On Capacity Sensitivity in Additive Vector Symmetric α-Stable Noise Channels

Malcolm Egan

On Capacity Sensitivity in Additive Vector Symmetric α-Stable Noise Channels

Malcolm Egan

***Abstract*—Due to massive numbers of uncoordinated devices present in wireless networks for the Internet of Things (IoT), interference is a key challenge. There is evidence both from experiments and analysis of statistical models that the uncoordinated nature of channel access leads to non-Gaussian statistics for the interference. A particularly attractive model in this scenario is the additive vector α-stable noise channel. In this paper, we study the capacity of this channel with fractional moment constraints. In particular, we establish well-posedness of the optimization problem for the capacity. We also study convergence of the capacity loss due to an additional constraint where input probability measures are concentrated on spherical shells, in addition to the fractional moment constraints.**

I. INTRODUCTION

A key challenge for wireless network deployments in the Internet of Things (IoT) is their scale. Even if devices operate at low power levels, interference can degrade the performance of other communication networks. As a consequence, guaranteeing coexistence is a difficult problem. Moreover, there is a need to obtain statistical characterizations of the interference, which has lead to a number of experimental studies [1]–[4].

Along with the theoretical work in [5]–[8], these experimental studies—in particular, [1] in the context of low power wide area (LPWA) networks—have revealed that interference in IoT networks is impulsive. In particular, high amplitude interference is significantly more likely than in Gaussian models. As a consequence, for network dimensioning, signal processing, and resource allocation it is necessary to move beyond Gaussian interference models.

In fact, numerous studies exploiting point process models for device locations have revealed the heavy-tailed nature of the interference. In [7], the interference statistics were studied for a general class of point process models. In the limiting case—where devices are located according to a Poisson point process, the guard zone radius tends to zero and the network radius tends to infinity—the marginal distribution for the interference on each given band is α-stable.

Due to the fact that several IoT protocols exploit narrowband transmissions over multiple frequencies, there is in general dependence between the interference on each band. In the case where each device (located according to a Poisson point process) transmits on the same set of bands, the interference vector is known to follow a sub-Gaussian α-stable distribution with uncorrelated underlying Gaussian random vector [8]. For more general classes of access protocols, there are presently few analytical characterizations of the interference random vector.

To model the statistical dependence between interference of different bands for general access protocols, copula models with α-stable marginals have been proposed in [9]–[11]. Copula models provide a very general framework in order to study interference in IoT networks. However, their generality makes an analytical study of network performance challenging and full characterizations of bit error rates and capacity are presently unknown.

In this paper, we study communication channels in IoT networks in the presence of interference vectors following vector α-stable distributions. This class of distributions contains the sub-Gaussian α-stable distributions in [8] and captures a wide range of dependence structures, with the benefit of being significantly more tractable than the copula models in [9] from an analytical point of view.

In particular, the main contributions in this paper concern the capacity of point-to-point memoryless and stationary vector α-stable noise channels. We first show that the capacity problem is well posed, admitting a unique optimal input distribution. This result generalizes known results for scalar α-stable noise channels, as well as vector Gaussian noise channels.

We then turn to the impact of constraining the input to lie on spherical shells, which generalizes the notion of discrete inputs in scalar channels. In order to study this problem, we show that it can be formalized within the capacity sensitivity framework developed in [12], [13]. In particular, the capacity sensitivity quantifies the effect of changing channel parameters or the constraint set. Extending results recently developed for scalar noise channels [13], we study convergence of the capacity sensitivity as the gap between each spherical shell tends to zero. This provides insight into approximations of the capacity subject to spherical shell constraints via the capacity with only moment constraints. The convergence result also provides a basis for further estimates of the capacity sensitivity, as was demonstrated in the case of scalar channels in [13].

A. Notation

Vectors are denoted by bold lowercase letters and random vectors by bold uppercase letters, respectively (e.g., $\mathbf{x}$, $\mathbf{X}$). We denote the distribution of a random vector $\mathbf{X}$ by $P_{\mathbf{X}}$. For vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, $\mathbf{x} \succeq \mathbf{y}$ if $x_i \geq y_i,\ i = 1, 2, \ldots, d$.

M. Egan is with the CITI laboratory, a joint laboratory between Université de Lyon, INSA Lyon and INRIA, France.

II. Problem Formulation

A. Preliminaries

Before studying vector symmetric α-stable channels, we overview key properties of α-stable random vectors that will be used in the following. The probability density function of an α-stable random variable is described by four parameters: the exponent $0 < \alpha \leq 2$; the scale parameter $\gamma \in \mathbb{R}_+$; the skew parameter $\beta \in [-1, 1]$; and the shift parameter $\delta \in \mathbb{R}$. As such, a common notation for an α-stable random variable X is $X \sim S_\alpha(\gamma, \beta, \delta)$. In the case $\beta = \delta = 0$, X is said to be a symmetric α-stable random variable.

