mu, \hat{\mu})$.

Definition 2. *Suppose $\overline{\mathbf{b}} \in \mathcal{B}$ and $\epsilon > 0$. An ϵ-neighborhood of $\Lambda(\overline{\mathbf{b}})$, denoted by $\eta_\epsilon(\Lambda(\overline{\mathbf{b}}))$ is defined by*

$$\eta_\epsilon(\Lambda(\overline{\mathbf{b}})) = \{\mu : \rho(\mu, \Lambda(\overline{\mathbf{b}})) < \epsilon\} = \cup_{\mu \in \Lambda(\overline{\mathbf{b}})} N_\epsilon(\mu). \tag{33}$$

The first notion of continuity is upper hemicontinuity, which is defined as follows.

Definition 3. *Suppose $\Lambda(\overline{\mathbf{b}})$ is compact. The point-to-set map Λ is upper hemicontinuous at $\overline{\mathbf{b}}$ if $\forall \epsilon > 0$, there exists $\delta > 0$ such that $\|\mathbf{b} - \overline{\mathbf{b}}\| < \delta$ implies that $\Lambda(\mathbf{b}) \subseteq \eta_\epsilon(\Lambda(\overline{\mathbf{b}}))$. Equivalently, Λ is upper hemicontinuous at $\overline{\mathbf{b}}$ if whenever $\mathbf{b}_n \in \mathcal{B}$ and $\mathbf{b}_n \to \mathbf{b}$ there exists an $n_0 \in \mathbb{N}$ such that $\Lambda(\mathbf{b}_n) \subseteq \eta_\epsilon(\Lambda(\overline{\mathbf{b}}))$ for all $n > n_0$.*

The second notion of continuity is lower hemicontinuity.

Definition 4. *Suppose $\Lambda(\overline{\mathbf{b}})$ is compact. The point-to-set map Λ is lower hemicontinuous at $\overline{\mathbf{b}}$ if $\forall \epsilon > 0$, there exists $\delta > 0$ such that $\|\mathbf{b} - \overline{\mathbf{b}}\| < \delta$ implies that $\Lambda(\overline{\mathbf{b}}) \subseteq \eta_\epsilon(\Lambda(\mathbf{b}))$. Equivalently, Λ is lower hemicontinuous at $\overline{\mathbf{b}}$ if whenever $\mathbf{b}^{(n)} \in \mathcal{B}$ and $\mathbf{b}^{(n)} \to \overline{\mathbf{b}}$ there exists an $n_0 \in \mathbb{N}$ such that $\Lambda(\overline{\mathbf{b}}) \subseteq \eta_\epsilon(\Lambda(\mathbf{b}^{(n)}))$ for all $n > n_0$.*

It is important to not to confuse upper and lower hemicontinuity of point-to-set maps with upper and lower semicontinuity of functions. Intuitively, upper hemicontinuity can be viewed as constraining the size of expansions of the set $\Lambda(\overline{\mathbf{b}})$, in the presence of small changes to $\overline{\mathbf{b}}$. Conversely, lower hemicontinuity can be viewed as constraining the size of contractions. When both upper and lower hemicontinuity hold for a map $\Lambda(\overline{\mathbf{b}})$ at the point $\overline{\mathbf{b}}$, the map is said to be hemicontinuous at $\overline{\mathbf{b}}$.

We begin the proof of Lemma 2 by showing that $\Lambda(\overline{\mathbf{b}})$ is upper hemicontinuous. The following lemma is instrumental in the proof.

Lemma 3. *Suppose $\Lambda(\tilde{\mathbf{b}})$ is compact for some $\tilde{\mathbf{b}} > \overline{\mathbf{b}}$. Then, for each $\epsilon > 0$ there is a $\mathbf{b}^* > \overline{\mathbf{b}}$ such that $\Lambda(\mathbf{b}^*) \subseteq \eta_\epsilon(\Lambda(\overline{\mathbf{b}}))$.*

Proof. Consider the sequence $\overline{\mathbf{b}} + \frac{1}{\mathbf{n}}$. For n sufficiently large, $\overline{\mathbf{b}} + \frac{1}{\mathbf{n}} < \tilde{\mathbf{b}}$ and by the weak continuity of $\mathbb{E}_\mu[f_i(|X|)]$, $\Lambda(\overline{\mathbf{b}} + \frac{1}{\mathbf{n}})$ is closed and hence compact. Let μ_n be a point in $\Lambda(\overline{\mathbf{b}} + \frac{1}{\mathbf{n}})$ that is at a maximum distance from $\Lambda(\overline{\mathbf{b}})$. Since the terms of the sequence μ_n eventually lie in a compact set, there is a convergent sequence $\mu_{n_j} \to \mu_0$. We then have $\mathbb{E}_{\mu_{n_j}}[f_i(|X|)] \leq \overline{\mathbf{b}} + \frac{1}{\mathbf{n}_j}$ for each $i = 1, 2, \ldots, m$. Since $\mathbb{E}_{\mu_{n_j}}[f_i(|X|)]$ is weakly continuous, $\mathbb{E}_{\mu_{n_j}}[f_i(|X|)] \to \mathbb{E}_{\mu_0}[f_i(|X|)]$, hence $\mu_0 \in \Lambda(\overline{\mathbf{b}})$. For n_j sufficiently large, μ_{n_j} is within ϵ of μ_0 and hence the entire set $\Lambda(\overline{\mathbf{b}} + \frac{1}{\mathbf{n}_j})$ is within ϵ of $\Lambda(\overline{\mathbf{b}})$. Taking $\mathbf{b}^* = \overline{\mathbf{b}} + \frac{1}{\mathbf{n}_j}$ completes the proof. □

We are now state the following theorem.

