AMERICAN UNIVERSITY OF BEIRUT

AN INFORMATION THEORETIC TREATISE ON UNIVARIATE ALPHA-STABLE DISTRIBUTIONS

by

JIHAD JAWAD FAHS

A dissertation
submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
to the Department of Electrical and Computer Engineering
of the Faculty of Engineering and Architecture
at the American University of Beirut

Beirut, Lebanon
April 2016

AMERICAN UNIVERSITY OF BEIRUT

AN INFORMATION THEORETIC TREATISE ON UNIVARIATE ALPHA-STABLE DISTRIBUTIONS

by

JIHAD JAWAD FAHS

Approved by:

Dr. Karim Kabalan, Professor — Chair of Committee
Department of Electrical and Computer Engineering

Dr. Ibrahim Abou-Faycal, Associate Professor — Advisor
Department of Electrical and Computer Engineering

Dr. Zaher Dawy, Professor — Member of Committee
Department of Electrical and Computer Engineering

Dr. Farouk Abi Khuzam, Professor — Member of Committee

Department of Mathematics

Dr. Bassam Shayya, Professor — Member of Committee

Department of Mathematics

Kai Kal'Dalan on behalf of Prof. Brian

Dr. Brian L. Evans, Professor — Member of Committee

University of Texas at Austin

Kai Kal'Dalan on behalf of Prof. Aslan

Dr. Aslan Tchamkerten, Associate Professor — Member of Committee

University of Telecom ParisTech

Date of dissertation defense: April 12, 2016

AMERICAN UNIVERSITY OF BEIRUT

THESIS, DISSERTATION, PROJECT RELEASE FORM

Student Name: ____

Last First Middle

○ Master's Thesis ○ Master's Project ○ Doctoral Dissertation

☐ I authorize the American University of Beirut to: (a) reproduce hard or electronic copies of my thesis, dissertation, or project; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes.

☐ I authorize the American University of Beirut, **three years after the date of submitting my thesis, dissertation, or project,** to: (a) reproduce hard or electronic copies of it; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes.

Signature Date

Acknowledgements

First, I thank God for His countless blessings.

A special thanks to my family. Words cannot express how grateful I am to my mother and my father for all of the sacrifices that you have made on my behalf. My thanks are extended to my brother, my mother-in law, my late father-in-law, Leina and Sandra for their encouragement and support. At the end I would like to express my greatest appreciation to my beloved wife –an exemplary judge and a dedicated professor– Samar who carried the burden during the last years and was always my support in moments of doubt and fatigue.

I would like to express my special appreciation and thanks to my advisor and friend Professor Ibrahim Abou-Faycal. Over the last seven years, he has been a tremendous mentor for me. I would like to thank him for encouraging my research and for allowing me to grow as a research scientist. His advice on both research as well as on my career have been priceless and without his efforts this dissertation would not have been possible.

I would like to offer my thanks to my committee members. I am particularly grateful for the assistance given by professor Karim Kabalan since the first days of my enrolment as a Masters student at the American University of Beirut. His support helped overcoming many obstacles. I would also like to thank Professor Zaher Dawy for his encouragement, his support and the valuable information that i have gained during our many encounters. I wish to acknowledge the help provided by Professors Bassam Shayya and Farouk Abi-Khuzam: their approach in making advanced mathematical tools rather simple and enjoyable is deeply appreciated. Finally, i would like to thank Professor Brian L. Evans and Professor Aslan Tchamkerten for accepting to serve on my dissertation committee and for their constructive feedback that helped enhancing the quality of this work.

My sincere respects goes to the General Directorate of General Security of Lebanon in the person of the General Director Major General Abbas Ibrahim for his encouragement and appreciation of higher education and innovation. My respects are extended to General Monah Sawaya, Lieutenant Colonel Jamal Kashmar and late Major Hassan Zeineddine for their confidence and understanding.

My deepest gratitude is expressed to the American University of Beirut and the Lebanese National Council for Scientific Research for their financial support that made this work possible.

At last I would like to thank my friends in need whose presence eased the heat of this journey: Professor Ahmad El Hajj, Professor Lise Safatly, Professor Dima Fares, Professor Rawad Abi Assi, Professor Oussama Ibrahim, Sara, Nadine, Karim, Zareh and all the others.

An Abstract of the Dissertation of

JIHAD FAHS for Doctor of Philosophy
Major: Electrical and Computer Engineering

Title: An Information Theoretic Treatise on Univariate Alpha-Stable Distributions

Many communication channels are reasonably modeled to be impaired by additive noise. A Central Limit Theorem (CLT) argument is widely adopted to model the noise as a Gaussian variable. A deeper investigation shows that the CLT motivation leads to noise models that are in general stable and not necessarily Gaussian. This is validated by recent studies suggesting that many channels are affected by additive noise that is impulsive in nature and is best explained by the heavy tailed non-totally skewed alpha-stable distributions.

Considering impulsive noise environments comes with an added complexity with respect to the standard Gaussian environment: the alpha-stable probability density functions do not possess closed-form expressions except in some special cases. Furthermore, they have an infinite second moment and the "nice" Hilbert space structure defined by the space of random variables having a finite second moment –which represents the universe in which the Gaussian theory is applica-

ble, is lost along with its tools and methodologies.

We study these probability models, their detrimental effect as noise variables and we investigate various bounds on the performance limits in classical problems arising from noisy observations of some quantity of interest. Our approach is from an information theory point of view and some related disciplines:

i) We study the channel capacity of channels affected by non-totally skewed alpha-stable noise models and other types of impulsive noise channels. We characterize capacity achieving inputs and argue that a suitable cost function to be imposed on the channel input is one that grows logarithmically.

ii) We define novel and appropriate notions of power in such contexts. These notions boil down in the Gaussian context to the second moment which is the standard notion of power in the space of finite second moment variables.

iii) In estimation theory, classical tools to quantify the estimator performance are tightly related to the assumption of a finite variance noise. In alpha-stable environments, expressions such as the mean square error and the Cramer-Rao bound –which relates the error power to the Fisher information– are hence non sensible. We develop novel tools that are tailored to the alpha-stable and heavy tailed noise scenarios and coincide with the standard tools adopted in the Gaussian setup: a generalized Fisher information, a generalized Cramer-Rao bound, etc...

iv) We generalize known information inequalities commonly used in the Gaussian context: the de-Bruijn's identity, the data processing inequality, the Fisher information inequality and the isoperimetric inequality for entropies.

We develop the theory and the tools in the most possible general frameworks that often are to various degrees strong enough to infer results on other types of distributions. Our theoretical findings are paralleled with numerical evaluations of some related quantities using developed *Matlab* packages.

Contents

List of Figures

List of Theorems

Abbreviations

AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
CDF	Cumulative Distribution Function
CDMA	Code Division Multiple Access
CLT	Central Limit Theorem
DCT	Dominated Convergence Theorem
EPI	Entropy Power Inequality
FII	Fisher Information Inequality
FLOM	Fractional Lower Order Moments
FLOS	Fractional Lower Order Statistics
FOM	Fractional Order Moments
GCLT	Generalized Central Limit Theorem
GFII	Generalized Fisher Information Inequality
GIIE	Generalized Isoperimetric Inequality for Entropies
GP	Geometric Power
HOM	Higher Order Moments
IID	Independent and Identically Distributed
IIE	Isoperimetric Inequality for Entropies
KKT	Karush-Kuhn-Tucker
LHS	Left-Hand Side
MAI	Multiple Access Interference

MCT	Monotone Convergence Theorem
ML	Maximum Likelihood
MMSE	Minimum Mean Square Error
MSE	Mean Square Error
MVT	Mean Value Theorem
PDF	Probability Density Function
PPP	Poisson Point Process
QoS	Quality of Service
RHS	Right-Hand Side
RV	Random Variable
SNR	Signal-to-Noise Ratio
SαS	Symmetric Alpha-Stable
WLOG	Without Loss Of Generality
ZOS	Zero Order Statistics

Chapter 1

Introduction and Thesis Outcomes

In modeling the noise effects in a communication channel, it is common to assume the presence of an additive term that is often modeled as Gaussian distributed for two main reasons:

- Among all the probability distributions with a given finite variance, the Gaussian distribution is the "worst" from an entropy perspective. Therefore, a communication system design for a Gaussian-modeled noise may be thought of as a "worst-case" design and any finite-power additive noise encountered in real implementations would yield potentially better results.
- If the noise is believed to be of finite power and due to multiple independent sources, by the results of various Central Limit Theorems (CLTs), their cumulative effect asymptotically approaches a Gaussian distribution. The CLT justification neglected the effects of the normalizing constants and those of the underlying assumption of finite variance.

However, many noise models in the literature were found to be better explained by non-Gaussian statistics [1]. In [2, 3], Middleton proposed his class A, B and C models for representing the electromagnetic interference and the nature versus the man-made noise. Though these models were also found to be suitable for modeling different types of interference such as the atmospheric noise, and recently for co-channel interference [4], their usage is limited due to the rather huge complexity they presented. Gaussian mixtures [5, 6] were also adopted as a more tractable models for non-Gaussian noise and as a simplification for some of Middleton's classes. In addition, generalized Gaussian, were proposed as a general setup of the well known Laplacian and Gaussian distributions [7]. They are exponential distributions were the exponent is to the power $a > 0$ ($a = 1$ for Laplacian and 2 for Gaussian). The statistical key intention behind formulating the above different type of noise models was the urge for finding good models for the impulsive nature of noise in communication channels [8–10] whereby impulsive is meant that extreme values of the noise signal are observed very frequently (i.e., with notable amount of probability) which cannot be captured by the rather fast exponential decay of the tail of the normal distribution. However, the above models namely the mixtures and the generalized Gaussians fail to capture this impulsiveness for several reasons the most important of which is that they do not possess the polynomial behavior of heavy tailed noise distributions encountered in typical communication channels [11]. One family of such distributions is the generalized Cauchy [7] which has an algebraic or polynomial tail behavior and is found to be reasonable in modeling the amplitude of atmospheric impulse noise [12, 13]. However, though these distributions are suitable candidates for modeling noise in impulsive scenarios, they lack some supporting theoretical reasoning such as the CLT which validated the usage of Gaussian noise since the early days of communication theory.

In this regard, Gaussian noise presents a short and simple story of the large and wide theory of stable distributions. By stable it is meant that, if certain constraints are satisfied, they are closed under convolution. These distributions, which are a subset of the set of infinitely divisible distributions, are the only ones that have the captivating property of being the resultant of a limit of normalized sums of Independent and Identically Distributed (IID) Random Variables (RV), a result which is referred to as the Generalized Central Limit Theorem (GCLT), and a property that constitutes one of the main reasons behind the adoption of Gaussian statistics for noise models in communication channels. Though the Gaussian is considered to be one of the stable laws, it represents the exception: it is unique in the sense that it is the only one that has a finite variance and an exponential tail, where the former result was given by G. Pólya in what is considered one of the first and most interesting results concerning stable distributions. All the others have an infinite variance and a polynomial tail. More elaborate results and properties concerning these distributions are stated in Section 2.2.1 and a more complete literature on the theory of stable distributions can be found in [14–18]. It was P. Lévy [19] in 1925 who first characterized this class while studying the limit of sums of IID RVs., a work carried on later by A.Ya. Khintchine [20]. However, the statistical interest in these models did not grow till the appearance of the work of B. Mandelbrot [21] who showed that the empirical asset returns have heavier tail than the Gaussian and are more suitably modeled by alpha-stable[1] densities. Since then stable distributions have acquired significant attention in physics, astrology and cosmology, economics, biology, genetics, chemistry, geology, computer science and engineering.

1.1 Alpha- Stable Distributions and Communications Theory

When it comes to communications, interference has been often found to have an impulsive nature and therefore found to be best explained by alpha-stable

[1]Throughout this dissertation, we will use the term alpha-stable to refer to the class of non-degenerate stable distributions excluding the Gaussian. Otherwise, only the term stable will be used.

statistics. This has been the case for telephone noise [22] and audio noise signals [23]. Furthermore, in many works that treated the multiuser interference in radio communication networks, a theoretical derivation, based on the assumption that the interferers are distributed over the entire plane and behave statistically as a Point Poisson Process (PPP), yielded an interference with alpha-stable statistics, starting with Sousa [24] who computed the self interference, considering only the pathloss effect for three spread spectrum schemes, direct sequence with binary phase shift keying (DS/BPSK), frequency hopping with M-ary frequency shift keying (FH/MFSK), and frequency hopping with on-off keying (FH/OOK), where a sinusoidal tone is transmitted as the on symbol. In [25], the authors introduced a novel approach to stable noise modeling based on the LePage series representation which permits the extension of the results on multiple access communications to environments with lognormal shadowing and Rayleigh fading. Continuous time multiuser interference was also found [26] to be well represented as an impulsive alpha-stable random process. Recently in [27], alpha-stable distributions were found to model well the aggregate interference in wireless networks: the authors treated the problem in a general framework that accounts for all the essential physical parameters that affect network interference with applications in cognitive radio, wireless packets, covert military schemes and networks where narrowband and ultrawideband systems coexist. In [4], Gulati et al. showed that the statistical-physical modeling of co-channel interference in a field of Poisson and Poisson-Poisson clustered interferers obeys an alpha-stable or Middleton class A statistics depending whether the interferers are spread in the entire plane, in a finite area or in a finite area with a guard zone with the alpha-stable being suitable for wireless sensor, ad-hoc and femtocells networks when both in-cell and out-of-cell interference are included. A generalization of the previous results for radio frequency interference in multiple antennas is found in [28] where joint statistical-physical interference from uncoordinated interfering sources is derived without any assumption on spatial independence or spatial isotropic interference.

Lastly, the alpha-stable model arises as a suitable noise model in molecular communications [29].

Impulsive noise environments have been treated in communication channels within the context of robust signal processing, detection and estimation theory. Robust statistics theory has been long established [30], which along with several other works [7, 31] showed the enhanced performance of non-linear detectors over the linear ones in non-Gaussian noise scenarios. For algebraic noise models, and more specifically for the alpha-stable class, a general theory of stable signal processing based on Fractional Lower Order Moments (FLOM) was presented in [11]; The "stable theory" was in accordance with the fact that second order methods and linear estimation theory were no longer suitable for infinite variance noise channels and new criteria based on the dispersion of alpha-stable random variables (the dispersion is defined explicitly in Section 2.2.1) and FLOM were investigated. The stable theory was also used in the treatment of various detection and estimation problems [32, 33], and the performance of optimum receivers designed to operate in environments of impulsive noise modeled as an alpha-stable random process were investigated in [34]. More recently [35–38], alpha-stable statistics were used as models of additive noise in multiple access interference networks and the performance of new receivers, mitigation and diversity techniques were investigated. The performance measure was the evaluation of the Bit Error Rate (BER) in terms of the ratio between the signal power in the usual sense and the noise dispersion. This ratio is known as the generalized SNR.

Gonzales et al. [39] presented a new approach for dealing with heavy tailed noise environments. After presenting the shortcomings of the FLOM approach, they presented a "general" unit of strength measure based on logarithmic moments where they motivated its usage within the framework of estimation and filtering under impulsive noise. The Zero Order Statistics theory (ZOS), which is a limiting case of the Fractional Lower Order Statistics (FLOS) theory, depends on three new parameters, namely the geometric power, the zero-order location and the zero-order dispersion. These quantities play similar role to those played by the power, the expected value and the standard deviation, in the theory of second-order processes. The ZOS model, was used in communications for signal processing [40], for decoding purposes [41] and for developing a novel mode-type estimator with important optimality properties under very impulsive noise [42].