In general, α-stable random variables do not have closed-form probability density functions. Instead, they are more compactly represented by their characteristic function, given by [14, Eq. 1.1.6]

$$\mathbb{E}[e^{i\theta X}] = \begin{cases} \exp\left\{-\gamma^\alpha |\theta|^\alpha (1 - i\beta (\text{sign}\theta) \tan \frac{\pi\alpha}{2}) + i\delta\theta\right\}, \\ \qquad \alpha \neq 1 \\ \exp\left\{-\gamma|\theta|(1 + i\beta \frac{2}{\pi} (\text{sign}\theta) \log|\theta|) + i\delta\theta\right\}, \\ \qquad \alpha = 1 \end{cases} \tag{1}$$

It is possible to extend the notion of an α-stable random variable to the multivariate setting. In particular, a random vector $\mathbf{X}$ is a symmetric α-stable random vector if for all $a, b > 0$ there exists $c > 0$ such that

$$a\mathbf{X}^{(1)} + b\mathbf{X}^{(2)} \stackrel{d}{=} c\mathbf{X}, \tag{2}$$

where $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}$ are independent copies of $\mathbf{X}$.

A sufficient condition for a random vector $\mathbf{X}$ in $\mathbb{R}^d$ to be a symmetric α-stable random vector is that all linear combinations of the elements of $\mathbf{X}$ are symmetric α-stable [14]. In general, d-dimensional symmetric α-stable random vectors are also represented via their characteristic function, given by [14]

$$\mathbb{E}[e^{i\boldsymbol{\theta}\cdot\mathbf{X}}] = \exp\left(-\int_{\mathbb{S}^{d-1}} \left|\sum_{k=1}^{d} \theta_k s_k\right|^\alpha \Gamma(d\mathbf{s})\right), \tag{3}$$

where Γ is the unique symmetric measure on the surface of the d-dimensional unit sphere.

In the case that a d-dimensional symmetric α-stable random vector $\mathbf{X}$ is *truly* d-dimensional, there exists a joint probability density function $p_{\mathbf{X}}(\mathbf{x})$ on $\mathbb{R}^d$. Note that a simple necessary and sufficient condition for $\mathbf{X}$ to be truly d-dimensional is for the support of the spectral measure Γ to span $\mathbb{R}^d$ [15]. This condition means that degenerate symmetric α-stable random vectors (e.g., when $X_i = X_j$ for some $i \neq j$, $i, j \in \{1, \ldots, d\}$) are not considered.

B. Communication Channel and Capacity

In this paper, we focus on memoryless, stationary, linear and point-point communication channels which have the additive vector symmetric α-stable noise structure ($1 < \alpha < 2$)

$$\mathbf{Y} = \mathbf{X} + \mathbf{N}, \tag{4}$$

where $\mathbf{N}$ is a *truly* d-dimensional symmetric α-stable random vector, admiting a multivariate probability density function $p_{\mathbf{N}}$, with $\mathbf{X}$ and $\mathbf{N}$ independent. This scenario models an interference-limited regime where the effect of thermal noise is negligible. Moreover, $\mathbf{X}$ is a random vector on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ satisfying the constraint

$$[\mathbb{E}[|X_1|^r], \ldots, \mathbb{E}[|X_n|^r]]^T \preceq \mathbf{c}, \tag{5}$$

where $1 < r < \alpha$.

Let $\mathcal{P}$ be the set of probability measures on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. By generalizations of Shannon's noisy channel coding theorem to vector non-Gaussian channels [16], the capacity may be interpreted as the maximum achievable rate with asymptotically zero average probability of error and is defined by

$$\begin{aligned} C = &\sup_{\mu \in \mathcal{P}} \quad && I(\mathbf{X}; \mathbf{Y}) \\ &\text{subject to} && \mu \in \Lambda, \end{aligned} \tag{6}$$

where

$$\Lambda = \{\mu \in \mathcal{P} : [\mathbb{E}_\mu[|X_1|^r], \ldots, \mathbb{E}_\mu[|X_n|^r]]^T \preceq \mathbf{c}\}, \tag{7}$$

with $1 < r < \alpha$. The set Λ constrains the fractional moment for each marginal of the input $\mathbf{X}$. This choice of constraint set may be justified in practice by viewing Λ as a relaxation of amplitude and fractional moment constraints, which guarantee finite power. Moreover, the formulation of the problem in (6) admits more tractable bounds in the special cases studied in [8], [17]

Amplitude constraints may be incorporated by restricting the support of the input distribution to lie on spherical shells. This additional constraint is a natural generalization of requiring discreteness for inputs in scalar channels. To this end, let $\Delta > 0$ and define a spherical shell in $\mathbb{R}^d$ by the set

$$S_r = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_2 = r\}. \tag{8}$$

To impose the support constraint on the input to ensure the input is restricted to spherical shells, we introduce the set $\mathcal{P}_\Delta$. In particular, $\mathcal{P}_\Delta$ consists of all probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ that lie on spherical shells S_r with radii in the set $\cup_{\Delta' > \Delta} \Delta'\mathbb{Z}$ where $\Delta\mathbb{Z} = \{r \in \mathbb{R}_{\geq 0} : r = 0, \pm\Delta, \pm 2\Delta, \ldots\}$.