Theorem 5. *The map Λ is upper hemicontinuous at $\overline{\mathbf{b}}$ if and only if there exists $\tilde{\mathbf{b}} > \overline{\mathbf{b}}$ such that $\Lambda(\tilde{\mathbf{b}})$ is compact.*

Proof. By Lemma 3, there exists $\mathbf{b}^* > \overline{\mathbf{b}}$ such that $\Lambda(\mathbf{b}^*) \subseteq \eta_\epsilon(\Lambda(\overline{\mathbf{b}}))$. Suppose that there exists $\tilde{\mathbf{b}} > \overline{\mathbf{b}}$ such that $\Lambda(\tilde{\mathbf{b}})$ is compact. Set $\delta = \min_i\{b_i^* - \overline{b}_i\}$ and suppose that $\mathbf{b}^{(n)} \to \overline{\mathbf{b}}$. Then, eventually $b_i^{(n)} < \overline{b}_i + \delta,\ i = 1, 2, \ldots, m$ which implies that eventually $\mathbf{b}^{(n)} < \mathbf{b}^*$. Hence, there exists $n_0 \in \mathbb{N}$ such that $n > n_0$ implies that

$$\Lambda(\mathbf{b}^{(n)}) \subseteq \Lambda(\mathbf{b}^*) \subseteq \eta_\epsilon(\Lambda(\overline{\mathbf{b}})). \tag{34}$$

To prove the other direction, suppose that Λ is upper hemicontinuous at $\overline{\mathbf{b}}$. Consider the sequence $\overline{\mathbf{b}} + \frac{1}{\mathbf{n}} \to \overline{\mathbf{b}}$. Since Λ is upper hemicontinuous, $\Lambda(\overline{\mathbf{b}} + \frac{1}{\mathbf{n}}) \subseteq \eta_\epsilon(\Lambda(\overline{\mathbf{b}}))$ for n sufficiently large and some $\eta_\epsilon(\Lambda(\overline{\mathbf{b}}))$ is bounded, there exists $\tilde{\mathbf{b}} > \overline{\mathbf{b}}$ such that $\Lambda(\tilde{\mathbf{b}})$ is compact. □

We now turn to obtaining conditions under which the point-to-set map Λ is lower hemicontinuous. We first require the notion of the δ-shrinkage.

Definition 5. *Suppose $\Lambda(\overline{\mathbf{b}})$ is compact and let $\hat{\Lambda}(\overline{\mathbf{b}}) = \{\mu \in \Lambda(\overline{\mathbf{b}}) : \mathbb{E}_\mu[f_j(|X|)] = \overline{b}_j, \text{ for some } j\}$. Let $\delta > 0$ and define the δ-shrinkage of $\Lambda(\overline{\mathbf{b}})$ as*

$$\Lambda_\delta(\overline{\mathbf{b}}) = \{\mu \in \Lambda(\overline{\mathbf{b}}) : \rho(\mu, \hat{\Lambda}(\overline{\mathbf{b}})) \geq \delta\}, \tag{35}$$

where ρ is the Lévy-Prokhorov metric.

To prove lower hemicontinuity of Λ, the first step is to obtain a characterization of the sets $\eta_\epsilon(\Lambda(\mathbf{b}^{(n)}))$.

Lemma 4. *Suppose $\Lambda(\overline{\mathbf{b}})$ is compact. $\mathcal{I}_{\overline{\mathbf{b}}} \neq \emptyset$, $\mathrm{cl}(\mathcal{I}_{\overline{\mathbf{b}}}) = \Lambda(\overline{\mathbf{b}})$. Then, given $\epsilon > 0$, there exists $\delta > 0$ such that $\Lambda(\overline{\mathbf{b}}) \subseteq \eta_\epsilon(\Lambda_\delta(\overline{\mathbf{b}}))$.*

Proof. Denote ∂A as the boundary of the set A. Cover $\hat{\Lambda}(\overline{\mathbf{b}})$ with a finite number of $\epsilon/4$ radius spheres, each sphere centered on a point in $\hat{\Lambda}(\overline{\mathbf{b}})$. Call the j-th sphere N_j. Note that $N_j \cap \mathcal{I}_{\overline{\mathbf{b}}} \neq \emptyset$ since $\mathrm{cl}(\mathcal{I}_{\overline{\mathbf{b}}}) = \Lambda(\overline{\mathbf{b}})$. Let $\mu_j \in N_j \cap \mathcal{I}_{\overline{\mathbf{b}}}$, $\delta_j = \rho(\mu_j, \hat{\Lambda}(\overline{\mathbf{b}}))$, $\delta = \min_j \delta_j$. Observe that $0 \leq \delta \leq \epsilon/4$.