Finally, alpha-stable distributions were treated in image and radar processing problems [43,44]. While the majority of the cited alpha-stable models whether in communications or other fields are impaired with the fact that they are symmetric (SαS), non-symmetric distributions can arise in some special cases [45, 46].

1.2 Thesis Objectives

Despite the fact that alpha-stable distributions are good models for additive noise in communication channels due to theoretical, statistical and empirical justifications, the fundamental limits governing classical applications in estimation theory and communications theory are far from being known. In [17], the authors cited three reasons for which the Gaussian distribution has been adopted as the most common type of noise models though it represents only one member of the uncountable family of stable distributions. The first two are in accordance with what is listed below. The third was the "shortage of knowledge" regarding the alpha-stable distributions. In fact, these noise models pose multiple challenges to system designers and we intend to address some of them in this work:

- Despite the fact that the Probability Density Function (PDF) of a stable R.V. was proven to exist and exhibit rather "nice" properties, its expression is not known except in three special cases: the Gaussian, the Cauchy and the Lévy distributions.

- Such noise distributions have infinite variance, which implies that the received signal has potentially infinite power (second moment) and any analysis that is based on a Hilbert space approach is not valid anymore.

In what follows we present three main objectives for this dissertation:

1- Channel capacity: The capacity problem was treated extensively in the literature for AWGN channels and occasionally for non-Gaussian additive noise. When it comes to alpha-stable noise models, no information theoretic studies are available. In fact, the channel capacity under an average input cost constraint of a basic linear channel where the output is simply a noisy version of the input and where the noise behaves statistically as an alpha-stable RV is not known and optimal signaling schemes are not known either. Furthermore, finding upper bounds on the entropy of independent sums where one of the variables is an alpha-stable one is not yet known. A few attempts were made along these directions and as far as the authors know, only numerical evaluation of some achievable rates have been conducted [47, 48]. Numerical computations of channel capacity without any characterization of the nature of the input distribution will be faced with many complications:

- Numerically-accurate computational tools, where due to the fact that the input distribution we are searching for is in an infinite dimensional (see uncountable) space, discretization is necessary.
- The exactness of the optimal solution and its general properties. Since discretization of the PDF is necessary, one is indirectly imposing the non-necessarily existent constraint of peak-power limitation. The resulting optimal input may turn out to be very different "in nature" from the actual one without discretization.

From this perspective, a numerical computation of channel capacity of the classical AWGN channel using the Blahut-Arimoto algorithm results in a discrete input while the true optimal input is Gaussian shaped. Therefore, in light of the encountered difficulties in the numerical computations, one must attempt to at least characterize the optimal input in order to tune the numerical computations to the type of input the optimal one has.

2- Power notion: The evaluation of performance measures in multiple applications in estimation and communications theory is normally done as a function the channel state or quality. A key quantity that summarizes the quality of the channel is the SNR which is a ratio between the power of a signal containing relevant information to that of noise signal. A standard measure of the signal power is made through the evaluation of the second moment. When working in alpha-stable noise environments, some information bearing signals will necessarily have an infinite second moment which eventually leads to having zero SNRs, a fact that masks the possibility to quantify the channel's state. Hence, when encountering impulsive noise models, new power and SNR definitions must be investigated for information theoretic suitability of these models.

3- Parameter estimation: A standard way of measuring the performance of an estimator is through the evaluation of the mean square error and comparing it to its lower bound given by the Cramer-Rao bound. Clearly, such a measure is not sensible unless the noise has a finite variance. No similar result exists in the case of infinite variance noise models.

1.3 Thesis Outcomes

In light of what is presented, we study in this dissertation information theoretic limits of problems where an additive independent alpha-stable noise is affecting a quantity of interest. Whenever feasible our approach is made as generic as possible. We present in what follows novel tools, approaches and solutions to three classical problems:

1- Channel capacity: whereby we consider generic input-output functions, generic cost functions and generic noise PDFs each satisfying rather technical conditions. We derive a simple relation between these three parameters, one which dictates the type of the capacity achieving input. In Layman terms we prove that the support of the optimal input is bounded whenever the cost grows faster than a"cut-off" rate equal to the logarithm of the noise PDF evaluated at the input-output function. Furthermore, we prove a converse statement that says whenever the cost grows slower than the "cut-off" rate, the optimal input has necessarily an unbounded support. In addition, we show how the discreteness of the optimal input is guaranteed whenever the triplet satisfy some analyticity properties. We show that channels affected with an alpha-stable noise where both tails decays polynomially fall under the general model for which our results are derived. The results do also apply to the case where the noise is a mixture of alpha-stable variables. Specifically, the generic results boil down to saying that the "cut-off" growth rate of the cost function is logarithmic. A fact which is used in a second stage when we argue that a suitable cost function to be imposed on the channel input is one that grows similarly to the "cut-off" rate. The characterization of the type of the optimal input is made possible by insuring the existence of a finite and achievable capacity. This guarantee of finiteness is of high significance as it is typically the first step one would undertake in order to quantify the capacity of a channel at hand. In Appendix B, we tackle this problem and provide a sufficient condition for such a constrained optimization problem (the capacity problem) to be both *well-defined* and yielding a finite and *achievable* solution. A sufficient condition that is satisfied by the considered channels in this dissertation. We use the fact that the optimal input is of a discrete nature and tune a developed *Matlab* package to compute the alpha-stable channel capacity when the characteristic exponent of the noise is $1 \leq \alpha \leq 2$. This is done by following the same methodology adopted in [49] for non-linear Gaussian channels with the additional complexity in computing the alpha-stable density functions due to the fact they do not possess closed-form expressions. This is done whether by implementing the fast Fourier transform in a manner akin to [50] or by using the specialized "*Stable*" package provided by Prof. J. P. Nolan. Our developed numerical package computes the capacity through the evaluation of the optimal input probability mass function in the form of two vectors of optimal positions and their respective optimal probabilities. Furthermore, we use a variation of the alpha-stable channels where a Gaussian noise is present in addition to the alpha-stable one. This model is widely known as the Middleton class B [51, 52] and is common in MAI channels where a Gaussian model is used for the thermal noise and an alpha-stable one for the MAI [24, 27, 37]. The optimal input is once more characterized and numerical results are provided.

2- Next, we proceed and discuss new measures of the signal strength when alpha-stable distributions and more generally polynomially tailed density functions are present. Since these distributions possess infinite second moments, whenever a channel is affected by alpha-stable noise, the received signal will have an infinite average power (in the classical sense using second moments) independently of the input power. An infinite power received signal does not go down well with neither theoreticians nor engineers, and we therefore investigate different notions of power as proposed by Shao [11] and Gonzalez [42]. The applications of these measures are multiple fold:

   i- Within the context of finding the channel capacity of additive independent alpha-stable noise channels, we relate these new measures to the type of "realistic" constraints one should impose on the transmitted signal over such channels and we capitalize on the channel capacity results to advocate that these suitable cost functions should have a logarithmic growth at infinity. This conclusion is further supported by showing that an alpha-stable input achieves the capacity of an alpha-stable channel under an average cost constraint where the cost function has a logarithmic growth. This comes in accordance with the fact that in the Gaussian setting, a Gaussian input is capacity achieving when the input is subjected to an average power constraint. Though, theoretical interests aside, it may seem unusual in a Gaussian setup to impose logarithmic constraints or any other type of input constraints that permits $\mathsf{E}\left[X^2\right]$ to be infinite. However, when the channel model features noise distributions having an infinite second moment, imposing a second moment constraint masks the characterization of the behaviour of the transmission rates function of the quality of the channel since the channel Signal-to-Noise Ratio (SNR) will constantly evaluate to zero. The new power measure is then shown to comply with generic properties that are satisfied by the standard deviation and is numerically evaluated for different types of probability densities. Separately, we tackle the problem of characterizing a RV's strength by working in the Fourier domain. Specifically, we define a power operator in terms of the characteristic function and we present the advantages and the drawbacks of such a definition.

ii- When measuring the "quality" of an estimator; A point that will be discussed later.

3- To parallel some classical information theoretic results that prove to be tight and "special" when a Gaussian component is present, we develop similar inequalities in the alpha-stable case. It is well-known that the Entropy Power Inequality (EPI) presents a lower bound on the entropy of independent sums, however the existence of generic upper bounds is not always guaranteed [53, proposition 4]. Even when they exist, the problem is rather challenging and case specific. In this dissertation, we find an upper bound on the entropy of the sum of two independent RVs when one of them is symmetric stable (including the Gaussian variable). The bound is achieved through defining a series of information theoretic quantities and inequalities that generalizes equivalent counter parts known and used in the Gaussian setting. Namely, we generalize the notion of Fisher information and accordingly we state a generalized de Bruin's identity. A Generalized Fisher Information Inequality (GFII) is shown to hold for the generalized Fisher information and is made possible through a data processing inequality argument. The upper bound has several implications even in the Gaussian setting.

4- The newly defined quantities are relevant in studying information and estimation theoretic problems involving symmetric alpha-stable noise variables. For example, let us consider the problem of estimating a parameter θ by looking at a noisy observation $\theta + N$: this estimation problem is well understood whenever the noise has a finite second moment. A Cramer-Rao bound relates the error variance to the Fisher information through a lower bound and the estimator's performance is measured via the tightness of its error variance to its lower bound . When the noise is alpha-stable for example, the usage of the Cramer-Rao bound is not sensible. Whenever the variance of the noise is infinite, a new Cramer-Rao bound is established in the form of a lower bound which relates the new power measure –studied in 2- of the error and the generalized Fisher information. This bound presents a novel analytical tool to measure the performance of estimators in impulsive noise environments.

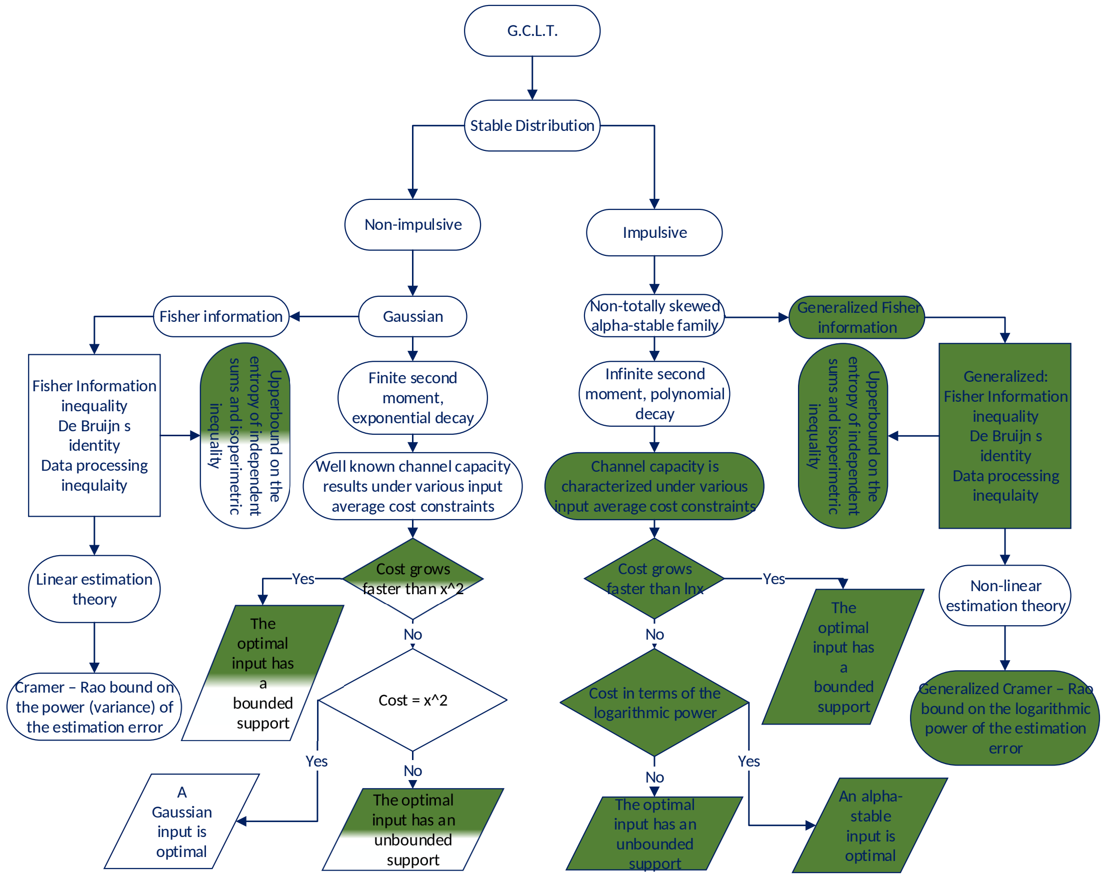

Figure 1-1: Flow chart of the comparison between the Gaussian theory and the new alpha-stable one. The boxes highlighted in green represent our contributions. The partially highlighted boxes represent partial contributions.

1.4 Applications

Based on the dissertation outcomes, we briefly present in what follows some of the applications of our results:

1- Our study gives sufficient conditions for the capacity achieving distribution of an additive noise channel to be discrete with finite number of mass points. Whenever these conditions hold true, our numerical package computes the optimal mass points, their respective probabilities and the capacity value at a given value of the average input cost. As presented, these sufficient conditions are mild and satisfied by the majority of the noise models and the input cost functions.

2- Novel quantities and identities are defined, stated and proven in this dissertation. Though some interesting applications of the generalized Fisher information and the new power definition are presented in this work –presented as separate items below, we believe that these quantities and identities will have "theoretical applications" of similar magnitude to those found for the Fisher information, the second moment and the multiple information theoretic identities involving these quantities in the fields of estimation and communications theory. As an example of such applications is the EPI which is found useful in finding bounds on capacity regions and in proving strong versions of CLTs.

3- Whenever the goal of a communications system is to maintain a Quality of Service (QoS) level for some or all of its users, that QoS for a user can be translated to a threshold rate (output entropy) or an output SNR. In both cases the QoS will be dependent on the output signal. Our "output"-based approach is tailored to this type of applications since it focuses on the output signal and takes into account the type of the encountered noise in the received signal to define sensible tools to quantify the QoS criteria.

4- Measuring the quality of an estimator is a fundamental tool to classify and rate the performance of estimators. The mean-square error measure does not fit well with impulsive noise models. In this context, our new power definition and the proven generalised Cramer-Rao bound make it possible to qualify and classify estimators and to search for optimality.

The rest of this dissertation is organized as follows: Chapter 2 is dedicated in a first part to general information theoretic notions and definitions, to the previously known channel capacity results and the adopted methodologies to solve the capacity problem. The last part is concerned with the stable distributions, their definition and some of their interesting properties. In Chapter 3, we give new capacity results for a multitude of input average cost constrained additive noise communication channels with applications on communications in impulsive noise environments, namely when the noise has an alpha-stable component. New signal strength measures are suggested in Chapter 4, where we argue the suitability of imposing logarithmically growing cost functions in the alpha-stable noise setting. We switch in Chapter 5 to deriving new information theoretic quantities and inequalities related to the alpha-stable model. Finally, Chapter 6 introduces a new estimation approach in infinite variance noise environments by providing a Cramer-Rao bound that relates the newly defined average power of the estimation error to the generalized Fisher information and Chapter 7 concludes the dissertation.