Accounting for the spherical shell constraint, the capacity of the channel in (4) is given by

$$\begin{aligned} C_\Delta = &\sup_{\mu \in \mathcal{P}} \quad && I(\mathbf{X}; \mathbf{Y}) \\ &\text{subject to} && \mu \in \Lambda_\Delta, \end{aligned} \tag{9}$$

where

$$\Lambda_\Delta = \Lambda \cap \mathcal{P}_\Delta. \tag{10}$$

C. Motivation for the IoT

To motivate the vector symmetric α-stable noise in the context of the IoT, consider the scenario of an infinite radius network of interfering devices, with guard bands of radius zero. The locations of the interfering devices form a homogeneous Poisson point process with intensity λ. Each interfering device transmits over a subset of orthogonal frequency bands $\mathcal{B} = \{1, 2, \ldots, K\}$.

At each time t, interfering devices independently transmit on a band $k \in \mathcal{B}$ with probability p. This probability p can be interpreted as being proportional to the quantity of data that must be transmitted by each device.

The interference received at the origin at time t on frequency k is given by

$$N_k(t) = \sum_{j \in \Phi_k(t)} r_j(t)^{-\eta/2} h_{j,k}(t) x_{j,k}(t), \tag{11}$$

where r_j is the distance from device $j \in \Phi_k(t)$ (the set of devices active on band k during time t), η is the path loss exponent, $h_{j,k}(t) \sim \mathcal{CN}(0,1)$, and $x_{j,k}(t)$ is the baseband emission.

Under this model, it is known (see e.g., [8]) that the real and imaginary parts of the interference at time t on each band is symmetric α-stable. Moreover, due to the fact that devices randomly access each band $k \in \mathcal{B}$, there is non-trivial statistical dependence between the interference on each band. This motivates the need for methods to model the dependence, of which symmetric α-stable random vectors are a general yet tractable choice.

III. Well-Posedness of the Capacity Problem

In this section, we show that the optimization problem for the capacity in (6) is well-posed in the sense that there exists a unique probability measure in Λ such that the supremum can be achieved. This problem has been widely studied in the context of point-to-point *scalar* additive noise channels [18], [19], but there are far fewer results in the vector channel scenario.

We first establish the existence of a capacity-achieving input distribution, with the result stated in the following theorem.

Theorem 1. *For the optimization problem in (6), there exists a unique input distribution μ^* corresponding to an input $\mathbf{X}^*$ such that $C = I(\mathbf{X}^*; \mathbf{Y})$.*

Proof. The details of the proof are provided in Appendix A. The strategy is to apply the extreme value theorem [20] by establishing compactness of Λ in the topology of weak convergence and the continuity of $I(\mathbf{X}; \mathbf{Y})$ on Λ. □

A similar argument can be readily applied to establish the following result concerning the capacity problem for C_Δ in (9).

Theorem 2. *For the optimization problem in (9), there exists a unique input distribution μ^* corresponding to an input $\mathbf{X}^*$ such that $C_\Delta = I(\mathbf{X}^*; \mathbf{Y})$.*

Having established the existence and uniqueness for the optimization problems in (6) and (9), we now turn to a sensitivity analysis of C_Δ.

IV. Capacity Sensitivity

The capacity of a general memoryless additive noise channel can be viewed as a map from the noise distribution $F_{\mathbf{N}}$, and the constraint set Λ to $\mathbb{R}_{\geq 0}$. That is, $(F_{\mathbf{N}}, \Lambda) \mapsto C$, where C is the optimal value function of the optimization problem in (6) or (9).

In order to study approximations of one channel by another, it is natural to introduce the capacity sensitivity [13], [21], [22]. In particular, the capacity sensitivity is the capacity gap between two channels, and is defined formally as follows.

Definition 1. *Let $\mathcal{K} = (F_{\mathbf{N}}, \Lambda)$ and $\hat{\mathcal{K}} = (\hat{F}_{\mathbf{N}}, \hat{\Lambda})$ 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*

$$C_{\mathcal{K} \to \hat{\mathcal{K}}} \triangleq |C(\mathcal{K}) - C(\hat{\mathcal{K}})|. \tag{12}$$

The capacity sensitivity problem can be viewed as a special case of analyzing the sensitivity of nonlinear optimization problems, where we identify the capacity as the *optimal 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 additive 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 (6) under general perturbations from one channel to another.