Now suppose that $\mu_0 \in \hat{\Lambda}(\overline{\mathbf{b}})$. Then, $\mu_0 \in N_j$ for some j. Since $\rho(\mu_j, \hat{\Lambda}(\overline{\mathbf{b}})) = \delta_j \geq \delta$, $\mu_j \in \Lambda_\delta(\overline{\mathbf{b}})$. Either $\mu_j \in \partial\Lambda_\delta(\overline{\mathbf{b}})$ or the line from μ_j to the center of N_j must pierce $\partial\Lambda_\delta(\overline{\mathbf{b}})$. In either case, there exists $\mu_j' \in N_j$ such that $\mu_j' \in \partial\Lambda_\delta(\overline{\mathbf{b}})$. Since both μ_0 and μ_j' are in the same $\epsilon/4$-sphere, an ϵ-sphere about μ_j' contains μ_0. Since $\Lambda_\delta(\overline{\mathbf{b}})$ is closed, $\partial\Lambda_\delta(\overline{\mathbf{b}}) \subseteq \Lambda_\delta(\overline{\mathbf{b}})$ and it follows that $\hat{\Lambda}(\overline{\mathbf{b}}) \subseteq \eta_\epsilon(\partial\Lambda_\delta(\overline{\mathbf{b}})) \subseteq \eta_\epsilon(\Lambda_\delta(\overline{\mathbf{b}}))$.

To complete the proof, suppose that $\mu_0 \in \mathcal{I}_{\overline{\mathbf{b}}}$. If $\rho(\mu_0, \hat{\Lambda}(\overline{\mathbf{b}})) \geq \delta$, $\mu_0 \in \Lambda_\delta(\overline{\mathbf{b}})$, and hence $\mu_0 \in \eta_\epsilon(\Lambda_\delta(\overline{\mathbf{b}}))$. Suppose $\rho(\mu_0, \hat{\Lambda}(\overline{\mathbf{b}})) < \delta$. Then there exists $\hat{\mu} \in \hat{\Lambda}(\overline{\mathbf{b}})$ such that $\rho(\mu_0, \hat{\mu}) < \delta$. The point $\hat{\mu}$ must be covered by some N_j, and there is also a $\mu_j' \in N_j$ where $\mu_j \in \partial\Lambda_\delta(\overline{\mathbf{b}})$. Then, $\rho(\mu_j', \mu_0) \leq \rho(\mu_j', \hat{\mu}) + \rho(\hat{\mu}, \mu_0) \leq \epsilon/2 + \delta < \epsilon$. It then follows that $\mu_0 \in \eta_\epsilon(\partial\Lambda_\delta(\overline{\mathbf{b}})) \subseteq \eta_\epsilon(\Lambda_\delta(\overline{\mathbf{b}}))$ and hence $\mathcal{I}_{\overline{\mathbf{b}}} \subseteq \eta_\epsilon(\Lambda_\delta(\overline{\mathbf{b}}))$. □

We now prove conditions under which Λ is lower hemicontinuous.

Theorem 6. *Suppose $\Lambda_{\overline{\mathbf{b}}}$ is compact and $\mathcal{I}_{\overline{\mathbf{b}}} \neq \emptyset$. Then the mapping Λ is lower hemicontinuous at $\overline{\mathbf{b}}$ if and only if $\mathrm{cl}(\mathcal{I}_{\overline{\mathbf{b}}}) = \Lambda(\overline{\mathbf{b}})$.*

Proof. Suppose $\mathrm{cl}(\mathcal{I}_{\overline{\mathbf{b}}}) = \Lambda(\overline{\mathbf{b}})$. By Lemma 4, there exists a δ such that $\Lambda(\overline{\mathbf{b}}) \subseteq \eta_\epsilon(\Lambda_\delta(\overline{\mathbf{b}}))$. Define

$$\begin{aligned} J_1 &= \{j : \mathbb{E}_\mu[f_j(|X|)] = \overline{b}_j,\ \text{some } \mu \in \Lambda(\overline{\mathbf{b}})\} \\ J_2 &= \{j : \mathbb{E}_\mu[f_j(|X|)] < \overline{b}_j,\ \text{all } \mu \in \Lambda(\overline{\mathbf{b}})\}. \end{aligned} \tag{36}$$

Then, $J_1 \cup J_2 = \{1, 2, \ldots, m\}$. Consider the convergent sequence $\mathbf{b}^{(n)} \to \overline{\mathbf{b}}$ with $n > 0$, and let $\Lambda_{nj} = \{\mu : \mathbb{E}_\mu[f_j(|X|)] \leq b_j^{(n)}\}$.