Chapter 2

Information Theoretic Background and Stable Distributions

2.1 Information Theory

2.1.1 Channel Capacity

The problem of determining the capacity of a communication channel is of high importance for system designers, if not essential. It characterizes the limit behavior, sets the bounds and indicates how to operate close to these theoretical bounds. The capacity problem is a part of the general theory of information whose foundation is due to Shannon [54, 55]. A key quantity in this theory was the notion of "entropy" as a measure of the uncertainty/information contained in a RV. As defined by Shannon, the entropy of a discrete RV X assuming discrete values $\{x_i\}_1^N$ with probability $\{p_i\}_1^N$ depends only on the RV's statistics and is defined as:

$$H(X) = -\sum_{i=1}^{N} p_i \log p_i,$$

and for a continuous RV with PDF, $p(\cdot)$:

$$h(X) = -\int p(x) \log p(x)\, dx$$

where the integration is over the support of $p(\cdot)$. Shannon also defined a quantity, called the "mutual information" between two RVs X and Y, that captures the amount of known information about Y when X is known and vice versa. In other words, it specifies by how much the knowledge of X reduces the uncertainty about Y. A key result of Shannon's work was setting the bounds for communicating reliably. This result which is known as the channel coding theorem states that capacity, i.e. the threshold that cannot be bypassed to communicate over a channel with arbitrarily low probability of error, is indeed the supremum of the mutual information between the output and the input of the channel over all the input probability measures or distribution functions (CDF). A complete interpretation and extensions of the work of Shannon can be found in [56].

Many communication channels in the literature are reasonably modeled to be impaired by additive noise N

$$Y = f(X) + N, \tag{2.1}$$

where X is the channel input whose alphabet is in $\mathcal{X}$, Y is the output with alphabet in $\mathcal{Y}$ and $f(\cdot)$ is a transformation of the input that could be deterministic or random. For all the above channel models, whenever the alphabets are discrete, the capacity problem is solved with either closed from expressions for the capacity or by using standard optimization tools and numerical packages. As for continuous channels, many complications are encountered with numerical computations and closed form expressions are sometimes impossible. When $f(\cdot)$ is deterministic, and whenever the additive noise has a distribution function $F_N(n)$ which is absolutely continuous (which means that the PDF $p_N(n) = \frac{dF_N(n)}{dn}$ of the noise exists), the output distribution is absolutely continuous [57] for all input distributions $F_X(x)$ and its derivative is $p_Y(y;F) = \int p_N\left(y - f(x)\right)\, dF_X(x)$. For these channels, the capacity is the supremum:

$$C = \sup_{F\in\Omega} I(F) \hat{=} \sup_{F\in\Omega} \iint p_N\left(y - f(x)\right) \log\left[\frac{p_N\left(y - f(x)\right)}{p(y;F)}\right] dy\, dF(x) \tag{2.2}$$

of the mutual information $I(F)$ between the input X and output Y over all input CDF F that meet the constraint Ω where $p(y;F)$ is the marginal output density induced by F. The above definition of the capacity is applicable to all deterministic, memoryless, additive and absolutely continuous noise channels. As it will be seen later, the alpha-stable channels treated in this work are all within the above class and hence the above definition of the channel capacity applies.

2.1.2 Related Work

One of the earliest results in communication theory is the derivation of the capacity of the Additive White Gaussian Noise channel (AWGN) under the average power constraint. This channel investigated by Shannon [54, 55], follows the model (2.1) where $f(x) = x$, $N \sim \mathcal{N}\left(0, \sigma_N^2\right)$[1], and by average power constraint, we mean that the input RVs are restricted to the set Ω:

$$\Omega = \left\{F : \int x^2\, dF(x) \leq A,\ A > 0\right\}.$$

For this channel, Shannon proved in his original paper that the capacity is achievable by using Gaussian statistics at the input i.e. $X \sim \mathcal{N}\left(0, A\right)$. Also the capacity has a closed form expression $C = \frac{1}{2}\log(1 + \text{SNR})$, which is probably one of the most used expressions to compute capacities in Gaussian deterministic channels. In the above expression, the SNR is defined as the ratio between the optimal input average power (signal power) and that of the noise by means of the cor-

[1] $\mathcal{N}\left(0, \sigma_N^2\right)$ means a Gaussian RV with zero mean and variance σ_N^2

responding second moments. The proof relied on the fact that, under second moment constraints, the Gaussian is an entropy maximizer. Smith [57] studied both, the peak power and the peak and average power constrained linear deterministic AWGN and proved that the optimal input is discrete. Shamai and Bar-David [58] extended the work of Smith to complex Gaussian channels and established the discreteness of the optimal input in their setup as well. Similarly, for fading channels, whether the fading is Rayleigh [59], Ricean [60] or whether the channel is considered to be non-coherent [61] the capacity problem were treated under the average power constraint and discrete input statistics were found to be again optimal. More recently, the capacity of Gaussian Channels with duty cycle and power constraints was found also to be achieved by a discrete input [62]. In all the above channel models, whether the channel was deterministic or not the input-output relation $f(\cdot)$ is assumed to be linear and the noise had Gaussian statistics. Even when the channel is modeled to have memory, the input signal is assumed to be distorted in a linear fashion and again the noise is assumed Gaussian [63] and [64]. Non-linear Gaussian channels were investigated by [65] where the authors proposed the usage of the Hermite polynomials as suitable basis for Hilbert space expansions for some capacity related information theoretic quantities. They treated the problem under a general setup of input constraints which included an even-moment, a compact support and a combination of both types of constraints. Again the optimal signaling schemes were found to be discrete in the huge majority of the treated cases with the exception of the average power constrained Gaussian linear channel and "equivalent" channels.

When it comes to non-Gaussian additive noise models, fewer attempts were made to characterize the channel capacity. It started with Smith [57] who extended his results from Gaussian noise to noise distributions satisfying some predefined "robustness" conditions. However these distributions were of the "Gaussian like" family. Later, Tchamkerten [66] considered a scalar additive channel whose input is amplitude constrained and derived sufficient conditions on the noise distributions that guarantee the discreteness of the optimal input. In addition, non-Gaussian additive noise channel capacity was treated under the average power constrained within a general setup on two occasions [67, 68] and for an exponential noise in [69]. In the first, Das imposed some technical conditions on the noise distribution and showed that the capacity-achieving distribution has bounded (respectively unbounded) support when the noise probability density function (PDF) decays at rate slower (respectively faster) than a Gaussian. In the second, and again after assuming that the noise statistics satisfies certain general conditions, yet different than those imposed by Das, the authors showed that the capacity-achieving distribution is discrete except when the noise is Gaussian. In light of the known results [57, 65, 66, 68], a linear channel, an average power constrained input and a Gaussian additive noise present an exceptional combination. Changes whether in terms of input conditions, channel linearity or in noise distribution will switch the optimal input from a continuous to a discrete one. Except of Tchamkerten's work [66], all the considered non-Gaussian noise models were assumed to have a finite variance, a property that alpha-stable distributions are not endowed with. When it comes to [66], the noise PDFs were assumed to abide by four general restrictions, one of which was the property of analytical extendability which is not satisfied by all stable distributions as it will be seen later, and even for the cases where the alpha-stable noise PDF is analytically extendable there is no clear proof that they satisfy another condition that states, in layman terms, that the amplitude noise PDF must be bounded by two non-increasing functions over a horizontal strip around the x-axis. Though alpha-stable distributions are known to have a polynomial tail behavior on the real axis, such a behavior is not clear when extended to the complex plane. In fact we prove in Appendix D a novel upper bound on the complex extension of the alpha-stable noise PDF when alpha is not less than 1. Finally, the input in [66] was amplitude constrained, i.e. the input RV was restricted to be in the set Ω:

$$\Omega = \left\{ F : \int_{-A}^{A} dF(x) = 1,\ A > 0 \right\}.$$

This condition is translated to the notion of imposing a peak power on the input signal, however it does not capture the notion of power for general signals where a finite support constraint is not existent. Studying the capacity problem when the input is not amplitude restricted will be seriously different. The characterization of the optimal input has been considered on numerous occasions in the literature [54, 57, 59–66, 68, 69] using the following methodology:

1. First, apply the theory of convex optimization from which necessary and sufficient conditions for the optimal distribution can be derived. These conditions are referred to as the Karush-Kuhn-Tucker (KKT) conditions.

2. Next, the optimal input is assumed to have an accumulation point. An extension to the complex domain is made which enables the usage of complex analysis tools.

3. Finally, either a solution is found or an elimination process on the type of the optimal input begins in order to characterize it.

In this dissertation, we follow a similar approach with possibly adopting slightly different alternatives in some cases to account for the different characteristics and some non-similar properties, such as the analyticity, which is not shared by all the alpha-stable distributions.

2.2 Stable Distributions

This section is dedicated to the general characteristics of stable distributions. While complete monographs were written in this regard, we will cite their most important properties, namely the ones related to or used in this dissertation. The theorems and properties presented hereafter will be listed without any proof and are selected from [11, 14–17, 20, 70–74].

2.2.1 General Properties

Definition 1 (Stability)**.** A distribution function $F(x)$ is said to be stable if to every $b_1 > 0$, $b_2 > 0$, and real c_1, and c_2, there corresponds a positive number b and a real number c such that the relation:

$$F\left(\frac{x-c_1}{b_1}\right) * F\left(\frac{x-c_2}{b_2}\right) = F\left(\frac{x-c}{b}\right)$$

holds.

The above definition is the basic property of stable distributions from which they got their name. It can be expressed differently as such: Let X_1 and X_2 be two independent copies of X, X is said to be stable if and only if, for any positive constants a and b we have

$$aX_1 + bX_2 = cX + d,$$

for some positive c and some $d \in \mathbb{R}$ and where by equality we mean in distribution. If $d = 0$ the distribution is said to be strictly stable.

Definition 2 (Standard Stable Characteristic Function)**.** The characteristic function [2] of a "standard" stable distribution $F(\cdot)$ is of the form:

$$\phi(\omega) = \exp\left[-\left(1 - i\beta\,\mathrm{sgn}(\omega)\Phi\right)|\omega|^{\alpha}\right]$$

[2]The characteristic function $\phi(\omega)$ of a distribution function $F(x)$ is defined by:

$$\phi(\omega) = \int_{\mathbb{R}} e^{i\omega x}\, dF(x)$$

where $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, $\text{sgn}(\omega)$ is the sign of ω and

$$\Phi = \begin{cases} \tan\left(\frac{\pi\alpha}{2}\right) & \alpha \neq 1 \\ -\frac{2}{\pi}\log|\omega| & \alpha = 1 \end{cases}$$

Definition 3 (Stable Characteristic Function)**.** A stable RV N is defined as:

$$N = \begin{cases} \gamma X + \delta & \alpha \neq 1 \\ \gamma X + \left(\delta + \frac{2}{\pi}\beta\gamma\log\gamma\right) & \alpha = 1 \end{cases}$$

where $\gamma > 0$, $\delta \in \mathbb{R}$ and X is a "standard" stable RV. Hence, the characteristic function of a stable distribution is:

$$\phi(\omega) = \exp\left[i\delta\omega - \gamma^{\alpha}\left(1 - i\beta\,\text{sgn}(\omega)\Phi\right)|\omega|^{\alpha}\right] \tag{2.3}$$

All RVs N having a characteristic function as in (2.3) are called stable, denoted $N \sim S(\alpha, \beta, \gamma, \delta)$, and their corresponding distribution functions $F_N(n)$ are named stable distributions. The parameters are subject to the restrictions:

$$0 < \alpha \leq 2 \qquad -1 \leq \beta \leq 1 \qquad \gamma > 0 \qquad \delta \in \mathbb{R}$$

In fact multiple parametrizations of stable distributions were used in the literature to qualify stable distributions, and different parameters names were adopted. These multiple parametrizations led to some confusion in the treatment of stable distributions and erroneous expressions of the characteristic function (the sign of the imaginary term for $\alpha \neq 1$) were observed in some references (regarding this remark see [17]). Throughout the dissertation, we adopt the above notation for information theoretic suitability. In addition, we adapt these erroneous statements to correct ones and to a format that corresponds to the adopted parametrization.

All stable laws are absolutely continuous and the the PDF of N is denoted by $p_N(\cdot)$. By means of the inversion formula:

$$p_N\left(n;\alpha,\beta,\gamma,\delta\right)=\frac{1}{2\pi}\int_{\mathbb{R}}e^{-in\omega}\phi(\omega)d\omega \tag{2.4}$$

where $\phi(\omega)$ is given by (2.3). The four parameters figuring in the definition of a stable distribution play important roles in determining the form of $p_N(\cdot)$. Indeed, the characteristic exponent α affects the tail behavior of the PDF and thus the impulsiveness of the RV ($\alpha = 2$ is for Gaussian, and as α decreases towards 0 the tail becomes heavier). As for the skewness β, it represents the non-symmetry of the distribution ($\beta = 0$ for Symmetric Alpha-Stable (SαS) PDFs, $\beta > 0$ the right tail is heavier than the left and vice versa for $\beta < 0$). For these reasons α & β are called the shape parameters. The dispersion of N is equal to γ^α. Finally, δ is a location parameter. It is worth noting that the last two parameters cannot be always related to the usual notions of variance and mean of a RV (in fact among all stable distributions, only the Gaussian has a finite variance).

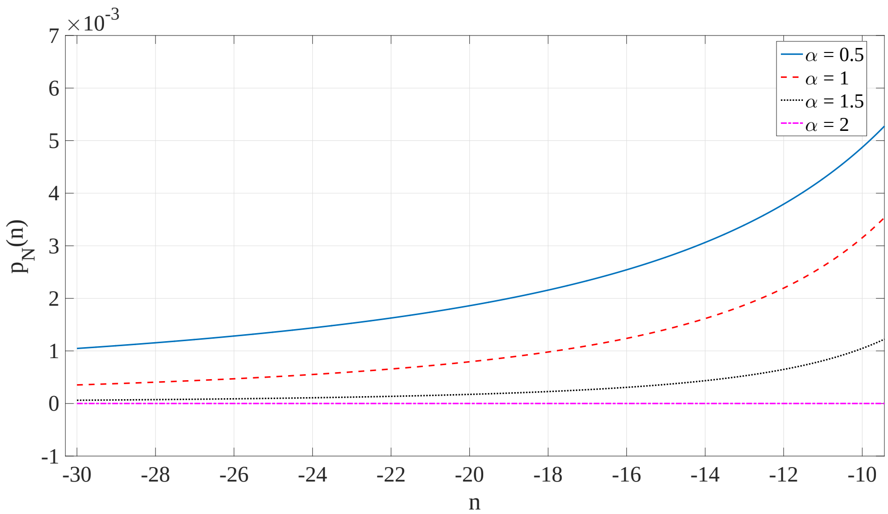

Figure 2-1: The left tail behaviour of the PDF $p_N(n)$ of $N \sim \mathcal{S}\left(\alpha,0,1,0\right)$ function of the characteristic exponent α.