Here, we are concerned with a class of *constraint perturbations*. In particular, the study of discrete input approximations of the capacity involves analyzing the effect of varying the constraint set Λ_Δ. The capacity sensitivity in this case therefore corresponds to

$$C_{\Lambda \to \Lambda_\Delta} = |C(\Lambda) - C(\Lambda_\Delta)|, \tag{13}$$

for $\Delta > 0$.

The first studies of general constraint perturbations for the capacity optimization problem were performed in [13], [21]. However, the analysis was limited to the case of scalar noise channels. We now establish an analogous result in the context of the vector channel defined in (4). Our main result establishes covergence of the capacity sensitivity.

Theorem 3. *Let C_Δ be as defined in (9) for the channel in (4). Then, $C_{\Lambda \to \Lambda_\Delta} \to 0$ as $\Delta \to 0$.*

Proof. See Appendix B. □

The result in Theorem 3 demonstrates that an arbitrarily good approximation of the capacity C in (6) can be obtained from C_Δ in (9). We remark that this property follows from a non-trivial proof relying on ideas from set-valued analysis and weak convergence. Moreover, it forms a basis for obtaining estimates of the capacity sensitivity $C_{\Lambda \to \Lambda_\Delta}$ using the ideas in [13].

V. Conclusion

Interference modeling for the IoT introduces the need to account for non-Gaussian vector channels. Symmetric α-stable random vector models form a realistic and tractable approach to capturing the resulting impulsive noise. In this paper, we have studied the fundamental aspects of additive vector symmetric α-stable noise channels. In particular, we have established the existence and uniqueness of the optimal input distribution.

Moreover, we have introduced the notion of capacity sensitivity for communication channels arising in the IoT. This forms a first step to understanding how model imperfections and practical constraints can affect estimates of data-rates achievable in this setting. We have established that inputs restricted to spherical shells can provide arbitrarily good approximations of the capacity of the channel under general moment constraints.

This work provides a basis for establishing further estimates of the capacity sensitivity as well as bounds on the capacity. An important research direction is to obtain practical insights from these bounds in order to understand whether designs based on tractable channel models are robust to noise or constraints arising from more realistic yet otherwise intractable models.

ACKNOWLEDGEMENTS

This work has been (partly) funded by the French National Agency for Research (ANR) under grant ANR-16-CE25-0001 - ARBURST.

APPENDIX A
PROOF OF THEOREM 1

The proof proceeds by establishing conditions under which the extreme value theorem [20] holds. The first step is to establish compactness of the constraint set, which is achieved by the application of Prokhorov's theorem [23]. The second step is to establish weak continuity of $I(\mathbf{X};\mathbf{Y})$ on Λ. The combination of these two steps ensures existence via the extreme value theorem. The final step is to establish uniqueness of the input.

A. Compactness of Λ

For any $\epsilon > 0$, there exists $\mathbf{a}_\epsilon = [a_{1,\epsilon}, \ldots, a_{d,\epsilon}]^T \succ 0$ such that for all $\mu \in \Lambda$

$$\Pr\left(|X_1|^r > a_{1,\epsilon}, \ldots, |X_d|^r > a_{d,\epsilon}\right) < \epsilon. \tag{14}$$

The inequality in (14) holds as a consequence of the generalized Markov inequality in [24, Example 2.3]. In more detail,

$$\begin{aligned}\Pr\left(|X_1|^r > a_{1,\epsilon}, \ldots, |X_d|^r > a_{d,\epsilon}\right) &\leq \min_{i=1,\ldots,d} \frac{\mathbb{E}[|X_i|^r]}{a_{i,\epsilon}} \\ &\leq \min_{i=1,\ldots,d} \frac{c_i}{a_{i,\epsilon}} < \epsilon.\end{aligned} \tag{15}$$

Now, choose $\mathcal{K}_\epsilon = [-a_{1,\epsilon}, a_{1,\epsilon}] \times \cdots \times [-a_{d,\epsilon}, a_{d,\epsilon}]$. Then, $\mathcal{K}_\epsilon$ is compact and $\mu(\mathcal{K}_\epsilon) \geq 1 - \epsilon$ for all $\mu \in \Lambda$. Hence, Λ is tight.

To establish closure, we apply a variation of the Portmanteau theorem [23]. Let $\{\mu_n\}_{n=1}^{\infty}$ be a weakly convergent sequence in Λ with limit μ_0. By a consequence of the Portmanteau theorem, it follows that

$$\begin{aligned}&[\mathbb{E}_{\mu_0}[|X_1|^r], \ldots, \mathbb{E}[|X_d|^r]]^T \\ &= \left[\int |x_1|^r d\mu_0(\mathbf{x}), \ldots, \int |x_d|^r d\mu_0(\mathbf{x})\right]^T \\ &\preceq \left[\liminf_{n\to\infty} \int |x_1|^r d\mu_n(\mathbf{x}), \ldots, \liminf_{n\to\infty} \int |x_d|^r d\mu_n(\mathbf{x})\right]^T \\ &\preceq \mathbf{c}.\end{aligned} \tag{16}$$

Hence, $\mu_0 \in \Lambda$. Since the choice of sequence is arbitrary, it follows that Λ is closed.