Consider the case $j \in J_1$. Note that $\Lambda_\delta(\overline{\mathbf{b}})$ is compact and $\Lambda_\delta(\overline{\mathbf{b}}) \subseteq \mathcal{I}_{\overline{\mathbf{b}}} \subseteq \Lambda(\overline{\mathbf{b}})$. Define $b_j^* = \max_{\mu \in \Lambda_\delta(\overline{\mathbf{b}})} \mathbb{E}_\mu[f_j(|X|)]$. Since $\Lambda_\delta(\overline{\mathbf{b}}) \subseteq \mathcal{I}_{\overline{\mathbf{b}}}$, $b_j^* < \overline{b}_j$. Let $\Lambda_{*j} = \{\mu : \mathbb{E}_\mu[f_j(|X|)] \leq b_j^*\}$. It then follows that $\mu \in \Lambda_\delta(\overline{\mathbf{b}})$, which implies that $\mathbb{E}_\mu[f_j(|X|)] \leq b_j^*$ for each $j \in J_1$, so $\Lambda_\delta(\overline{\mathbf{b}}) \subseteq \cap_{j \in J_1} \Lambda_{*j}$. But since $\mathbf{b}^{(n)} \to \overline{\mathbf{b}}$, for all n sufficiently large, $\Lambda_\delta(\overline{\mathbf{b}}) \subseteq \cap_{j \in J_1} \Lambda_{nj}$.

Now consider $j \in J_2$ and define $b_j^* = \max_{\mu \in \Lambda(\overline{\mathbf{b}})} \mathbb{E}_\mu[f_j(|X|)]$. By the definition of J_2, $b_j^* < \overline{b}_j$ for each $j \in J_2$ and hence $\Lambda(\overline{\mathbf{b}}) \subseteq \cap_{j \in J_2} \Lambda_{*j}$. Since $\mathbf{b}^{(n)} \to \overline{\mathbf{b}}$, for n sufficiently large $b_{nj} > b_j^*$ and hence $\Lambda_{*j} \subseteq \Lambda_{nj}$ for each $j \in J_2$. As such, $\Lambda_\delta(\overline{\mathbf{b}}) \subseteq \Lambda(\overline{\mathbf{b}}) \subseteq \cap_{j \in J_2} \Lambda_{*j} \subseteq \cap_{j \in J_2} \Lambda_{nj}$. This means that

$$\Lambda_\delta(\overline{\mathbf{b}}) \subseteq \cap_{j \in J_1 \cup J_2} \Lambda_{nj} = \Lambda(\mathbf{b}^{(n)}) \tag{37}$$

for all n sufficiently large and hence eventually $\Lambda(\overline{\mathbf{b}}) \subseteq \eta_\epsilon(\Lambda(\mathbf{b}^{(n)}))$ and Λ is lower hemicontinuous at $\overline{\mathbf{b}}$.

Conversely, suppose Λ is lower hemicontinuous at $\overline{\mathbf{b}}$. Clearly, $\mathrm{cl}(\mathcal{I}_{\overline{\mathbf{b}}}) \subseteq \Lambda(\overline{\mathbf{b}})$. It therefore remains to show that $\Lambda(\overline{\mathbf{b}}) \subseteq \mathrm{cl}(\mathcal{I}_{\overline{\mathbf{b}}})$. Let $\overline{\mu} \in \Lambda(\overline{\mathbf{b}})$. If $\overline{\mu} \in \mathcal{I}_{\overline{\mathbf{b}}}$, then $\overline{\mu} \in \mathrm{cl}(\mathcal{I}_{\overline{\mathbf{b}}})$. Suppose $\overline{\mu} \notin \mathcal{I}_{\overline{\mathbf{b}}}$ and select $\epsilon > 0$. Since $\mathcal{I}_{\overline{\mathbf{b}}} \neq \emptyset$, there exists a sequence $\overline{\mathbf{b}} - \frac{1}{\mathbf{n}}$ which eventually lies in $\mathcal{B}$. Since Λ is lower hemicontinuous at $\overline{\mathbf{b}}$, $\Lambda(\overline{\mathbf{b}}) \subseteq \eta_\epsilon(\Lambda(\overline{\mathbf{b} - \frac{1}{\mathbf{n}}}))$, $n > n_0$. This means that there exists a $\mu' \in N_\epsilon(\overline{\mu})$, and since $\mu' \in \Lambda(\overline{\mathbf{b}} - \frac{1}{\mathbf{n}})$, μ' is also in $\mathcal{I}_{\overline{\mathbf{b}}}$. As such, in every neighborhood of $\overline{\mu}$, there exists $\mu' \in \mathcal{I}_{\overline{\mathbf{b}}}$, which implies that $\overline{\mu} \in \mathrm{cl}(\mathcal{I}_{\overline{\mathbf{b}}})$. □

Finally, we require the following theorem from [19].

Theorem 7. *If $I(X;Y)$ is weakly upper semicontinuous and Λ is upper hemicontinuous at $\overline{\mathbf{b}}$, then the capacity C in (8) is weakly upper semicontinuous at $\overline{\mathbf{b}}$. Similarly, if $I(X;Y)$ is weakly lower semicontinuous and Λ is lower hemicontinuous at $\overline{\mathbf{b}}$, then the capacity C in (8) is weakly lower semicontinuous at $\overline{\mathbf{b}}$.*

The desired result then follows by applying Theorem 5 and Theorem 6 in Theorem 7.

C Proof of Lemma 1

Consider the class of optimization problems denoted by (P_u) in the form

$$\begin{aligned} \min_{x \in X} \quad & f(x, u) \\ \text{subject to} \quad & x \in \Phi(u), \end{aligned} \tag{38}$$

where X is a Banach space and u lies in a metric space U. The constraint set is restricted to have the finitely constrained form

$$\Phi(u) = \{x : g_i(x) \leq u, \ i = 1, 2, \ldots, q\}. \tag{39}$$

We denote the set of solutions to (P_u) as $S(u)$.