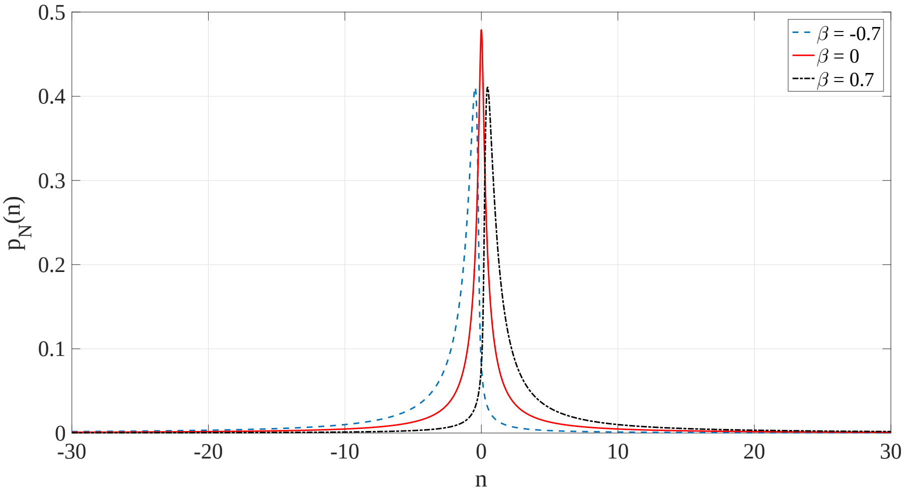

Figure 2-2: The PDF $p_N(n)$ of $N \sim \mathcal{S}(0.6, \beta, 1, 0)$ for various values of the skewness parameter β.

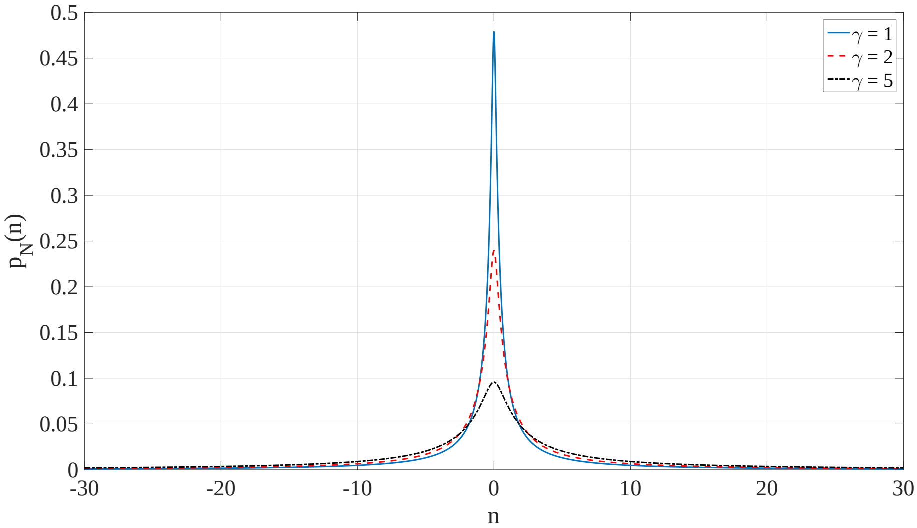

Figure 2-3: The PDF $p_N(n)$ of $N \sim \mathcal{S}(0.6, 0, \gamma, 0)$ for various values of the scale parameter γ.

Using (2.4) and (2.3) , the following relations can be derived:

$$p_N\left(n;\alpha,\beta,\gamma,\delta\right) = p_N\left(n-\delta;\alpha,\beta,\gamma,0\right) \tag{2.5}$$

$$p_N\left(n;\alpha,\beta,\gamma,0\right) = p_N\left(-n;\alpha,-\beta,\gamma,0\right) \tag{2.6}$$

and

$$p_N(n;\alpha,\beta,\gamma,0) = \begin{cases} \dfrac{1}{\gamma} p_N\left(\dfrac{n}{\gamma};\alpha,\beta,1,0\right) & \alpha \neq 1 \\ \dfrac{1}{\gamma} p_N\left(\dfrac{n-\frac{2}{\pi}\beta\gamma\ln\gamma}{\gamma};\alpha,\beta,1,0\right) & \alpha = 1 \end{cases} \tag{2.7}$$

Equation (2.6) is known as the reflection property. Using the above set of equations it is sufficient to only characterize the analytical properties of the stable PDF for the standard case $\gamma = 1$, $\delta = 0$ and map the results accordingly. Also due to the reflection property, we can restrict the analysis to the range $n > 0$ or to the range $0 \leq \beta \leq 1$. The standard stable variable is denoted $S(\alpha,\beta)$ and the corresponding PDF is $p_N(n;\alpha,\beta)$. As stated earlier, one of the reasons for the non-popularity of stable distributions is the inability to express them as closed form expressions. Indeed, equation (2.4) is expressible in only three special cases:

1- Gaussian

$$p_N(n;2,0) = \frac{1}{2\sqrt{\pi}} e^{-\frac{n^2}{4}}$$

2- Cauchy

$$p_N(n;1,0) = \frac{1}{\pi\left(1+n^2\right)}$$

3- Lévy

$$p_N(n;1/2,1) = \begin{cases} 0 & n < 0 \\ \dfrac{1}{\sqrt{2\pi}} n^{-\frac{3}{2}} e^{\frac{-1}{2n}} & n > 0 \end{cases}$$

Besides the above cases, no closed-form formulas are available for stable PDFs. However, except when $\alpha = 1, \beta \neq 0$, the series expansions of these PDFs is known:

$$p_N(n;\alpha,\beta) = \begin{cases} \dfrac{1}{\pi n} \displaystyle\sum_{k=1}^{+\infty} \dfrac{(-1)^{k-1}}{k!} \Gamma(\alpha k+1) \left(\dfrac{n}{p}\right)^{-\alpha k} \sin\left[\dfrac{k\pi}{2}(\alpha+\zeta)\right] & 0 < \alpha < 1 \\ \dfrac{1}{\pi n} \displaystyle\sum_{k=1}^{+\infty} \dfrac{(-1)^{k-1}}{k!} \Gamma\left(\dfrac{k}{\alpha}+1\right) \left(\dfrac{n}{p}\right)^{k} \sin\left[\dfrac{k\pi}{2\alpha}(\alpha+\zeta)\right] & 1 < \alpha \leq 2 \end{cases}$$

where $p = (1+\eta^2)^{\frac{-1}{2\alpha}}$, $\zeta = \frac{2}{\pi}\arctan(\eta)$ and $\eta = \beta\tan\left(\frac{\pi\alpha}{2}\right)$. Finally, concerning the representation of the alpha-stable PDFs, some of those were found to be expressible in terms of special functions such as the modified Bessel, hypergeometric, Whittaker, and Lommel functions. Also some were found to be expressible in terms of Fresnel integrals. For a complete reference on the subject, we refer the reader to [17, 72] and references within.

In what follows, we will state some important properties of the stable laws.

Lemma 1 (Support of Stable Density Functions). *The support of $p_N(n)$ where $N \sim S(\alpha,\beta,\gamma,\delta)$ is*

$$\mathcal{S}(n;\alpha,\beta,\gamma,\delta) = \begin{cases} [\delta,+\infty) & \alpha<1, \beta=1 \\ (-\infty,\delta] & \alpha<1, \beta=-1 \\ (-\infty,+\infty) & o.w. \end{cases}$$

Property 1 (Moments of Stable Variables). *Every stable law with characteristic exponent $0<\alpha<2$ has finite absolute moments $E[|X|^r]$ of order r, $0<r<\alpha$. On the other hand, all absolute moments of order greater or equal α are infinite.*

Hence, in particular, it follows that among all stable laws only the normal law has a finite variance (in fact all the moments of the normal law are finite). For $1<\alpha<2$, the stable laws have mathematical expectations, for $0<\alpha\leq 1$ they have neither variance nor mathematical expectation.

Property 2 (Unimodality, Continuity and Differentiability of Stable Density Functions). *All stable laws are unimodal, continuous and have derivatives of all orders at every point.*

Theorem 1 (Tail Behaviour of Stable Distributions). *Let $N \sim S(\alpha,\beta,\gamma,\delta)$ with*

$0 < \alpha < 2$, $-1 < \beta \leq 1$. Then as $n \to +\infty$,

$$\begin{aligned} 1 - F_N(n) &\sim 4\frac{k_N}{\alpha}\, n^{-\alpha} \\ p_N(n) &\sim 4k_N\, n^{-(\alpha+1)} \end{aligned}$$

where $4\,k_N = \alpha\gamma^\alpha(1+\beta)\sin\left(\frac{\pi\alpha}{2}\right)\frac{\Gamma(\alpha)}{\pi}$.[3] Using the reflection property (2.6), the lower tail properties are similar: for $-1 \leq \beta < 1$ as $n \to +\infty$

$$\begin{aligned} F_N(-n) &\sim 4\frac{k'_N}{\alpha}\, n^{-\alpha} \\ p_N(-n) &\sim 4k'_N\, n^{-(\alpha+1)} \end{aligned}$$

where $4\,k'_N = \alpha\gamma^\alpha(1-\beta)\sin\left(\frac{\pi\alpha}{2}\right)\frac{\Gamma(\alpha)}{\pi}$.

The above result shows that alpha-stable distributions have polynomial tails in all cases except for the left (resp. right) tail when $\beta = 1$ (resp. $\beta = -1$). For these cases $|\beta| = 1$ and when $\alpha < 1$ these tails do not exist (see Lemma 1). The latter behavior, and that of the non-polynomial tail of the totally skewed distribution ($|\beta| = 1$) when $\alpha \geq 1$ can be found in [74] (in fact it is faster than the polynomial one) where asymptotic expansions of the standard stable PDFs are expressed for all the different cases.

Theorem 2 (Analyticity, Order and Type of Stable Density Functions)**.** *Let $\lambda = \gamma^\alpha\sqrt{1+\beta^2\tan^2\left(\frac{\pi\alpha}{2}\right)} > 0$. The PDF of a stable distribution with a characteristic exponent $\alpha > 1$ is an entire function of order $\rho = \frac{\alpha}{\alpha-1}$ and type $\tau = \lambda^{-\frac{\rho}{\alpha}}(\alpha - 1)\alpha^{-\frac{\alpha}{\alpha-1}}$. When $\alpha = 1$, they are entire functions of infinite order if $\beta \neq 0$ but are rational if $\beta = 0$ with poles at the points $i\gamma$ and $-i\gamma$. Finally, for $\alpha < 1$, they*

[3] *Let $f(\cdot)$, $g(\cdot)$ be two real valued functions. We say that $f(x) \sim g(x)$ as $x \to a$ if* $\lim_{x\to a}\frac{f(x)}{g(x)} = 1$

have the form:

$$p_N(n;\alpha,\beta,\gamma,\delta) = \begin{cases} \dfrac{1}{\pi n}\Phi_1\left(n^{-\alpha}\right) & n > 0 \\ \dfrac{1}{\pi n}\Phi_2\left(|n|^{-\alpha}\right) & n < 0 \end{cases}$$

where $\Phi_1(z)$ *and* $\Phi_2(z)$ *are entire functions of order* $\rho = \frac{1}{1-\alpha}$ *and type* $\tau = \lambda^\rho (1-\alpha)^{\frac{\alpha}{(1-\alpha)}}$.

2.2.2 Central Limit Theorems and Entropy Maximization

Domains of Attraction of Stable Laws

Perhaps one of the most captivating property and the primary theoretical motivation for this work is the double sided relation between stable laws and CLTs. Formally, this can be stated in the following theorem [15, Th. p.162]:

Theorem 3 (Generalized Central Limit Theorem (GCLT))**.** *Let*

$$Z_n = \frac{X_1 + X_2 + ... + X_n}{B_n} - A_n \tag{2.8}$$

be a normalized sum of IID RV $X_1, X_2, ..., X_n$. *For the distribution function* $F(x)$ *to be a limit distribution for sums (2.8) of IID distributed summands, it is necessary and sufficient that it be stable.*

By stable we mean formally satisfying Definition 1, and the normalizing constants A_n and B_n are chosen to insure convergence. This result which appeared first in the work of Lévy [19] and Khintchine [20] is indeed the main characteristic of stable laws. In fact, and based on the theory of infinitely divisible laws, the general form of the characteristic function (2.3) is a direct consequence of this theorem. It is interesting to note that in the derivation, the restriction on the characteristic exponent α was obtained separately for the value 2 and by using opposite conditions on some intermediate functions than those used for the case

$0 < \alpha < 2$. This would explain, in a way the unique behavior of the Gaussian distribution in the time domain though it appears to be a limiting case in the Fourier domain. Statements such that Gaussian is a limiting case for alpha-stable distributions would be misleading since it is indeed an exception rather than the norm. Going back to equation (2.8), we define $G(x)$ to be the distribution function according to which all the X_is behave. Then if the sum converges to $F(x)$ for some A_n and B_n, we say that $G(x)$ is within the domain of attraction of $F(x)$. By the result of Theorem 3, it is clear that only stable distribution have non-empty domains of attraction. An important problem is to determine these domains of attraction for stable distributions. This is established in the literature ([15, 75] and references within) and we would like make a few comments in this regard:

- There are two different forms of the domains of attraction for stable distributions, one is for the Gaussian and the other is for alpha-stable. This fact comes in consistency with the exceptional nature of the Gaussian distribution among all stable laws.

- Necessary and sufficient conditions are provided in order for $G(x)$ to be in domain of attraction of stable laws [15]. Interestingly, it is shown that in the Gaussian case, $G(x)$ is not restricted to have a finite variance. However, for the alpha-stable case the domain of attraction is a subset of the infinite variance space.

- A narrower domain of attraction is defined when some restrictions are made on the normalizing constants. A special choice of B_n would be of the form $an^{\frac{1}{\alpha}}$ where a is some constant and $0 < \alpha \leq 2$. Under this choice of B_n, $G(x)$ is said to be in the domain of normal attraction of stable distribution. The Gaussian domain of normal attraction is the space of finite variance distributions. In this case, we have $B_n = a\sqrt{n}$. As for the alpha-stable ones, it consists of all laws that have a polynomial tail of order α i.e $\lim_{x\to\infty} |x|^\alpha Pr(|X| > |x|) = \text{cte} \neq 0$ where $B_n = an^{\frac{1}{\alpha}}$ and $0 < \alpha < 2$.

For completeness, we list in this section some formal statements about the domains of attraction of alpha-stable variables.

Theorem 4 (Domains of Attraction of Stable Laws)**.** *[75, p.76 Sec.6 Th.2.6.1] In order that a distribution $F(x)$ belong to the domain of attraction of a stable law $V(x)$ with characterisitc exponent α $(0 < \alpha < 2)$, it is necessary and sufficient that, as $|x| \to +\infty$,*

$$F(x) = \begin{cases} (c_1 + o(1))\, \dfrac{1}{(-x)^\alpha} h(-x) & x < 0 \\ 1 - (c_2 + o(1))\, \dfrac{1}{x^\alpha} h(x) & x > 0 \end{cases} \tag{2.9}$$

where the function $h(x)$ is slowly varying in the sense of Karamata [75, Appendix 1] and c_1 and c_2 are constants with $c_1,\ c_2 \geq 0$, $c_1 + c_2 > 0$. The parameters β and γ of the stable law $V(x)$ are determined as follows:

$$\beta = \frac{c_2 - c_1}{c_2 + c_1}$$

and

$$\gamma = \begin{cases} -\alpha L(\alpha)(c_1 + c_2)\cos\left(\dfrac{\pi}{2}\alpha\right) & 0 < \alpha < 1 \\ -\alpha M(\alpha)(c_1 + c_2)\cos\left(\dfrac{\pi}{2}\alpha\right) & 1 < \alpha < 2 \\ (c_1 + c_2)\dfrac{\pi}{2} & \alpha = 1 \end{cases}$$

where

$$L(\alpha) = \int_0^{+\infty} (e^{-y} - 1)\frac{dy}{y^{1+\alpha}} = -\frac{\Gamma(1-\alpha)}{\alpha} < 0$$

$$M(\alpha) = \int_0^{+\infty} (e^{-y} - 1 + y)\frac{dy}{y^{1+\alpha}} = \frac{\Gamma(2-\alpha)}{\alpha(\alpha-1)} > 0$$

Furthermore, it is proven [75, page 46] that for a RV X to be in the domain of attraction of a stable law with characteristic exponent α it is necessary that

$B_n = \sqrt[\alpha]{n}\, h(n)$, where $h(n)$ is a slowly varying function at 0. The special case when $B_n = a\sqrt[\alpha]{n}$, $a > 0$, defines a sub-domain called the domain of normal attraction $\mathbb{D}_{\alpha,\beta}$. The set $\mathbb{D}_{\alpha,\beta}$, $0 < \alpha < 2$ can be characterized as follows:

Theorem 5 (Domains of Normal Attraction of Stable Laws). *[15, p.181 Sec.35 Th.5] In order that the law $F(x)$ belong to the domain of normal attraction of the stable law $V(x)$ with characterisitc exponent α ($0 < \alpha < 2$) and given constants c_1 and c_2, with $B_n = a\sqrt[\alpha]{n}$, it is necessary and sufficient that*

$$F(x) = \begin{cases} (c_1 a^\alpha + \alpha_1(x))\,\dfrac{1}{|x|^\alpha} & x < 0 \\ 1 - (c_2 a^\alpha + \alpha_2(x))\,\dfrac{1}{x^\alpha} & x > 0 \end{cases} \tag{2.10}$$

where a is a positive constant and the functions $\alpha_1(x)$ and $\alpha_2(x)$ satisfy the conditions

$$\lim_{x\to-\infty} \alpha_1(x) = \lim_{x\to+\infty} \alpha_2(x) = 0$$

Though Theorems 4 and 5 relate the alpha-stable laws to the tail behaviour of their attracted laws via their characteristic exponents, one would be interested to capture if there are any relations when it comes to the Fourier domain, more specifically between characteristic functions. A task that was also fulfilled in [15, 75].