B. Continuity of $I(\mathbf{X};\mathbf{Y})$ on Λ

The second step is to establish that $I(\mathbf{X};\mathbf{Y})$ is weakly continuous on Λ. In particular, we need to show that for any weakly convergent sequence of probability measures $(\mu_n)_{n=1}^{\infty}$ with limit μ_0

$$\lim_{n\to\infty} -\int p_{\mathbf{Y}_n}(\mathbf{y}) \log p_{\mathbf{Y}_n}(\mathbf{y}) d\mathbf{y} = -\int p_{\mathbf{Y}_0}(\mathbf{y}) \log p_{\mathbf{Y}_0}(\mathbf{y}) d\mathbf{y}, \tag{17}$$

where $\mathbf{Y}_n$ is the output corresponding to an input $\mathbf{X}_n$ with probability measure μ_n. Note that $\mathbf{Y}_n = \mathbf{X}_n + \mathbf{N}$ admits a probability density function since $\mathbf{N}$ is truly d-dimensional.

Observe that if the limit and the integral in (17) can be swapped, the result follows from the definition of weak convergence if the probability density function of $\mathbf{N}$, $p_{\mathbf{N}}$, is bounded and continuous. Note that this is indeed the case since it can be shown that the characteristic function $\Phi_{\mathbf{N}}$ is integrable when $\mathbf{N}$ is truly d-dimensional.

Hence to complete the proof, we must justify swapping of the limit and integral in (17). This is achieved as follows. Let $1 < r < \alpha$. We need to establish that for all $n \geq 0$ and any $\delta > 0$, there exists $R(\delta) > 0$ such that

$$\left| \int_{\|\mathbf{y}\|_r > R(\delta)} p_{\mathbf{Y}_n}(\mathbf{y}) \log p_{\mathbf{Y}_n}(\mathbf{y}) d\mathbf{y} \right| < \delta. \tag{18}$$

To proceed, let

$$q(\mathbf{y}) = \frac{1}{\pi^d} \frac{1}{\prod_{i=1}^d (1 + y_i^2)}, \; \mathbf{y} \in \mathbb{R}^d \tag{19}$$

and observe that

$$\begin{aligned}&-\int_{\|\mathbf{y}\|_r > R(\delta)} p_{\mathbf{Y}_n}(\mathbf{y}) \log p_{\mathbf{Y}_n}(\mathbf{y}) d\mathbf{y} \\ &= -\int_{\|\mathbf{y}\|_r > R(\delta)} p_{\mathbf{Y}_n}(\mathbf{y}) \log q(\mathbf{y}) d\mathbf{y} \\ &\quad -\int_{\|\mathbf{y}\|_r > R(\delta)} p_{\mathbf{Y}_n}(\mathbf{y}) \log \frac{p_{\mathbf{Y}_n}(\mathbf{y})}{q(\mathbf{y})} d\mathbf{y} \\ &\leq \int_{\|\mathbf{y}\|_r > R(\delta)} p_{\mathbf{Y}_n}(\mathbf{y}) \left(\log \pi^d + \log \prod_{i=1}^d (1 + y_i^2) \right) d\mathbf{y} \\ &\quad + \frac{1}{e} \int_{\|\mathbf{y}\|_r > R(\delta)} q(\mathbf{y}) d\mathbf{y}.\end{aligned} \tag{20}$$

Note that

$$\int_{\|\mathbf{y}\|_r > R(\delta)} \log \pi^d p_{\mathbf{Y}_n}(\mathbf{y}) d\mathbf{y} \leq d \log \pi \frac{\mathbb{E}[\|\mathbf{Y}\|_r]}{R(\delta)} \leq d \log \pi \frac{L}{R(\delta)}, \tag{21}$$

which tends to zero as $R(\delta) \to \infty$. Here, $L < \infty$ since

$$\mathbb{E}[\|\mathbf{Y}\|_r] \leq \mathbb{E}[\|\mathbf{X}\|_r] + \mathbb{E}[\|\mathbf{N}\|_r] < L < \infty, \tag{22}$$

using the constraints in (6).