Our proof relies on the following theorem given in [11, Theorem 4.3], which applies to problem (38).

Theorem 8. *Let $u^0 \in U$. Suppose that (i) the problem (P_{u^0}) is convex, (ii) the optimal set $S(u^0)$ is non-empty and compact, (iii) the directional regularity condition holds for all $x^0 \in S(u^0)$, and (iv) for sufficiently small $t > 0$ the program (P_{u^0+td}) possesses an $o(t)$-optimal solution $x(t)$ such that $\mathrm{dist}(x(t), S(u^0)) \to 0$ as $t \to 0^+$. Then the optimal value function is directionally differentiable at u^0 in the unit direction d and*

$$v'(u^0, d) = \inf_{x \in S(u^0)} \sup_{\lambda \in \Lambda(u^0)} D_u L(x, \lambda, u^0) d. \tag{40}$$

It follows that if we can show that Theorem 8 holds for the capacity optimization problem, then we obtain the desired result. To proceed, note that the mutual information is a convex functional of the input probability measure μ and that the constraint set $\Lambda(\overline{\mathbf{b}})$ is a convex set, which ensures that the condition (i) is satisfied. Moreover, the optimal input probability measure μ^* exists and is unique by assumption, which yields condition (ii).

To prove condition (iii), we use the fact that if Slater's condition for the problem (P_{u^0}), then the directional regularity condition is guaranteed to hold [11, pg. 17]. In our setting, Slater's condition states that there exists a probability measure $\overline{\mu} \in \Lambda(\overline{\mathbf{b}})$ such that $\mathbb{E}_{\overline{\mu}}[f_i(|X|)] < \overline{b}_i,\ i = 1, 2, \ldots, m$. Clearly, Slater's condition holds as each f_i is positive and non-decreasing.

Condition (iv) is a consequence of the fact that the constraint perturbations are finitely constrained and [11, Theorem 4.2]. The result then follows from using the identity $D_{\mathbf{d}} L(\mu^*, \lambda; \overline{\mathbf{b}}) = \nabla L(\mu^*, \lambda; \overline{\mathbf{b}}) \cdot \mathbf{d}$.

D Proof of Theorem 3

We apply the same argument as in the proof of Berge's maximum theorem, restricted to the sequence $\{p_N^i\}_{i=1}^{\infty}$. Let $\Lambda^*(p_N^i)$ be the set of optimal input distributions corresponding to a noise probability density function p_N^i. To prove C is continuous at p_N^0, consider the sequence $\{p_N^i\}_{i=1}^{\infty}$ which converges to p_N^0. We wish to show that $C(p_N^i) \to C(p_N^0)$. Observe that $C(p_N^i)$ has a subsequence $C(p_N^{i_k}) \to \limsup_{i \to \infty} C(p_N^i)$. Now pick any $\mu_{i_k} \in \Lambda^*(p_N^{i_k})$ so that $C(p_N^{i_k}) = I(\mu_{i_k}, p_N^{i_k})$ for each i_k. Since Λ^* is compact valued and upper hemicontinuous at μ_0^*, we can find a subsequence of $(\mu_{i_k}^*)$ that converges to a point μ_0^* in $\Lambda^*(p_N^0)$. Using condition (**C5**), then

$$C(p_N^{i_k}) = I(\mu_{i_k}^*, p_N^{i_k}) \overset{i_k \to \infty}{\to} I(\mu_0^*, p_N^0) = C(p_N^0) \tag{41}$$

which proves that $C(p_N^0) = \limsup_{i \to \infty} C(p_N^i)$. But the same argument also shows that $C(p_N^0) = \liminf_{i \to \infty} C(p_N^i)$, which completes the proof.

E Proof of Theorem 4

We seek an upper bound on $|C(p_N^0) - C(p_N^1)|$. To this end, observe that

$$\begin{aligned} C(p_N^0) - C(p_N^1) &= I(\mu_0^*, p_N^0) - I(\mu_1^*, p_N^1) \\ &= I(\mu_0^*, p_N^0) - I(\mu_0^*, p_N^1) + I(\mu_0^*, p_N^1) - I(\mu_1^*, p_N^1) \\ &\leq I(\mu_0^*, p_N^0) - I(\mu_0^*, p_N^1), \end{aligned} \tag{42}$$

which follows from the fact that μ_1^* maximizes the mutual information for the noise distribution p_N^1. Similarly,

$$\begin{aligned} C(p_N^0) - C(p_N^1) &= I(\mu_0^*, p_N^0) - I(\mu_1^*, p_N^0) + I(\mu_1^*, p_N^0) - I(\mu_1^*, p_N^1) \\ &\geq I(\mu_1^*, p_N^0) - I(\mu_1^*, p_N^1). \end{aligned} \tag{43}$$

It then follows that

$$|C(p_N^0) - C(p_N^1)| \leq \max\{|I(\mu_0^*, p_N^0) - I(\mu_0^*, p_N^1)|, |I(\mu_1^*, p_N^0) - I(\mu_1^*, p_N^1)|\}. \tag{44}$$