Theorem 6 (Domains of Attraction of Stable Laws: Characteristic Functions). *[75, Ch 2 p.85 Th. 2.6.5] In order that the distribution with characteristic function $\phi(\omega)$ belong to the domain of attraction of the stable law $S(\alpha,\beta,\gamma,0)$, it necessary and sufficient that, in the neighbourhood of the origin,*

$$\log\phi(\omega) = i\omega\delta - \gamma^\alpha|\omega|^\alpha \tilde{h}(|\omega|)\,[1 - i\beta\,\mathrm{sgn}(\omega)\Phi] \tag{2.11}$$

where

$$\Phi = \begin{cases} \tan\left(\dfrac{\pi\alpha}{2}\right) & \alpha \neq 1 \\ -\dfrac{2}{\pi}\log|\omega| & \alpha = 1 \end{cases}$$

and δ is a constant. $\tilde{h}(\omega) : \mathbb{R}^+ \to \mathbb{R}^+$ is a slowly varying function [75, Appendix 1] at 0.

When it comes to $\mathbb{D}_{\alpha,\beta}$ for which $h(n) = 1$ and $B_n = \sqrt[\alpha]{n}$, an immediate task would be to determine, under this condition, the expansion of $\phi(\omega)$ around 0. To this end, we state the following theorem

Theorem 7 (Domains of Normal Attraction of Stable Laws: Characteristic Functions)**.** *In order that the distribution with characteristic function $\phi(\omega)$ belong to the domain of normal attraction $\mathbb{D}_{\alpha,\beta}$ of the stable type $S(\alpha, \beta, \gamma, 0)$, it is necessary and sufficient that, in the neighbourhood of the origin,*

$$\log \phi(\omega) = i\omega\delta - \gamma^{\alpha}|\omega|^{\alpha} [1 - i\beta \operatorname{sgn}(\omega)\Phi] [1 + o(1)] \tag{2.12}$$

Proof. The proof is done exactly as in [75] to prove theorem 6. The only difference is that the tail behavior of the distribution function $F(x)$ that belongs to the domain of normal attraction $\mathbb{D}_{\alpha,\beta}$ is given by equation (2.10) instead of (2.9). □

Stable Distributions as Entropy Maximizers: A Central Limit Theorem Approach

In the rest of this section, we will restrict our treatment of CLTs in the context of domain of normal attraction exclusively. Accordingly, we consider the following basic CLT:

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i, \tag{2.13}$$

where the X_is are independent copies of a random variable X with zero mean and unit variance. By the CLT results, Z_n approaches the standard Gaussian as n tends to infinity, in a variety of probabilistic senses (one of them is in distribution). This would mean in layman terms, that a Gaussian distribution is the cumulative effect of infinite RVs (up to a certain normalizing constant) each with a finite variance. This would have some intuitive properties from an information theoretic perspective: Since entropy is considered as a measure of randomness, adding RVs would increase entropy. This fact, in addition to the CLT would come in accordance with the known fact that Gaussian distribution is an entropy maximizer among all RV having a given variance. The formal link between the CLT and entropy maximization started with the Entropy Power Inequality (EPI) stated by Shannon and proved rigorously by Stam [76] and later by Blachman [77]:

$$N(X+Y) \geq N(X) + N(Y), \tag{2.14}$$

where X and Y are two independent RVs each of finite entropy. $N(Z)$ is the entropy power of Z and is defined as such:

$$N(Z) = \frac{1}{2\pi e} e^{2\,h(Z)}, \tag{2.15}$$

where $h(Z)$ is the entropy of Z. Equality in (2.14) is achieved when both X and Y are Gaussian distributed. The EPI would imply that for the sequences of power of 2, $h\left(Z_{2^k}\right) \geq h\left(Z_{2^{k-1}}\right)$, $k \geq 1$. A stronger result [78] showed that $h(Z_n)$, $n \geq 1$ is an increasing sequence and hence proving the assertion stated by Lieb [79]. It shows that in the CLT the entropy increases at every step until it reaches its maximum, which is that of the Gaussian. This CLT approach in showing that Gaussian PDFs are entropy maximizers among the RV which are in their domain of normal attraction leads to the following question concerning the GCLT: Are alpha-stable distributions entropy maximizers among their domain of normal attraction i.e. RVs having a polynomial tail? More formally, if we consider this version of the GCLT [71, Th. 3.54]:

Suppose X is a random variable with tail probabilities that satisfy $x^{\alpha}\text{Pr}(X > x) \to \eta$ and $x^{\alpha}\text{Pr}(X < -x) \to \eta$ as $x \to +\infty$, with $\eta > 0$ and $1 < \alpha < 2$. Then

$\mu = E[X]$ must be finite and

$$\frac{1}{an^{\frac{1}{\alpha}}}\sum_{i=1}^{n} X_i - A_n \to Z \sim S(\alpha, 1) \tag{2.16}$$

where the convergence is in distribution. The X_is are IID according to X, $a = \left(\frac{\Gamma(\alpha)\sin\left(\frac{\pi\alpha}{2}\right)}{\pi\eta}\right)^{-\frac{1}{\alpha}}$, and $A_n = \frac{\mu}{a}n^{1-\frac{1}{\alpha}}$.

This theorem is considered to be the analog of that considered in equation (2.13), but whether the entropy of the sum (2.16) increases or not is not clear and needs more consideration. An immediate investigation shows that the EPI would not yield the desired result due to the presence of $\frac{1}{\alpha}$ instead of $\frac{1}{2}$ as power of n in the normalizing constants. The author believe that it would not be the case since the domain of normal attraction of stable variable does not involve any average constraint on its elements while the domain of normal attraction of the Gaussian type imposes a constraint of finiteness on the second moment.

Chapter 3

Generic Capacity Results: Applications to the Alpha-Stable Channel

3.1 Background

In communication systems design, a key engineering objective is to build systems that operate close to channel capacity. Needless to say that this quantity, as defined by Shannon [54, 55] in his pioneering work, is the cutoff value which delimits the achievable region for "reliable" communications. Clearly, the channel capacity and how it can be achieved are intimately related to the channel model. Despite the well-known capacity results for discrete memoryless channels, closed-form capacity expressions are rarely found in the literature for continuous ones. The most well-understood –and perhaps important– continuous channel is the linear Additive White Gaussian Channel (AWGN) subjected to an average power constraint. In the literature, multiple channel models were investigated by making variations to the Shannon setup in the following aspects:

- *The input-output relationship*: While Shannon considered a deterministic linear input-output relationship, many studies assumed a non-deterministic relationship [59–61] or generally a non-linear deterministic one [49].

- *The input constraint or cost function*: One of the main reasons of the popularity of the second moment constraint $\mathsf{E}\left[X^2\right]$ –which corresponds to a cost function $\mathcal{C}(x) = x^2$, is that it represents the average power of the discrete time transmitted signal which is equal to the average power of the corresponding white continuous process assuming that the transmitted signals are square integrable. Nevertheless, other input constraints were studied starting with Smith [57] who considered peak power constraints and a combination of peak and average power constraints. More recently, the capacity of Gaussian Channels with duty cycle and average power constraints was studied in [62] .

- *The noise distribution*: Though Gaussian statistics of the noise can be motivated by the Central Limit Theorem (CLT), it also has an appealing property of being the worst case noise from an entropy perspective among finite second moment Random Variables (RVs). Nevertheless, Non-Gaussian average power constrained communication channels have some applications and their channel capacities were investigated under a general setup in the work of Das [67] where the noise is assumed to have a finite second moment, a condition that was not imposed on the non-Gaussian noise distributions in [68].

- *Combinations of more than one aspect* were also considered in the literature. Smith [57] extended his capacity results for the peak power constrained Gaussian channel to non-Gaussian ones where the noise statistics are Gaussian like. Later, Tchamkerten [66] considered a scalar additive channel whose input is amplitude constrained and for which the additive noise is assumed to satisfy some general properties however not necessarily having a finite second moment. Lately, Fahs and Abou-Faycal [49, 80] investigated non-linear Gaussian channels under a general setup of input constraints such as even moments, compact support constraints and a combination of both types.

Nearly, for all the cited models above, and whenever the noise PDF is assumed to have an analytical extendability property, the optimal input is proven to be of a discrete nature and in most cases with a finite number of mass points. Additionally, channel capacity could not be written in closed-form. In this sense, the linear AWGN channel and some "equivalent" channels [49] seem to be an exception, along with a few channel models such as the additive exponential noise channel under a mean constraint with non-negative inputs [81]. For these channel models, the optimal input distribution is found to be of the same nature of the noise and capacity is described in closed-form.

One is tempted to study whether there is a general relation between the input-output function $f(\cdot)$, the input cost function $\mathcal{C}(\cdot)$, and the noise PDF $p_N(\cdot)$ that governs the type of the capacity-achieving input. In this work, we conduct this study and consider general, real, deterministic and memoryless discrete-time additive noise channels. By "general", we mean that the input-output relationship $f(\cdot)$ may not be linear but required nevertheless to satisfy some rather mild conditions. Additionally, instead of formulating the problem in terms of the average power constraint or the FOM constraints, we use generic input cost functions $\mathcal{C}(\cdot)$ that are also required to satisfy some technical conditions. We emphasize that our results cover all cost functions which are "super-logarithmic"[1] which is a rather very large set. When it comes to the noise statistics, the noise is as-

[1]A "super-logarithmic moment" is an expectation of the form $\mathsf{E}\left[f(|X|)\right]$ for some function $f(|x|) = \omega(\ln(|x|))$.
We say that $f(x) = \omega\left(g(x)\right)$ if and only if $\forall\, \kappa > 0, \exists\, c > 0$ such that $|f(x)| \geq \kappa |g(x)|, \forall |x| \geq c$. Equivalently, we say that $g(x) = o\left(f(x)\right)$. We say that $f(x) = \Omega\left(g(x)\right)$ if and only if $\exists\, \kappa > 0, c > 0$ such that $|f(x)| \geq \kappa |g(x)|, \forall |x| \geq c$. Equivalently, we say that $g(x) = O\left(f(x)\right)$. We say that $f(x) = \Theta\left(g(x)\right)$ if and only if $f(x) = O\left(g(x)\right)$ and $f(x) = \Omega\left(g(x)\right)$.

sumed to be absolutely continuous with positive and continuous PDFs that are with or without monotonic tails and have a finite logarithmic-type of moments. Two conditions are however imposed on the noise PDF and are subsequently presented. The first guarantees the finiteness of the noise differential entropy. The second concerns the tail behavior of a lower envelope to the noise PDF. These two conditions are "easily satisfied" such as whenever the PDF has a dominant exponential or a dominant polynomial component. Despite the apparent long list of requirements, we emphasize that the considered functions $f(\cdot)$, input costs $\mathcal{C}(\cdot)$ and noise PDFs cover the vast majority of the known models found in the literature.

Though our main interest in this dissertation is to study communications in impulsive noise environments specifically channels affected by alpha-stable additive noise, we conduct this generic study simply because the methodology remains unchanged with no additional complexity nor constraints. We showcase our generic results by applying them to channels impaired by a non totally skewed alpha-stable additive noise or a mixture of alpha-stable noise variables . Furthermore, we consider a different setup where the noise is a composite one. Namely, the noise is modeled as an independent sum of a Gaussian variable and a non-totally skewed alpha-stable. We note that the totally skewed alpha-stable noise models ($|\beta| = 1$) which are one sided ($\alpha < 1$) or posses a single polynomially decaying tail ($\alpha \geq 1$) are left out from this study since they are not frequent nor natural noise models and do not fit under the general two sided polynomially tailed distributions.

The results give new insights for communicating over alpha-stable noise channels by characterizing the type of optimal signaling schemes in order to maximize the input transmission rates when dealing with these impulsive noise scenarios. Our main results - stated in Theorems 8 and 9, imply in the alpha-stable setup that whenever $\mathcal{C}(x) = \omega\left(\ln|f(x)|\right)$, the support[2] of the capacity-achieving input is necessarily bounded. In addition, we state and prove a converse statement that says that whenever $\mathcal{C}(x) = o\left(\ln|f(x)|\right)$, the optimal input is necessarily unbounded.

Using a developed *Matlab* numerical package, we compute the alpha-stable channel capacity. The numerical package searches for the optimal mass function: optimal location of the mass points and their corresponding optimal probabilities. For validation, the resulting function is checked whether it satisfies the necessary and sufficient KKT conditions of optimality (equation (3.4)). Furthermore, the package is used to evaluate the capacity of the composite noise channel and some related quantities. We mention that the numerical package returns a discrete optimal input and hence surely finds such one whenever the channel capacity is achieved with discrete statistics. This is the case for the alpha-stable channel whenever $\alpha \geq 1$ and for the composite noise for all the ranges of $0 < \alpha < 2$. We note that the alpha-stable densities are computed using the "*Stable*" package provided by professor John P. Nolan.

3.2 A Generic Channel Model

We consider a generic memoryless real discrete-time noisy communication channel where the noise is additive and where the input and output are possibly non-linearly related as follows:

$$Y_i = f(X_i) + N_i, \tag{3.1}$$

where i is the time index. We denote by $Y_i \in \mathbb{R}$ the channel output at time i. The input at time i is denoted X_i and is assumed to have an alphabet $\mathcal{X} \subseteq \mathbb{R}$. The channel's input is distorted according to the deterministic and possibly non-

[2]We define the support of a RV as being the set of its points of increase i.e. $\{x \in \mathbb{R} : \Pr(x - \eta < X < x + \eta) > 0 \text{ for all } \eta > 0\}$.

linear function $f(x)$. Additionally, the communication channel is subjected to an additive noise process that is independent of the input. The variables $\{N_i\}_i$ are also assumed to be Independent and Identically Distributed (IID) RVs.