Moreover,

$$\begin{aligned} &\int_{\|\mathbf{y}\|_r > R(\delta)} p_{\mathbf{Y}_n}(\mathbf{y}) \log \prod_{i=1}^{d} (1 + y_i^2) d\mathbf{y} \\ &\leq 2 \int_{\|\mathbf{y}\|_r > R(\delta)} p_{\mathbf{Y}_n}(\mathbf{y}) \log \prod_{i=1}^{d} (1 + |y_i|) d\mathbf{y} \\ &\leq 2 \sup_{\|\mathbf{y}\|_r > R(\delta)} \left\{ \frac{\log \prod_{i=1}^{d} (1 + |y_i|)}{\|\mathbf{y}\|_r} \right\} \\ &\quad \times \int_{\|\mathbf{y}\|_r > R(\delta)} \|\mathbf{y}\|_r p_{\mathbf{Y}_n}(\mathbf{y}) d\mathbf{y}, \end{aligned} \tag{23}$$

which via the weighted arithmetic-geometric mean inequality is finite and tends to zero as $R(\delta) \to \infty$. Finally,

$$\frac{1}{e} \int_{\|\mathbf{y}\|_r > R(\delta)} q(\mathbf{y}) d\mathbf{y} \leq \frac{1}{eR(\delta)} \mathbb{E}_q[\|\mathbf{Y}\|_r], \tag{24}$$

which tends to zero as $R(\delta) \to \infty$.

After an application of the dominated convergence theorem (details omitted due to space constraints) for any $\delta > 0$

$$\begin{aligned} &\lim_{n \to \infty} -\int_{\|\mathbf{y}\|_r \leq R(\delta)} p_{\mathbf{Y}_n}(\mathbf{y}) \log p_{\mathbf{Y}_n}(\mathbf{y}) d\mathbf{y} \\ &= -\int_{\|\mathbf{y}\|_r \leq R(\delta)} p_{\mathbf{Y}_0}(\mathbf{y}) \log p_{\mathbf{Y}_0}(\mathbf{y}) d\mathbf{y}. \end{aligned} \tag{25}$$

Since the identities in (21), (23), (24) and (25) hold for all $\delta > 0$, weak continuity of $I(\mathbf{X}; \mathbf{Y})$ follows by taking $\delta \to 0$ (and hence $R(\delta) \to \infty$). The existence part of Theorem 1 then holds by applying the extreme value theorem.

C. Uniqueness

The uniqueness of the optimal input follows from the fact that the entropy $h(\mathbf{Y})$ is a strictly concave function of $P_{\mathbf{Y}}$. By the fact that the characteristic function of $\mathbf{N}$ is strictly positive, $P_{\mathbf{Y}}$ is a one-to-one function of $P_{\mathbf{X}}$. Hence, $h(\mathbf{Y})$ is a strictly concave function of $P_{\mathbf{X}}$. As the mutual information can be written as

$$I(\mathbf{X}; \mathbf{Y}) = h(\mathbf{Y}) - h(\mathbf{N}) \tag{26}$$

it follows that $I(\mathbf{X}; \mathbf{Y})$ is a strictly concave function of $P_{\mathbf{X}}$ since $h(\mathbf{N})$ does not depend on $P_{\mathbf{X}}$. Since this holds for any input lying in Λ, it follows then that the optimal input distribution is unique.

Appendix B
Proof of Theorem 3

A. Preliminaries

Before presenting the proof of Theorem 3, we present preliminaries that will be required. In particular, the proof relies on definitions and results in the theory of set-valued maps.

Let (Θ, d) and (S, d_S) be metric spaces. A set-valued map $\Gamma : \Theta \rightrightarrows S$ is a map from Θ to a subset in S such that for each point $\theta \in \Theta$ the set $\Gamma(\theta)$ is compact. Let $s \in S$, $\mathcal{S} \subseteq S$ and define $d_S(s, \mathcal{S}) = \inf_{\hat{s} \in \mathcal{S}} d_S(\hat{s}, s)$. Furthermore, for any $\epsilon > 0$ define the ϵ-ball centered at $s \in S$ by $B_\epsilon(s) = \{\hat{s} \in S : d_S(s, \hat{s}) < \epsilon\}$.

Definition 2. *Let $\theta \in \Theta$ and $\epsilon > 0$. The ϵ-neighborhood of the set $\Gamma(\theta)$ is defined by*

$$\eta_\epsilon(\Gamma(\theta)) = \{s \in S : d_S(s, \Gamma(\theta)) < \epsilon\} = \bigcup_{\overline{s} \in \Gamma(\theta)} B_\epsilon(\overline{s}). \tag{27}$$

There are two notions of continuity for set-valued maps [25], which are defined as follows.

Definition 3. *A set-valued map $\Gamma : \Theta \rightrightarrows S$ is upper hemicontinuous at $\theta \in \Theta$ if for all $\epsilon > 0$ there exists $\delta > 0$ such that $d(\overline{\theta}, \theta) < \delta$ implies that $\Gamma(\overline{\theta}) \subseteq \eta_\epsilon(\Gamma(\theta))$.*

Definition 4. *A set-valued map $\Gamma : \Theta \to S$ is lower hemicontinuous at $\theta \in \Theta$ if for all $\epsilon > 0$ there exists $\delta > 0$ such that $d(\overline{\theta}, \theta) < \delta$ implies that $\Gamma(\theta) \subseteq \eta_\epsilon(\Gamma(\overline{\theta}))$.*

If a set-valued map $\Gamma : \Theta \rightrightarrows S$ is both upper and lower hemicontinuous at θ, then it is said to be continuous. Intuitively, upper hemicontinuity can be viewed as constraining the size of expansions of the set $\Gamma(\theta)$, in the presence of small changes to θ. Conversely, lower hemicontinuity can be viewed as constraining the size of contractions of $\Gamma(\theta)$.