Let $0 \leq \theta \leq 1$ and define $q_N^i(\theta) = (1-\theta)p_N^0 + \theta p_N^i$. By hypothesis, we then have for each i, $M_i < \infty$ and $N_i < \infty$ such that

$$\begin{aligned} \left| \lim_{\theta \to 0^+} \frac{I(\mu_0^*, q_N^i) - I(\mu_0^*, p_N^0)}{\theta \|p_N^0 - p_N^i\|_{TV}} \right| &= M_i, \\ \left| \lim_{\theta \to 0^+} \frac{I(\mu_1^*, p_N^0) - I(\mu_1^*, q_N^i)}{\theta \|p_N^0 - p_N^i\|_{TV}} \right| &= N_i, \end{aligned} \tag{45}$$

This implies that

$$\begin{aligned} |I(\mu_0^*, q_N^i) - I(\mu_0^*, p_N^0)| &= M_i \theta \|p_N^0 - p_N^i\|_{TV} + o(\theta) \\ |I(\mu_1^*, p_N^0) - I(\mu_1^*, q_N^i)| &= N_i \theta \|p_N^0 - p_N^i\|_{TV} + o(\theta). \end{aligned} \tag{46}$$

Since $M = \sup_i M_i < \infty$ and $N = \sup_i N_i < \infty$, it then follows that for all i

$$\begin{aligned} |I(\mu_0^*, q_N^i) - I(\mu_0^*, p_N^0)| &\leq M \theta \|p_N^0 - p_N^i\|_{TV} + o(\theta) \\ |I(\mu_1^*, p_N^0) - I(\mu_1^*, q_N^i)| &\leq N \theta \|p_N^0 - p_N^i\|_{TV} + o(\theta). \end{aligned} \tag{47}$$

The result then follows by using the above inequalities with (44).

F Verification of (C7) for Truncated α-Stable Noise

We verify that condition (**C7**) in Theorem 4 holds for the channels in Section 5; namely, truncated α-stable noise channels subject to a fractional moment constraint $\mathbb{E}_\mu[|X|^r] \leq c$ with $0 < r < \alpha$. Fix the parameters α, γ and choose the sequence of probability density functions $\{p_N^i\}_{i=1}^\infty$ to be the sequence of level i truncations defined in (20), where p_N^0 is the probability density function of symmetric α-stable noise with scale parameter γ.

We proceed in two steps. First, we establish the existence of certain Gâteaux differentials which in turn guarantee M_i and N_i in Theorem 4 are finite for fixed i. We then prove that $M = \sup_i M_i$ and $N = \sup_i N_i$ are finite.

In this appendix, we denote the output probability density function for Y as $p_Y(y; p_N, \mu)$, where p_N is the noise probability density function and μ is the input probability measure.

F.1 Existence of Gâteaux Differentials

Fix $i \geq 1$ and for $0 \leq \theta \leq 1$ define $q_{N,\theta}^i = (1-\theta)p_N^0 + \theta p_N^i$. Consider the Gâteaux differential

$$\lim_{\theta \to 0^+} \frac{I(\mu_0^*, q_{N,\theta}^i) - I(\mu_0^*, p_N^0)}{\theta} = G_{1,i}. \tag{48}$$

The first step is to show that $G_{i,1} < \infty$. By the definition of mutual information,

$$\begin{aligned} G_{i,1} = \lim_{\theta \to 0^+} \Bigg[&- \int_{-\infty}^{\infty} p_Y(y; q_{N,\theta}^i, \mu_0^*) \log p_Y(y; q_{N,\theta}^i, \mu_0^*) \mathrm{d}y + \int_{-\infty}^{\infty} q_{N,\theta}^i(y) \log q_{N,\theta}^i(y) \mathrm{d}y \\ &+ \int_{-\infty}^{\infty} p_Y(y; p_N^0, \mu_0^*) \log p_Y(y; p_N^0, \mu_0^*) \mathrm{d}y - \int_{-\infty}^{\infty} p_N^0(y) \log p_N^0(y) \mathrm{d}y \Bigg]. \end{aligned} \tag{49}$$

Using the fact that the continuity condition in (**C5**) for the mutual information holds and L'Hôpital's rule,

$$M_i = \lim_{\theta\to 0^+} -\left[\int_{-\infty}^{\infty} p_Y(y; q^i_{N,\theta}, \mu^*_0)\log p_Y(y; q_{N,\theta}, \mu^*_0)\mathrm{d}y - \int_{-\infty}^{\infty} q^i_{N,\theta}(y)\log q^i_{N,\theta}(y)\mathrm{d}y\right]'. \quad (50)$$

Using the definition of the derivative

$$\begin{aligned}&\left[\int_{-\infty}^{\infty} p_Y(y; q^i_{N,\theta}, \mu^*_0)\log p_Y(y; q_{N,\theta}, \mu^*_0)\mathrm{d}y - \int_{-\infty}^{\infty} q^i_{N,\theta}(y)\log q^i_{N,\theta}(y)\mathrm{d}y\right]'\\ &= \lim_{h\to 0}\left[\frac{\int_{-\infty}^{\infty} p_Y(y; q^i_{N,\theta+h}, \mu^*_0)\log p_Y(y; q_{N,\theta+h}, \mu^*_0) - q^i_{N,\theta+h}(y)\log q^i_{N,\theta+h}(y)\mathrm{d}y}{h}\right.\\ &\quad\left.-\frac{\int_{-\infty}^{\infty} p_Y(y; q^i_{N,\theta}, \mu^*_0)\log p_Y(y; q_{N,\theta}, \mu^*_0) - q^i_{N,\theta}(y)\log q^i_{N,\theta}(y)\mathrm{d}y}{h}\right] \end{aligned}\quad (51)$$