Finally, we subject the input to an average cost constraint of the form: $\mathsf{E}\left[\mathcal{C}\left(|X_i|\right)\right] \leq A$, for some $A \in \mathbb{R}^{+*}$ where $\mathcal{C}(\cdot)$ is some cost function:

$$\mathcal{C} : \mathbb{R}^{+} \longrightarrow \mathbb{R}.$$

Accordingly, we define for $A > 0$

$$\mathcal{P}_A = \left\{ \text{ Probability distributions } F \text{ of } X : \int \mathcal{C}\left(|x|\right)\, dF(x) \leq A \right\}, \tag{3.2}$$

the set of all distribution functions satisfying the average cost constraint.

Given that the channel model is stationary and memoryless, the capacity-achieving statistics of X_i are also memoryless (IID), therefore we suppress the time index and write

$$Y = f(X) + N, \tag{3.3}$$

where the noise is assumed to be absolutely continuous with PDF $p_N(\cdot)$. This implies that the channel transition probability density function is given by

$$p_{Y|X}(y|x) = p_N(y - f(x)), \quad y \in \mathbb{R},\ x \in \mathcal{X}. \tag{3.4}$$

We characterize the tail behavior of $p_N(\cdot)$ by considering the following positive functions which are non-increasing for $x \geq 0$ and non-decreasing for $x < 0$:

$$T_{\mathrm{l}}\left(x\right) = \begin{cases} \inf\limits_{0 \leq n \leq x} p_N(n) & x \geq 0 \\ \inf\limits_{x \leq n \leq 0} p_N(n) & x < 0, \end{cases} \qquad T_{\mathrm{u}}\left(x\right) = \begin{cases} \sup\limits_{n \geq x}\ p_N(n) & x \geq 0 \\ \sup\limits_{n \leq x}\ p_N(n) & x < 0. \end{cases}$$

Considering the tail behavior of $T_{\mathrm{l}}\left(x\right)$ and $T_{\mathrm{u}}\left(x\right)$ instead of $p_N(x)$ allows us to include in our analysis PDFs which do not possess a monotonic tail. For those that do, $p_N(x)$, $T_{\text{l}}(x)$ and $T_{\text{u}}(x)$ will be identical for large values of $|x|$.

The main results of this work are based on relating the tail behavior of $\mathcal{C}(\cdot)$ to that of $T_{\text{l}}(\cdot)$ and $T_{\text{u}}(\cdot)$ in order to characterize the capacity-achieving input distributions of channel (3.3). More explicitly we prove that, whenever $\mathcal{C}(|x|) = \omega\left(\ln\left[\frac{1}{T_{\text{l}}(f(x))}\right]\right)$, the optimal input has necessarily a bounded support. Furthermore, we prove a converse statement: whenever $\mathcal{C}(|x|) = o\left(\ln\left[\frac{1}{T_{\text{u}}(f(x))}\right]\right)$, the capacity-achieving input is not bounded.

When the noise PDF has a monotonic tail, our results infer that cost functions which are $\Theta\left(\ln\left[\frac{1}{p_N(f(x))}\right]\right)$ form somehow a "transition" between bounded and unbounded optimal inputs. For example, whenever the noise is Gaussian, the "transitional" cost is of the form $\Theta\left(f^2(x)\right)$. The discreteness –and hence the finiteness of the number of mass points of the optimal input in the bounded case– is a direct consequence of the analyticity properties of $p_N(\cdot)$ and $\mathcal{C}(\cdot)$ whenever these properties exist.

3.2.1 Assumptions

In this chapter, we make the following rather-mild assumptions:

- The function $f(\cdot)$:

  C1- The function is continuous.

  C2- The absolute value of the function $|f(\cdot)|$ is a non-decreasing function of $|x|$ and $|f(x)| \to +\infty$ as $|x| \to +\infty$.

- The cost function $\mathcal{C}(\cdot)$:

  C3- The cost function is lower semi-continuous and non-decreasing. Without Loss of Generality (WLOG) we assume that $\mathcal{C}(0) = 0$: if it were not, define $\mathcal{C}_0(|x|) = \mathcal{C}(|x|) - \mathcal{C}(0)$ and adjust the input space under the cost $\mathcal{C}_0(|x|)$ to $\mathcal{P}_{A-\mathcal{C}(0)}$. Note that necessarily $A - \mathcal{C}(0) \geq 0$.

C4- $\mathcal{C}(|x|) = \omega(\ln|f(x)|)$.

- The noise PDF $p_N(\cdot)$:

C5- The PDF is positive and continuous on $\mathbb{R}$. Note that this automatically implies that $p_N(\cdot)$ is upper bounded.

C6- There exits a non-decreasing function

$$\mathcal{C}_N : \mathbb{R}^+ \longrightarrow \mathbb{R}$$

such that $\mathcal{C}_N(|n|) = \omega(\ln|n|)$, and

$$\mathsf{E}_N\left[\mathcal{C}_N(|N|)\right] = L_N < \infty.$$

This necessarily implies that $\mathrm{E}_N\left[\ln(1+|N|)\right] < \infty$. Note that, for example, the above condition holds true for any noise PDF whose tail is faster than $\frac{1}{x(\ln x)^3}$.

Since from an information theoretic perspective, the general channel model (3.1) is invariant with respect to output scaling, we consider WLOG that the noise PDF is less than "1" for technical reasons. Furthermore, the boundedness of $p_N(\cdot)$ along with the fact that it has a finite logarithmic moment insure that its differential entropy exists and is finite $h(N) < \infty$ (see [82, Proposition 1]).

Restrictions C1 to C6 are "technical" in the sense that they represent sufficient conditions for the existence of a solution to the capacity problem as defined in [57] and enables the formulation of the Karush Kuhn Tucker (KKT) conditions as being necessary and sufficient for optimality of the input probability distribution.

- The lower and upper bounds $T_{\mathrm{l}}(\cdot)$ and $T_{\mathrm{u}}(\cdot)$:

Note that by definition, $0 < T_{\rm l}(x) \le p_N(x) \le T_{\rm u}(x) \le 1$ for all $x \in \mathbb{R}$. We assume that $T_{\rm l}(\cdot)$ and $T_{\rm u}(\cdot)$ satisfy the following properties:

C7- The function $L(x) = \ln\left[\frac{1}{T_{\rm l}(x)}\right]$ which is positive, non-decreasing for $x \ge 0$ and non-increasing in $x < 0$, satisfies the following inequality:

$$L(x+y) \le \kappa_{\rm l}\left(L(x) + L(y)\right), \tag{3.5}$$

for some positive constant $\kappa_{\rm l}$, whenever $|x|$, $|y|$ are sufficiently large.

We note that functions that satisfy condition C7 define a convex set. In fact, let $f(x)$, $g(x)$ be two positive, non-decreasing functions on $\mathbb{R}^+$ non-increasing on $\mathbb{R}^{-*}$. Let $\alpha \in [0,1]$ and define $h = \alpha f + (1-\alpha)g$. The function $h(x)$ is positive, having the same monotonic properties. Then, whenever there exists $\kappa_{\rm f}$ and $\kappa_{\rm g} > 0$ for which f and g satisfy condition C7, we have

$$h(x+y) = \alpha f(x+y) + (1-\alpha)g(x+y) \le \kappa_{\rm h}(h(x) + h(y)),$$

where $\kappa_{\rm h} = \max\{\kappa_{\rm f}; \kappa_{\rm g}\} > 0$.

We clarify that condition C7 is for example satisfied by all noise distribution functions where $T_{\rm l}(x)$ is any linear combinations of:

$$T_{\rm l}(x) = \Theta\left(s(x)e^{r(x)}\right) \qquad T_{\rm l}(x) = \Theta\left(\frac{s(x)}{r(x)}\right),$$

where

$$r(x) = |x|^a \underbrace{\log\ldots\log(|x|)}_{\beta \text{ times}}, \qquad s(x) = |x|^{a'} \underbrace{\log\ldots\log(|x|)}_{\beta' \text{ times}},$$

and where the parameters a, $a' \in \mathbb{R}^+$, and β, $\beta' \in \mathbb{N}$, chosen so that $T_{\rm l}(x)$ is positive, its total integral is no greater than one, and conserves its monotonic behavior[3]. The fact that these two general types satisfy condition C7 is based on the following basic identities [83]:

- For all x, y and $r \in \mathbb{R}$,

$$|x + y|^r \leq \max\{1; 2^{r-1}\} \left(|x|^r + |y|^r\right).$$

- For any $x_0 > 0$, there exist $y_0 > 0$ such that

$$|x| + |y| \leq |xy|^p, \text{ for some } p > 1 \text{ whenever } |x| > x_0, \, |y| > y_0.$$

Finally, we also assume that

C8- The integral $-\displaystyle\int_{-\infty}^{+\infty} T_{\mathrm{u}}\left(x\right) \ln T_{\mathrm{l}}\left(x\right) \, dx$ exists and is finite.

Note that whenever the tail of $p_N(\cdot)$ is monotone, condition C8 is not necessary and boils down to saying that noise differential entropy is finite which is a byproduct of properties C5 and C6 of the noise PDF.

When it comes to conditions C5 through C8 –and specifically C7 and C8–, they are satisfied by a rather large class of noise probability functions that includes most of the known probability models such as Gaussian, generalized Gaussian, generalized t, alpha-stable distributions and all of their possible mixtures.

3.3 Preliminaries

In this section we establish some preliminary results that are needed in subsequent sections: we derive lower and upper bounds on the output probability and a quantity of interest presented hereafter.

[3]The values $\beta = 0$ and $\beta^{'} = 0$ imply that respectively $r(x)$ and $s(x)$ have no logarithmic component.

We start by noting that for channel (3.3), the existence of a positive, continuous transition PDF such as in (3.4), implies the existence, for any input distribution F, of an induced output probability density function $p_Y(y) = p(y; F)$ which is also continuous (hence upper-bounded) [49] and is given by:

$$p_Y(y;F) = p(y;F) = \int p_N(y - f(x))\, dF_X(x) \leq 1. \tag{3.6}$$

Furthermore, equation (3.6) along with the fact that $f(\cdot)$ is continuous insures that the property that $p_N(\cdot)$ is bounded away from zero on compact subsets of $\mathbb{R}$ is conserved as well for $p_Y(y;F)$. This in turns implies that $p_Y(\cdot)$ is also positive on $\mathbb{R}$.

3.3.1 Bounds on $p(y;F)$

In what follows, we derive upper and lower bounds on the output probability distribution induced by an input distribution F.

Lemma 2 (Lower Bound on the Output PDF)**.** *Let $y_0 > 0$ be sufficiently large. For an input distribution F, the PDF $p(y;F)$ of the output of channel (3.3) is lower bounded by*

$$p(y;F) \geq \begin{cases} \dfrac{T_l(y - y_0)}{2} & y \leq -y_0 \\ \dfrac{T_l(y + y_0)}{2} & y \geq \quad y_0, \end{cases}$$

Proof. Given an input probability distribution F, we define the following:

- We denote by d_F a positive constant such that $\Pr(|X| \leq d_F) \geq \frac{1}{2}$.
- We denote by $f_{\max} = \sup_{|x| \leq d_F} |f(x)|$, the existence of which is guaranteed by the assumption that $f(\cdot)$ is continuous on $\mathbb{R}$.

Let $y_0 > f_{\max}$. In what follows, we only present in detail the case $y \geq y_0$ as the proof in the other range follows similar steps.

$$\begin{aligned} p_Y(y;F) &\geq \int_{x:|x|\leq d_F} p_N\big(y - f(x)\big)\,dF(x) \\ &\geq \int_{x:|x|\leq d_F} T_{\rm l}\,(y - f(x))\,dF(x) &(3.7)\\ &\geq \frac{1}{2}T_{\rm l}\,(y + f_{\max}) \geq \frac{1}{2}T_{\rm l}\,(y + y_0)\,, &(3.8)\end{aligned}$$

where equation (3.7) is due to the fact that $T_{\rm l}\,(\cdot)$ is a lower bound on $p_N(\cdot)$ by definition and inequalities (3.8) are justified since $T_{\rm l}\,(\cdot)$ is non-increasing on the considered interval. □

We also derive an upper bound on the output law whenever the input is bounded within $[-B, B]$ for some $B > 0$:

Lemma 3 (Upper Bound on the Output PDF)**.** *For an input distribution F that has a bounded support within $[-B, B]$ for some $B > 0$, the PDF $p(y;F)$ of the output of channel (3.3) is upper bounded by*

$$p(y;F) \leq \begin{cases} T_u\left(y + y_0^B\right) & y \leq -y_0^B \\ T_u\left(y - y_0^B\right) & y \geq \quad y_0^B, \end{cases}$$

for any large-enough y_0^B.

Proof. Let $f_{\max}^B = \sup_{[-B;B]} |f(x)|$, the existence of which is guaranteed by the fact that $f(\cdot)$ is continuous on $\mathbb{R}$. Also let $y_0^B \geq f_{\max}^B$. For $y \geq y_0^B$, since $T_{\rm u}\,(\cdot)$ is an upper bound on $p_N(\cdot)$, we have,

$$\begin{aligned} p(y;F) &= \int p_N(y - f(x))\,dF(x) \\ &= \int_{-B}^{B} p_N(y - f(x))\,dF(x) &(3.9)\\ &\leq \int_{-B}^{B} T_{\rm u}\,(y - f(x))\,dF(x) \\ &\leq T_{\rm u}\left(y - f_{\max}^B\right) \leq T_{\rm u}\left(y - y_0^B\right) &(3.10)\end{aligned}$$

where equations (3.10) are due to the fact that $T_{\mathrm{u}}(x)$ is non-increasing on the positive semi-axis. A similar derivation yields the result for $y \leq -y_0^B$. □

We emphasize that this lower bound on $p(y;F)$ is only possible under the assumption that the support of F is bounded (as seen in equation (3.9)).

3.3.2 Bounds on $i(x;F)$

In this section we analyze the function of interest

$$\begin{aligned} i(x;F) &= -\int_{-\infty}^{+\infty} p_N(y-x) \ln p_Y(y;F)\, dy \qquad (3.11) \\ &= -\int_{-\infty}^{+\infty} p_N(y) \ln p_Y(y+x;F)\, dy. \end{aligned}$$

Lemma 4 (Upper Bound on $i(x;F)$)**.** *For any probability distribution F,*

$$i(x;F) = O\left(\ln\left[\frac{1}{T_l(x)}\right]\right).$$

Proof. Consider a large-enough y_0 so that Lemma 2 holds, and let x be such that $x > y_0$. For a probability distribution F on the input we compute,

$$i(x;F) = -\int_{-\infty}^{+\infty} p_N(y) \ln p_Y(y+x;F)\, dy = I_1 + I_2 + I_3,$$

where the interval of integration is divided into three sub-intervals: $(-\infty, -x - y_0)$, $[-x-y_0, y_0]$, $(y_0, +\infty)$.

We study the growth rate in x of the integral terms I_1, I_2 and I_3 function of the rate of decay of $T_{\mathrm{l}}(\cdot)$.