Although set-valued maps are widely studied in the case the set S is $\mathbb{R}^n$, the definitions also apply to other metric spaces and can even be extended to more general topological spaces. For the purposes of this paper, the set S corresponds to the set of probability measures with the Lèvy-Prokhorov metric [23].

In order to establish convergence of the spherical shell approximation, we will require the following theorem due to Berge [25] that provides conditions ensuring continuity of the optimal value function in terms of the upper and lower hemicontinuity of the constraint map.

Theorem 4. *Let Θ and S be two metric spaces, $\Gamma : \Theta \rightrightarrows S$ a compact-valued correspondence, and $\psi : S \times \Theta \to \mathbb{R}$ be a continuous function on $S \times \Theta$. Define*

$$\begin{aligned} \sigma(\theta) &= \arg\max\{\psi(s, \theta) : s \in \Gamma(\theta)\}, \ \forall \theta \in \Theta \\ \psi^*(\theta) &= \max\{\psi(s, \theta) : s \in \Gamma(\theta)\}, \ \forall \theta \in \Theta \end{aligned} \tag{28}$$

and assume that Γ is continuous at $\theta \in \Theta$. Then,

(i) *$\sigma : \Theta \rightrightarrows S$ is compact-valued and upper hemicontinuous at θ.*

(ii) *$\psi^* : \Theta \to \mathbb{R}$ is continuous at θ.*

Intuitively, Theorem 4 shows that if the constraint set varies continuously and the objective function is also continuous at a point, then the optimal value function is also continuous at a point.

B. Proof

We will require the following theorem characterizing dense subsets in the topology of weak convergence [26].

Theorem 5. *Let X be a separable metric space and $E \subset X$ be a dense subset of X. Then, the set of all probability measures whose support are finite subsets of E is dense on X in the space of probability measures equipped with the topology of weak convergence.*

Using Theorem 5, $\mathcal{P}_0$ with $\Delta = 0$ is dense in the topology of weak convergence. In particular, let $X = \mathbb{R}^d$ in Theorem 5. Then, the set of all probability measures whose supports are finite subsets of $\mathbb{R}^d$ is dense in the space of probability measures on $\mathbb{R}^d$. We therefore need to show that the set

$$\bigcup_{\Delta > 0} \bigcup_{r \in \Delta\mathbb{Z}} S_r \tag{29}$$

is dense in $\mathbb{R}^d$. Indeed, this holds since for any element α in $\mathbb{R}$ there exists a convergent sequence in

$$\bigcup_{\Delta > 0} \Delta\mathbb{Z} \tag{30}$$

such that the limit is α. Moreover, (29) clearly contains its own finite subsets. The claim that $\mathcal{P}_0$ is dense in $\mathcal{P}$, the set of all probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, then follows from Theorem 5.

By exploiting the fact that $\mathcal{P}(\mathbb{R}^d)$ forms a metric space with the Lèvy-Prokhorov metric ρ [23], from Berge's theorem (Theorem 4) and the weak continuity of the mutual information (established for Theorem 1 in Appendix A-B), it follows that if Λ_Δ is continuous as a set-valued map at $\Delta = 0$, then $C(\Lambda_\Delta) \to C(\Lambda_0)$.

To establish that Λ_Δ is continuous as a set-valued map, note that Λ_Δ is upper hemicontinuous (see Definition 3) since Λ_Δ is increasing as Δ decreases. Moreover, Λ is compact by Appendix A-A. This means that Λ is also separable.

Let μ_0 be an element of Λ. By Theorem 5, μ_0 can be obtained as the limit of a sequence $(\mu_n)_n$ of probability measures with support on finite subsets of $\mathbb{R}^d$. As a consequence, for all $\epsilon > 0$, there exists $\delta > 0$ such that $\Delta < \delta$ implies $\rho(\mu_0, \Lambda_\Delta) < \epsilon$. This implies that Λ_Δ is lower hemicontinuous (see Definition 4) at $\Delta = 0$ and hence Λ_Δ is continuous at $\Delta = 0$.

Since $C(\Lambda_\Delta) \to C(\Lambda_0)$ as $\Delta \to 0$, all that remains is to establish that $C(\Lambda) = C(\Lambda_0)$. Indeed, this holds since $\mathcal{P}_0$ is dense in the topology of weak convergence and the mutual information is weakly continuous.

REFERENCES

[1] M. Lauridsen, B. Vejlgaard, I. Kovács, H. Nguyen, and P. Mogensen, "Interference measurements in the European 868 MHz ISM band with focus on LoRa and SigFox," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2017.