We now consider the case $h\to 0^+$ and note that the case $h\to 0^-$ can be treated similarly. By the mean value theorem, there exists a $0\le c(h)\le h$ such that

$$\begin{aligned}&\lim_{h\to 0}\left[\frac{\int_{-\infty}^{\infty} p_Y(y; q^i_{N,\theta+h}, \mu^*_0)\log p_Y(y; q_{N,\theta+h}, \mu^*_0) - q^i_{N,\theta+h}(y)\log q^i_{N,\theta+h}(y)\mathrm{d}y}{h}\right.\\ &\left.-\frac{\int_{-\infty}^{\infty} p_Y(y; q^i_{N,\theta}, \mu^*_0)\log p_Y(y; q_{N,\theta}, \mu^*_0) - q^i_{N,\theta}(y)\log q^i_{N,\theta}(y)\mathrm{d}y}{h}\right]\\ &= \lim_{h\to 0^+}\int_{-\infty}^{\infty}\left[p_Y(y; q^i_{N,\theta}, \mu^*_0)\log p_Y(y; q_{N,\theta}, \mu^*_0) - q^i_{N,\theta}(y)\log q^i_{N,\theta}(y)\right]'_{\theta+c(h)}\mathrm{d}y\\ &= \lim_{h\to 0^+}\int_{-\infty}^{\infty}\Big[\big(p_Y(y;p^i_N,\mu^*_0) - p_Y(y;p^0_N,\mu^*_0)\big)\log p_Y(y; q^i_{N,\theta+c(h)}, \mu^*_0) + p_Y(y;p^i_N,\mu^*_0) - p_Y(y;p^0_N,\mu^*_0)\\ &-\big(p^i_N(y) - p^0_N(y)\big)\log q^i_{N,\theta+c(h)}(y) + p^i_N(y) - p^0_N(y)\Big]\,\mathrm{d}y\\ &= \int_{-\infty}^{\infty}\big(p_Y(y;p^i_N,\mu^*_0) - p_Y(y;p^0_N,\mu^*_0)\big)\log p_Y(y; q^i_{N,\theta}, \mu^*_0) - \big(p^i_N(y) - p^0_N(y)\big)\log q^i_{N,\theta}(y)\mathrm{d}y,\end{aligned}\quad (52)$$

where the last inequality follows from the dominated convergence theorem. In particular, observe that

$$|(p_Y(y;p^i_N,\mu^*_0) - p_Y(y;p^0_N,\mu^*_0))\log p_Y(y;q^i_{N,\theta+c(h)},\mu^*_0)| \le (p_Y(y;p^i_N,\mu^*_0) + p_Y(y;p^0_N,\mu^*_0))|\log p_Y(y;q^i_{N,\theta+c(h)},\mu^*_0)|. \quad (53)$$

We then have

$$\begin{aligned}p(y;q^i_{N,\theta+c(h)},\mu^*_0) &= (1-\theta-c(h))p_Y(y;p^0_N,\mu^*_0) + (\theta+c(h))p_Y(y;p^i_N,\mu^*_0)\\ &\ge (1-\theta-c(h))p_Y(y;p^0_N,\mu^*_0) \ge \frac{1}{2}p_Y(y;p^0_N,\mu^*_0),\end{aligned}\quad (54)$$

whenever $\theta + c(h)\le\frac{1}{2}$, which holds since $\theta + c(h)$ is arbitrarily small. Therefore,

$$|(p_Y(y;p^i_N) + p_Y(y;p^0_N,\mu^*_0))||\log p_Y(y;q^i_{N,\theta+c(h)},\mu^*_0)| \le -(p_Y(y;p^i_N,\mu^*_0) + p_Y(y;p^0_N),\mu^*_0)\log\frac{1}{2}p_Y(y;p^0_N,\mu^*_0). \quad (55)$$

Since $-\int_{-\infty}^{\infty} p_Y(y;p_N^0,\mu_0^*)\log p_Y(y;p_N^0,\mu_0^*) < \infty$, it follows that $-p_Y(y;p_N^0,\mu_0^*)\log p_Y(y;p_N^0,\mu_0^*)$ is integrable. This means that we only need to show that $-p_Y(y;p_N^i,\mu_0^*)\log p_Y(y;p_N^0,\mu_0^*)$ is integrable. To this end, recall that

$$p_N^i(x) = \begin{cases} \frac{p_N^0(x)}{\kappa_i}, & |x| \leq T \\ 0, & \text{otherwise.} \end{cases} \tag{56}$$