Using Lemma 2,

$$
\begin{aligned}
I_1 &= -\int_{-\infty}^{-x-y_0} p_N(y) \ln p_Y(y+x;F)\, dy \\
&\leq -\int_{-\infty}^{-x-y_0} p_N(y) \ln \left[\frac{T_l\,(y+x-y_0)}{2}\right] dy = \int_{-\infty}^{-x-y_0} p_N(y) \ln \left[\frac{2}{T_l\,(y+x-y_0)}\right] dy \\
&\leq \ln 2 + \kappa \int_{-\infty}^{-x-y_0} p_N(y) \left(\ln\left[\frac{1}{T_l\,(y)}\right] + \ln\left[\frac{1}{T_l\,(x)}\right] + \ln\left[\frac{1}{T_l\,(-y_0)}\right]\right) dy \qquad (3.12)\\
&\leq \ln 2 + \kappa \, \ln\left[\frac{1}{T_l\,(-y_0)}\right] + \kappa \, \ln\left[\frac{1}{T_l\,(x)}\right] + \kappa \int_{-\infty}^{+\infty} p_N(y) \ln\left[\frac{1}{T_l\,(y)}\right] dy \qquad (3.13)\\
&\leq 2\kappa \, \ln\left[\frac{1}{T_l\,(x)}\right],
\end{aligned}
$$

for some positive κ and for $x > y_0$ large-enough. Equation (3.12) is due to property C7 since both x and y_0 are large enough and so is $|y|$. The integral term in (3.13) is finite by property C8 and the last equation is valid since $\ln\left[\frac{1}{T_l(x)}\right]$, which is positive, is increasing to $+\infty$.

Similarly,

$$
\begin{aligned}
I_3 &= -\int_{y_0}^{\infty} p_N(y) \ln p_Y(y+x;F)\, dy \\
&\leq -\int_{y_0}^{\infty} p_N(y) \ln \left[\frac{T_l\,(y+x+y_0)}{2}\right] dy = \int_{y_0}^{\infty} p_N(y) \ln \left[\frac{2}{T_l\,(y+x+y_0)}\right] dy \\
&\leq \ln 2 + \kappa \int_{y_0}^{\infty} p_N(y) \left(\ln\left[\frac{1}{T_l\,(y)}\right] + \ln\left[\frac{1}{T_l\,(x)}\right] + \ln\left[\frac{1}{T_l\,(y_0)}\right]\right) dy \\
&\leq 2\kappa \, \ln\left[\frac{1}{T_l\,(x)}\right],
\end{aligned}
$$

As for I_2,

$$\begin{aligned}
I_2 &= -\int_{-x-y_0}^{y_0} p_N(y) \ln p_Y(y+x;F)\,dy \\
&= -\int_{-x-y_0}^{-x+y_0} p_N(y) \ln p_Y(y+x;F)\,dy - \int_{-x+y_0}^{y_0} p_N(y) \ln p_Y(y+x;F)\,dy \\
&\leq \sup_{|y|\leq y_0} \ln\left[\frac{1}{p_Y(y;F)}\right] + \int_{-x+y_0}^{y_0} p_N(y) \ln\left[\frac{2}{T_l\,(y+x+y_0)}\right] dy \\
&\leq \sup_{|y|\leq y_0} \ln\left[\frac{1}{p_Y(y;F)}\right] + \ln 2 + \ln\left[\frac{1}{T_l\,(x+2y_0)}\right] \qquad (3.14) \\
&\leq \sup_{|y|\leq y_0} \ln\left[\frac{1}{p_Y(y;F)}\right] + \ln 2 + \kappa \ln\left[\frac{1}{T_l\,(x)}\right] + \kappa \ln\left[\frac{1}{T_l\,(2y_0)}\right] \qquad (3.15) \\
&\leq 2\,\kappa\,\ln\left[\frac{1}{T_l\,(x)}\right].
\end{aligned}$$

The supremum is finite since it is taken over a compact set where $p_Y(y)$ (which is less than one) is positive, continuous and hence positively lower bounded. Equation (3.14) is due to the fact that $\ln\left[\frac{1}{T_l(\cdot)}\right]$ is non-decreasing on the positive axis, equation (3.15) is given by property C7 since both x and y_0 are large enough and the last equation is justified since $\ln\left[\frac{1}{T_l(x)}\right]$ is increasing to $+\infty$ as $|x| \to +\infty$.

A similar procedure can be adopted to prove this result when $x \to -\infty$ by adjusting the intervals of integration to the following: $(-\infty, -y_0)$, $[-y_0, -x+y_0]$, $(-x+y_0, +\infty)$ where $x < -y_0$ such that $|x|$ is large enough. This would imply that for any probability distribution F, $i(x;F) = I_1 + I_2 + I_3 = O\left(\ln\left[\frac{1}{T_l(x)}\right]\right)$. □

We also derive a lower bound whenever the input is bounded within $[-B, B]$ for some $B > 0$:

Lemma 5 (Lower Bound on $i(x;F)$)**.** *For an input distribution F that has a bounded support within $[-B, B]$ for some $B > 0$,*

$$i(x;F) = \Omega\left(\ln\left[\frac{1}{T_u\,(x)}\right]\right).$$

Proof. We proceed in a manner akin to the proof of Lemma 4: For an input distribution F that has a bounded support within $[-B, B]$ for some $B > 0$, we consider a large-enough y_0^B so that Lemma 3 holds, and let x be such that $x > y_0^B$.

$$
\begin{aligned}
i(x;F) &= -\int_{-\infty}^{\infty} p_N(y) \ln p(y+x;F)\, dy \\
&\geq -\int_{y_0^B}^{+\infty} p_N(y) \ln p(y+x;F)\, dy \\
&\geq \int_{y_0^B}^{+\infty} p_N(y) \ln \left[\frac{1}{T_{\mathrm{u}}\left(y+x-y_0^B\right)} \right] dy \qquad (3.16) \\
&\geq \left(1 - F_N(y_0^B)\right) \ln \left[\frac{1}{T_{\mathrm{u}}(x)} \right] > 0. \qquad (3.17)
\end{aligned}
$$

In order to write equation (3.16) we use the upper bound in Lemma 3. Equation (3.17) is justified since $\ln \left[\frac{1}{T_{\mathrm{u}}(\cdot)}\right]$ is non-decreasing on the non-negative semi-axis and the end result is positive since the support of N is $\mathbb{R}$. A similar analysis may be conducted for the case when $x < -y_0^B < 0$. □

3.4 The Karush-Kuhn-Tucker (KKT) Theorem

The capacity of channel (3.1) is the supremum of the mutual information $I(\cdot)$ between the input X and output Y over all input probability distributions F that meet the constraint $\mathcal{P}_A$:

$$
C = \sup_{F \in \mathcal{P}_A} I(F) = \sup_{F \in \mathcal{P}_A} \iint p_N\left(y - f(x)\right) \ln \left[\frac{p_N\left(y - f(x)\right)}{p(y;F)} \right] dy\, dF(x). \quad (3.18)
$$

Conditions C1 to C6 guarantee that this optimization problem is well-defined and that its solution –the capacity– is finite and is achievable (see Theorem 2 in Appendix B). Indeed, the conditions are sufficient for $\mathcal{P}_A$ to be convex and compact (Theorem 3, Appendix B) and for $I(\cdot)$ to be concave and continuous (in the weak sense [84, Sec.III.7]) (Theorems 4,5, Appendix B).

When dealing with constrained optimization problems, the Lagrangian theorem [85] is a useful tool as it transforms the problem to an unconstrained one when some convexity conditions are satisfied by the objective function and the constraints. In our problem these conditions are satisfied as the mutual information is concave and the cost is linear - and hence convex. The theorem states that there exists a non-negative parameter ν_A such that the optimization problem (3.18) can be written as:

$$C = \sup_{F \in \mathcal{P}_A} I(F) = \sup_{F} \left\{ I(F) - \nu_A \big(\mathsf{E}_F \left[\mathcal{C} \left(|X| \right) \right] - A \big) \right\} \tag{3.19}$$
$$= I(F^*) - \nu_A \, \mathsf{E}_{F^*} \left[\mathcal{C} \left(|X| \right) \right] + \nu_A A,$$

where the last equality is true since the solution is finite and achievable by an optimal F^*. Furthermore,

$$\nu_A \left(\mathsf{E}_{F^*} \left[\mathcal{C} \left(|X| \right) \right] - A \right) = 0.$$

For every positive A, denote by $C(A)$ the capacity of the channel under the constraint $F \in \mathcal{P}_A$, and consider the function $C(A)$ for $A > 0$. The significance of the Lagrange parameter ν_A is addressed in the following Lemma.

Lemma 6 (Non-Binding Constraint)**.** *Whenever for some positive A the parameter $\nu(A) = 0$, then $C(A') = C(A)$ for all $A' \geq A$.*

Proof. We start by noting that the channel capacity $C(A)$ is a non-decreasing function of A, due to the fact $\mathcal{P}_A \subseteq \mathcal{P}_{A'}$, for $0 < A \leq A'$. Now assume that $\nu(A) = 0$ for some $A > 0$. For this value of A, equation (3.19) becomes

$$C = \sup_{F \in \mathcal{P}_A} I(F) = \sup_{F} \left\{ I(F) - \nu_A \big(\mathsf{E}_F \left[\mathcal{C} \left(|X| \right) \right] - A \big) \right\} = \sup_{F} I(F),$$

which is a maximal value over all probability distributions irrespective of the constraint. This observation along with the fact that $C(A)$ is non-decreasing establish the result. □

In our setup, a value of $\nu(A) = 0$ can be ruled out. Said differently, the cost constraint in equation (3.18) is binding. The argument we make is similar to the one used in [59]: we consider a family of input signals composed of N discrete levels with equal probabilities at locations $\{1, L, L^2, \cdots, L^{2^{N-2}}\}$. When L increases, the probability of error of a minimum probability of error receiver goes to zero, which implies by Fano's inequality that the mutual information approaches $\ln(N)$. Therefore, as $A \to \infty$, the achievable rates in our setup are arbitrarily large and $C(A)$ increases to infinity; a fact that is not possible if $\nu(A)$ were equal to zero for some A by Lemma 6. This conclusion is corroborated by the fact that the capacity for general continuous-input, continuous-output channels is achieved by a boundary input when the input power is assumed to be the second moment [86]. Our statement here holds for a general input constraint $\mathcal{C}(\cdot)$.

Whenever weak (Gateaux) differentiability is guaranteed, one can further write necessary and sufficient conditions on the maximum achieving distribution; conditions that are commonly referred to as the KKT conditions [85]. More formal statements on the theory of convex optimization are summarized in Appendix A. The KKT approach was used previously in many studies [54,57,59–66, 68,69] in order to solve the capacity problem and for the purpose of this work, we follow similar steps. We indeed prove in Appendix C the weak differentiability of $I(\cdot)$ at any optimal input F^* and proceeding as in [59], we write the KKT conditions as being necessary and sufficient conditions for the optimal input to satisfy. These conditions state that an input RV X^* with probability distribution F^* achieves the capacity C of an average cost constrained channel if and only if there exists $\nu \geq 0$ such that,

$$\begin{aligned} \nu(\mathcal{C}(|x|) - A) + C + H + \int p_N(y - f(x)) \ln p(y; F^*)\, dy \\ = \nu(\mathcal{C}(|x|) - A) + C + H - i(f(x); F^*) \geq 0, \quad (3.20) \end{aligned}$$

for all x in $\mathbb{R}$, with equality if x is a point of increase of F^*, and where H is the entropy of the noise.

3.5 Main Results

Theorem 8 (Compactly Supported Capacity Achieving Input). *Whenever* $\mathcal{C}(|x|) = \omega\left(\ln\left[\frac{1}{T_l(f(x))}\right]\right)$, *the support of the capacity-achieving input of channel (3.3) is compact.*

Proof. We consider the necessary and sufficient conditions of optimality (3.20), and we study the behavior of the expression function of the variable x as its magnitude goes to infinity.

These conditions state that for the optimal input X^*, condition (3.20) is satisfied with equality for any point of increase x_0 of the capacity-achieving distribution $F^* \in \mathcal{P}_A$. For such an x_0 we obtain,

$$\nu(\mathcal{C}(|x_0|) - A) + C + H = i(f(x_0); F^*).$$

Using Lemma 4, $i(f(x); F) = O\left(\ln\left[\frac{1}{T_l(f(x))}\right]\right)$ since $|f(x)| \to +\infty$ as $|x| \to +\infty$ and therefore

$$\nu(\mathcal{C}(|x_0|) - A) + C + H = O\left(\ln\left[\frac{1}{T_l(f(x))}\right]\right).$$

This implies that whenever $\mathcal{C}(|x|) = \omega\left(\ln\left[\frac{1}{T_l(f(x))}\right]\right)$ the points of increase of X^* could not assume arbitrarily large values unless $\nu = 0$, which has been ruled out in Section 3.4, which implies that the support of X^* is bounded. Finally, we note that the support is always closed, as its complement is open. Therefore, X^* is compactly supported. $\square$

3.5.1 A Converse Theorem

Now we make use of the upper bound on the noise PDF. In this section, we state and prove a converse formulation of Theorem 8. Indeed we prove that whenever $\mathcal{C}(|x|) = o\left(\ln\left[\frac{1}{T_{\mathrm{u}}(f(x))}\right]\right)$, the capacity-achieving input is not bounded.

Theorem 9 (Unbounded Support Capacity Achieving Input). *Whenever* $\mathcal{C}(|x|) = o\left(\ln\left[\frac{1}{T_{u}(f(x))}\right]\right)$, *the support of the capacity-achieving input of channel (3.3) is unbounded.*

Proof. Suppose that the optimal input X^* with distribution function F^* has a bounded support within $[-B, B]$ for some $B > 0$. The KKT conditions imply that there exists $\nu \geq 0$ such that,

$$\nu(\mathcal{C}(|x|) - A) + C + H + \int p_N(y - f(x)) \ln p(y; F^*)\, dy \geq 0,$$

for all x in $\mathbb{R}$, with equality if x is a point of increase of F^*. Using Lemma 4, the integral term $i(f(x); F^*) = \Omega\left(\ln\left[\frac{1}{T_{\mathrm{u}}(f(x))}\right]\right)$ and hence, equation (3.20) necessarily implies that,

$$\nu(\mathcal{C}(|x|) - A) + C + H = \Omega\left(\ln\left[\frac{1}{T_{\mathrm{u}}(f(x))}\right]\right),$$

which is impossible whenever $\mathcal{C}(|x|) = o\left(\ln\left[\frac{1}{T_{\mathrm{u}}(f(x))}\right]\right)$. □

3.5.2 Discreteness

In what follows, we further characterize the capacity-achieving input statistics when the cost function, the noise PDF and the channel distortion function have an additional analyticity property. This property guarantees the type of the optimal bounded input to be a discrete one, and hence with a finite number of mass points by virtue of compactness. This characterization permits to proceed to numerical computations in order to compute channel capacity and find the achieving input.

In this section, let $\eta > 0$ denote a positive scalar and let $\mathcal{S}_\eta = \{z \in \mathbb{C} : |\Im(z)| < \eta\}$ be a horizontal strip in the complex domain. We adopt in this section an alternative definition of $T_{\mathrm{u}}(x)$:

$$T_{\mathrm{u}}(x) = \begin{cases} \sup\limits_{\zeta \in \mathcal{S}_\eta : \Re(\zeta) \geq x} |p_N(\zeta)| & x \geq 0 \\ \sup\limits_{\zeta \in \mathcal{S}_\eta : \Re(\zeta) \leq x} |p_N(\zeta)| & x < 0, \end{cases} \tag{3.21}$$

and we assume that the following condition holds: The integral

$$-\int_{-\infty}^{+\infty} T_{\mathrm{u}}(x) \ln T_{\mathrm{l}}(x)\, dx$$

exists and is finite. Note that this condition is similar to C8 but it is function of a redefined $T_{\mathrm{u}}()$. One may think of the condition as more restrictive. However, this strengthened condition is needed only to establish discreteness. In the remainder of this document we will refer to this condition as "the strengthened-C8". We present hereafter, a lemma that guarantees the analyticity of $i(\cdot; F)$ on $\mathcal{S}_\eta$:

Lemma 7 (Analyticity of $i(\cdot; F)$)**.** *Whenever there exists an $\eta > 0$ such that $p_N(\cdot)$ admits an analytic extension on $\mathcal{S}_\eta$, the function $i(\cdot; F) : \mathcal{S}_\eta \to \mathbb{C}$ defined by:*

$$z \to i(z; F) = -\int_{-\infty}^{\infty} p_N(y - z)\, \ln\, p(y; F)\, dy, \tag{3.22}$$

is analytic.