[2] A. Moro, "Potential interference generated by UHF RFID systems on military telecommunication devices," *Italian National Agency for New Technologies*, vol. 3, 2007.

[3] M. Simek, M. Fuchs, L. Mraz, P. Moravek, and M. Botta, "Measurement of LowPAN network coexistence with home microwave appliances in laboratory and home environments," *Broadband and Wireless Computing, Communication and Applications*, vol. 10, pp. 292–299, 2011.

[4] B. Vejlgaard, M. Lauridsen, H. Nguyen, I. Kovács, P. Mogensen, and M. Sørensen, "Interference impact on coverage and capacity for low power wide area IoT networks," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2017.

[5] J. Ilow and D. Hatzinakos, "Analytic alpha-stable noise modeling in a poisson field of interfers or scatterers," *IEEE Transactions on Signal Processing*, vol. 46, no. 6, pp. 1601–1611, 1998.

[6] P. Pinto and M. Win, "Communication in a Poisson field of interferers-part I: interference distribution and error probability," *IEEE Transactions on Wireless Communications*, vol. 9, no. 7, pp. 2176–2186, 2010.

[7] K. Gulati, B. Evans, J. Andrews, and K. Tinsley, "Statistics of co-channel interference in a field of Poisson-Poisson clustered interferers," *IEEE Transactions on Signal Processing*, vol. 58, no. 12, pp. 6207–6222, 2010.

[8] M. Egan, L. Clavier, C. Zheng, M. de Freitas, and J.-M. Gorce, "Dynamic interference for uplink SCMA in large-scale wireless networks without coordination," *EURASIP Journal on Wireless Communications and Networking*, vol. 1, 2018.

[9] X. Yan, L. Clavier, G. Peters, N. Azzaoui, F. Septier, and I. Nevat, "Skew-t copula for dependence modelling of impulsive (alpha-stable) interference," in *IEEE International Conference on Communications (ICC)*, 2015.

[10] C. Zheng, M. Egan, L. Clavier, G. Peters, and J.-M. Gorce, "Copula-based interference models for IoT wireless networks," in *Proc. IEEE International Conference on Communications (ICC)*, 2019.

[11] J. Kitchen and W. Moran, "Copula techniques in wireless communications," *ANZIAM Journal*, vol. 51, pp. 526–540, 2010.

[12] M. Egan, S. Perlaza, and V. Kungurtsev, "Capacity sensitivity in additive non-gaussian noise channels," in *Proc. IEEE International Symposium on Information Theory*, 2017.

[13] M. Egan and S. Perlaza, "Capacity approximation of continuous channels by discrete inputs," in *Proc. Annual Conference on Information Sciences and Systems (CISS)*, 2018.

[14] G. Samorodnitsky and M. Taqqu, *Stable Non-Gaussian Random Processes*. New York, NY: CRC Press, 1994.

[15] T. Byczkowski, J. Nolan, and B. Rajput, "Approximation of multidimensional stable densities," *Journal of Multivariate Analysis*, vol. 46, pp. 13–31, 1993.

[16] T. Han, *Information-Spectrum Methods in Information Theory*. Berlin Heidelberg: Springer-Verlag, 2003.

[17] M. de Freitas, M. Egan, L. Clavier, G. Peters, and N. Azzaoui, "Capacity bounds for additive symmetric alpha-stable noise channels," *IEEE Transactions on Information Theory*, vol. 63, pp. 5115–5123, Aug. 2017.

[18] A. Dytso, M. Goldenbaum, H. Poor, and S. Shamai (Shitz), "When are discrete channel inputs optimal?optimization techniques and some new results," in *Annual Conference on Information Sciences and Systems (CISS)*, 2018.

[19] J. Fahs and I. Abou-Faycal, "On the finiteness of the capacity of continuous channels," *IEEE Transactions on Communications*, vol. 64, no. 1, pp. 166–173, 2016.

[20] D. Luenberger, *Optimization By Vector Space Methods*. New York, NY: Wiley, 1969.

[21] M. Egan, L. Clavier, M. de Freitas, L. Dorville, J.-M. Gorce, and A. Savard, "Wireless communciation in dynamic interference," in *IEEE Global Communciations Conference (GLOBECOM)*, 2017.

[22] M. Pinsker, V. Prelov, and S. Verdú, "Sensitivity of channel capacity," *IEEE Transactions on Information Theory*, vol. 41, pp. 1877–1888, Nov. 1995.

[23] P. Billingsley, *Convergence of Probability Measures*. John Wiley and Sons, 1999.

[24] A. Marshall, "Markov's inequality for random variables taking values in a linear topological space," *Inequalities in Statistics and Probability*, vol. 5, pp. 104–108, 1984.

[25] C. Berge, *Topological Spaces*. Mineola, NY: Dover, 1963.

[26] K. Parthasarathy, *Probability Measures on Metric Spaces*. New York, NY: Academic Press, 1967.