As such,

$$\begin{aligned} p_Y(y;p_N^i,\mu_0^*) &= \int_{-\infty}^{\infty} p_N^i(y-x)d\mu_0^*(x) \\ &= \int_{x:|y-x|\leq T} \frac{p_N^0(y-x)}{\kappa_i} d\mu_0^*(x) \\ &\leq \frac{1}{\kappa_i}\int_{-\infty}^{\infty} p_N^0(y-x)d\mu_0^*(x) \\ &= \frac{1}{\kappa_i} p_Y(y;p_N^0,\mu_0^*). \end{aligned} \tag{57}$$

This implies that

$$-p_Y(y;p_N^i,\mu_0^*)\log p_Y(y;p_N^0,\mu_0^*) \leq -\frac{1}{\kappa_i} p_Y(y;p_N^0,\mu_0^*)\log p_Y(y;p_N^0,\mu_0^*), \tag{58}$$

for y sufficiently large, which in turn means that $-p_Y(y;p_N^i,\mu_0^*)\log p(y;p_N^0,\mu_0^*)$ is integrable. Moreover, observe that a similar argument also guarantees the integrability of $-p_N^i(y)\log p_N^0(y)$, which justifies the interchange of the limit and the integral in (52).

Hence,

$$\begin{aligned} G_{1,i} &= \lim_{\theta\to 0^+}\int_{-\infty}^{\infty} \left(p_Y(y;p_N^i,\mu_0^*) - p_Y(y;p_N^0,\mu_0^*)\right)\log p_Y(y;q_{N,\theta}^i,\mu_0^*) - \left(p_N^i(y) - p_N^0(y)\right)\log q_{N,\theta}^i(y)\mathrm{d}y \\ &= \int_{-\infty}^{\infty} \left(p_Y(y;p_N^i,\mu_0^*) - p_Y(y;p_N^0,\mu_0^*)\right)\log p_Y(y;p_N^0,\mu_0^*) - \left(p_N^i(y) - p_N^0(y)\right)\log p_N^0(y)\mathrm{d}y, \end{aligned} \tag{59}$$

where the interchange of the limit and the integral again follows from the dominated convergence theorem.

Since for a fixed i, $\|p_N^0 - p_N^i\|_{TV} < \infty$ it then follows that $M_i < \infty$. A similar argument guarantees

$$G_{2,i} = \lim_{\theta\to 0^+}\frac{I(\mu_1^*,p_N^0) - I(\mu_1^*,q_N^i)}{\theta} < \infty, \tag{60}$$

which in turn implies that $N_i < \infty$.

F.2 Finiteness of M and N

The last step is to show that $\sup_i M_i < \infty$ and $\sup_i N_i < \infty$. To do this, we show that $G_{1,i} = O(i^{-\alpha})$ and $G_{2,i} = O(i^{-\alpha})$. We then show that[1] $\|p_N^0 - p_N^i\|_{TV} = \Omega(i^{-\alpha})$, which yields the desired result.

[1]We say that $f(x) = \Omega(g(x))$ if and only if $\exists k > 0, c > 0$ such that $|f(x)| \geq \kappa|g(x)|,\ \forall |x| \geq c$.

To show that $G_{k,i} = O(i^{-\alpha}),\ k = 1,2$, we consider the two integrals in (59). Let $\mu_k^* \in \{\mu_0^*, \mu_1^*\}$ and consider

$$\begin{aligned} I_1 &= \left| \int_{-\infty}^{\infty} (p_Y(y; p_N^i, \mu_k^*) - p_Y(y; p_N^0, \mu_k^*)) \log p_Y(y; p_N^0, \mu_k^*) \mathrm{d}y \right| \\ &= \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (p_N^i(y-x) - p_N^0(y-x)) d\mu_k^*(x) \log p_Y(y; p_N^0, \mu_k^*) \mathrm{d}y \right| \\ &\leq \left| \frac{1}{\kappa_i} - 1 \right| \int_{-\infty}^{\infty} p_Y(y; p_N^0, \mu_k^*) |\log p_Y(y; p_N^0, \mu_k^*)| \mathrm{d}y \\ &= O(i^{-\alpha}). \end{aligned} \tag{61}$$

Similarly,

$$\begin{aligned} I_2 &\leq \left| \frac{1}{\kappa_i} - 1 \right| \left| \int_{-\infty}^{\infty} p_N^0(y) \log p_N^0(y) \mathrm{d}y \right| \\ &= O(i^{-\alpha}). \end{aligned} \tag{62}$$

To complete the proof, we show that $1 - \kappa_T = \Omega(T^{-\alpha})$. Indeed, there exists a $c_l > 0$ such that

$$\begin{aligned} 1 - \kappa_T &\geq 2 \int_{y > T} \frac{c_l}{y^{-1-\alpha}} \mathrm{d}y \\ &= 2 K_\alpha T^{-\alpha}, \end{aligned} \tag{63}$$

with $K_\alpha > 0$. This implies that $\sup_i M_i < \infty$ and $\sup_i N_i < \infty$, as required.

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

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

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

[14] 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.

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

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

[17] 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.

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

[19] J. Evans and F. Gould, “Stability in nonlinear programming,” *Operations Research*, vol. 18, no. 1, pp. 107–118, 1970.

RESEARCH CENTRE
GRENOBLE – RHÔNE-ALPES

Inovallée
655 avenue de l'Europe Montbonnot
38334 Saint Ismier Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399