Proof. To prove this lemma, we will make use of Morera's theorem:

a) We start first by proving the *continuity* of $i(\cdot; F)$. In fact, let $\rho > 0$, z_0

and $z \in \mathcal{S}_\eta$ such that $|z - z_0| \leq \rho$,

$$\begin{aligned}
\lim_{z \to z_0} i(z; F) &= - \lim_{z \to z_0} \int p_N(y - z) \, \ln p(y; F) \, dy \\
&= - \int \lim_{z \to z_0} p_N(y - z) \, \ln p(y; F) \, dy \qquad (3.23) \\
&= - \int p_N(y - z_0) \ln p(y; F) \, dy = i(z_0; F). \qquad (3.24)
\end{aligned}$$

Equation (3.24) is justified by $p_N(y - z)$ being a continuous function of z on $\mathcal{S}_\eta$ by virtue of its analyticity and equation (3.23) by Lebesgue's Dominated Convergence Theorem (DCT). Indeed, in what follows we find an integrable function $r(y)$ such that,

$$\big|p_N(y - z) \, \ln \, p(y; F)\big| = -\big|p_N(y - z)\big| \ln \, p(y; F) \leq r(y),$$

for all $z \in \mathcal{S}_\eta$ such that $|z - z_0| \leq \rho$ and for all $y \in \mathbb{R}$. We upper bound first $|p_N(y - z)|$: let y_0 be large enough so that Lemma 2 holds

- If $y \leq -(y_0 + |\Re(z_0)| + \rho)$, then $y \leq -y_0 + \Re(z_0) - \rho$ (where y_0 has been defined in Lemma 2) and

$$\begin{aligned}
|p_N(y - z)| \leq T_\mathrm{u}\left(y - \Re(z)\right) &\leq \max_{\zeta \in \mathcal{S}_\eta : |\zeta - z_0| \leq \rho} T_\mathrm{u}\left(y - \Re(\zeta)\right) \\
&= T_\mathrm{u}\left(y - \Re(z_0) + \rho\right),
\end{aligned}$$

  where the last equality is due to the fact that for $x \leq 0$, $T_\mathrm{u}(x)$ is non-decreasing, and for $\zeta \in \mathcal{S}_\eta ; |\zeta - z_0| \leq \rho$, $(y - \Re(\zeta)) \leq (y - \Re(z_0) + \rho) < 0$.

- Similarly, for $y \geq (y_0 + |\Re(z_0)| + \rho) \geq (y_0 + \Re(z_0) + \rho)$,

$$\big|p_N(y - z)\big| \leq T_\mathrm{u}\left(y - \Re(z_0) - \rho\right).$$

Next, using Lemma 2 we also upper bound $-\ln p(y;F)$ to obtain:

$$r(y) = \begin{cases} T_{\mathrm{u}}\left(y - \Re(z_0) + \rho\right) \ln\left[\dfrac{2}{T_{\mathrm{l}}\left(y - y_0\right)}\right] & y \leq -(y_0 + |\Re(z_0)| + \rho) \\ -M \ln M^{'} & |y| < y_0 + |\Re(z_0)| + \rho \\ T_{\mathrm{u}}\left(y - \Re(z_0) - \rho\right) \ln\left[\dfrac{2}{T_{\mathrm{l}}\left(y + y_0\right)}\right] & y \geq y_0 + |\Re(z_0)| + \rho, \end{cases}$$

where

$$M = \max_{\{|y| \leq (y_0 + |\Re(z_0)| + \rho)\}} \max_{\{\zeta \in \mathcal{S}_\eta : |\zeta - z_0| \leq \rho\}} |p_N(y - \zeta)| \quad \& \quad M^{'} = \min_{\{|y| \leq (y_0 + |\Re(z_0)| + \rho)\}} p_Y(y;F).$$

Note that M is finite since $p_N(\cdot)$ is analytic and the maximization is taken over a compact set, and $0 < M^{'} < 1$, since $p_Y(\cdot;F)$ is positive, continuous and less than 1. Properties C7 and strengthened-C8 insure the integrability of $r(y)$ which concludes the proof of continuity of $i(z;F)$.

b) To continue the proof of analyticity, we need to integrate $i(\cdot;F)$ on the boundary $\partial\Delta$ of a compact triangle $\Delta \subset \mathcal{S}_\eta$. We denote by $|\Delta|$ its perimeter, $\eta_0 = \min_{z\in\partial\Delta} \Re(z)$, $\eta_1 = \max_{z\in\partial\Delta} \Re(z)$ and $\phi = y_0 + \max\{|\eta_0|, |\eta_1|\}$. By similar arguments as above, we have

$$\begin{aligned} &\int_{\mathbb{R}} \int_{\partial\Delta} |p_N(y - z)| |\ln p(y;F)| dz\, dy \\ &\leq |\Delta|\, M^{''} \int_{|y|\leq\phi} \left|\ln p(y;F)\right| dy + |\Delta| \int_{y\leq-\phi} T_{\mathrm{u}}\left(y - \eta_0\right) \ln\left[\frac{2}{T_{\mathrm{l}}\left(y - y_0\right)}\right] dy \\ &\qquad + |\Delta| \int_{y\geq\phi} T_{\mathrm{u}}\left(y - \eta_1\right) \ln\left[\frac{2}{T_{\mathrm{l}}\left(y - y_0\right)}\right] dy < \infty, \end{aligned}$$

where

$$M^{''} = \max_{y:|y|\leq\phi} \max_{\xi\in\partial\Delta} |p_N(y - \xi)| < \infty.$$

Using Fubini's theorem to interchange the order of integration,

$$\begin{aligned}\int_{\partial\Delta} i(z;F)dz &= -\int_{\partial\Delta}\int_{\mathbb{R}} p_N(y-z)\ln p(y;F)\,dy\,dz \\ &= -\int_{\mathbb{R}}\int_{\partial\Delta} p_N(y-z)\ln p(y;F)\,dz\,dy \\ &= -\int_{\mathbb{R}} \ln p(y;F)\int_{\partial\Delta} p_N(y-z)\,dz\,dy = 0, \end{aligned} \tag{3.25}$$

where (3.25) is justified by the fact that $p_N(y-z)$ is analytic for all $z \in \mathcal{S}_\eta$ and $y \in \mathbb{R}$. Equation (3.25) in addition to the continuity of $i(\cdot;F)$ insure its analyticity on $\mathcal{S}_\eta$. □

Theorem 10 (Discreteness of the Capacity Achieving Input)**.** *Assume there exists an $\eta > 0$ such that $p_N(x)$ is analytically extendable on $\mathcal{S}_\eta$, and let $\mathcal{I}$ be an unbounded closed interval of $\mathbb{R}$[4]. The capacity-achieving input of channel (3.3) is compactly supported and discrete with finite number of mass points on $\mathcal{I}$, whenever the following conditions hold:*

- $\mathcal{C}\left(|x|\right) = \omega\left(\ln\left[\frac{1}{T_l(f(x))}\right]\right)$.
- *The restrictions of $f(x)$ and $\mathcal{C}\left(|x|\right)$ on $\mathcal{I}$ admit analytic extensions to $\mathcal{I}\times\mathbb{R}$, denoted $f_{\mathcal{I}}(\cdot)$ and $\mathcal{C}_{\mathcal{I}}(\cdot)$ respectively.*
- *The inverse map $f_{\mathcal{I}}^{-1}(\cdot)$ of $f_{\mathcal{I}}(\cdot)$ conserves connectedness.*

Before we prove the theorem, we note that a necessary condition for analytical extendability is to have $\mathcal{C}\left(|x|\right)$ an explicit function of the variable x on $\mathcal{I}$ which can be possibly realized when $\mathcal{I}$ is for example a subset of either $\mathbb{R}^+$ or $\mathbb{R}^-$.

Proof. We start by setting some notation and making a few remarks:

- Define $\mathcal{J}$ to be the image of interval $\mathcal{I}$ by $f_{\mathcal{I}}(\cdot)$.

[4] *We consider that $\mathbb{R}$ is both closed and open.*

Since by analyticity $f_{\mathcal{I}}(\cdot)$ is continuous, then $\mathcal{J}$ is an interval of $\mathbb{R}$ because $f_{\mathcal{I}}(\cdot)$ is identical to $f(\cdot)$ on $\mathcal{I}$, and is real valued.

- Let $\mathcal{J}_\eta = \{z \in \mathcal{S}_\eta : \Re(z) \in \mathcal{J}\}$ and define $\mathcal{I}_\eta = f_{\mathcal{I}}^{-1}(\mathcal{J}_\eta)$, the inverse image of $\mathcal{J}_\eta$ by $f_{\mathcal{I}}(\cdot)$.

  Note that since $\mathcal{J}$ is an interval, $\mathcal{J}_\eta$ is connected and so is $\mathcal{I}_\eta$ by virtue of the properties of $f_{\mathcal{I}}^{-1}(\cdot)$. Additionally, since $f_{\mathcal{I}}(\mathcal{I}) = \mathcal{J}$ then $\mathcal{I} \subset \mathcal{I}_\eta$.

  In what follows, we work using the induced topology on $\mathcal{I}_\eta$. Under this topology, $\mathcal{I}_\eta$ is both open and closed.

We proceed with the proof and assume that the optimal input X^* with distribution function F^* has at least one point of increase in $\mathcal{I}$ for otherwise the result becomes trivial. *Assume that the points of increase of F^* in $\mathcal{I}$ are accumulating*, and let

$$s(z) = \nu\left(\mathcal{C}_{\mathcal{I}}(z) - A\right) + C + H - i(f_{\mathcal{I}}(z); F^*).$$

By the result of Lemma 7, $i(f_{\mathcal{I}}(z); F^*)$ is analytic on $\mathcal{I}_\eta$ since it is the composition of two analytic functions: $f_{\mathcal{I}}(\cdot)$ on $\mathcal{I}_\eta$ and $i(\cdot; F^*)$ on $\mathcal{J}_\eta = f_{\mathcal{I}}(\mathcal{I}_\eta) \subset \mathcal{S}_\eta$. This implies that the function $s(z)$ is analytic on $\mathcal{I}_\eta$. Since by assumption the points of increase of F^* have an accumulation point on $\mathcal{I}$ then by the KKT conditions, $s(z)$ has accumulating zeros on $\mathcal{I} \subset \mathcal{I}_\eta$, which necessarily implies by the identity Theorem [87, sec. 66] that $s(\cdot)$ is identically null on $\mathcal{I}_\eta$, since $\mathcal{I}_\eta$ is open and connected. Therefore,

$$\nu(\mathcal{C}\left(|x|\right) - A) + C + H = -\int p_N(y) \ln p(y - f(x); F^*)\, dy, \qquad \forall x \in \mathcal{I}.$$

Since $\mathcal{I}$ is unbounded, this equality is impossible for large values of x by the result of Theorem 8 unless $\nu = 0$ which is non sensible. This leads to a contradiction and rules out the assumption of having an accumulation point on $\mathcal{I}$. Since $\mathbb{R}$ is Lindelof, X^* is necessarily discrete on $\mathcal{I}$. Knowing that $\mathcal{C}\left(|x|\right) =$

$\omega\left(\ln\left[\frac{1}{T_{\rm l}(f(x))}\right]\right)$ then the support of X^* is compact (Theorem 8). The fact that $\mathcal{I}$ is closed in $\mathbb{R}$ implies that X^* has necessarily a finite number of mass points on $\mathcal{I}$. $\square$

Before moving to applying our general theorems to the alpha-stable channels, we would like to state that some conditions were only considered for either the sake of the clarity of the proofs, or for conserving the general aspect of the results. Many such conditions could be relaxed while conserving some or all of the found conclusions. For example,

- The notions of ω, Ω, o and O used in this document are defined as $|x| \to +\infty$, i.e., in such a way to capture a symmetric rate of decay for both tails. However, one can only consider left or right tail behaviors separately. The results of boundedness and discreteness could be given in terms of each tail where for example for non-symmetric noise PDFs or non-symmetric cost functions, the optimal input could only be bounded on one of the semi-axis.

- For Theorems 8 and 9, the assumption that $p_N(\cdot)$ is positive could be relaxed to one sided noise PDFs. These theorems are still valid on one side of the axis.

- The proven theorems –stated in terms of $T_{\rm l}\left(x\right)$ and $T_{\rm u}\left(x\right)$– could be stated in terms of any two functions having the same properties and providing lower and upper bounds on $p_N(x)$ for large values of $|x|$.

3.6 The Alpha-Stable Channel Capacity

In this section we apply our results –in the form of Theorems 8, 9, and 10 as follows:

1- We verify the results by applying them to the Gaussian channel that has been previously studied in the literature.

2- We give new capacity results and compute numerically the channel capacity for two types of channels of interest that present practical models when communicating in impulsive noise environments:

   – The additive noise is a non-totally skewed alpha-stable RV or a mixture of a finite number of them.

   – The additive noise is the sum of two independent RVs: a Gaussian one and a non-totally skewed alpha-stable one.

We note that in all the examples presented subsequently the considered functions $f(\cdot)$ and the cost constraints satisfy the general conditions C1 through C4 in Section 3.2.1. The noise distributions are absolutely continuous with positive, continuous PDFs with tails that have "at least" a polynomial decay and hence satisfying the assumptions C5 and C6. Finally, in all the provided examples the noise PDFs possess a monotonic tail and a finite differential entropy and therefore, condition C8 is satisfied. It remains to check for each example condition C7 and possibly the strengthened-C8.

For the purpose of verifying condition C7, we note that one can use $p_N(x)$ instead of $T_{\text{l}}(x)$ since they are identical at large values of $|x|$. When it comes to discreteness, whenever $|x|$ is large enough the function $T_{\text{u}}(x)$ defined in (3.21) becomes

$$T_{\text{u}}(x) = \sup_{\{z:\, \Re(z)=x \,\&\, |\Im(z)|<\eta\}} |p_N(z)|\,,$$

because $|p_N(z)|$ is decreasing with $|\Re(z)|$ at large values for all the given examples.

For each model we consider in what follows, we will check whether the appropriate conditions are satisfied, state the results –specialized to the channel at hand, and compare with the known results in the literature, if any.

3.6.1 The Gaussian Model

For a Gaussian noise distribution with mean zero and variance σ^2, the PDF is $p_N(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$ and we write $N \sim \mathcal{N}(0, \sigma^2)$.

Checking the conditions: Condition C7 is validated as follows: for large values of $|x|$ and $|y|$,

$$\begin{aligned} L(x+y) &= \ln\left[\frac{1}{p_N(x+y)}\right] = \ln\sqrt{2\pi\sigma^2} + \frac{(x+y)^2}{2\sigma^2} \\ &\leq 2\left(\ln\sqrt{2\pi\sigma^2} + \frac{x^2}{2\sigma^2} + \ln\sqrt{2\pi\sigma^2} + \frac{y^2}{2\sigma^2}\right) - 3\ln\sqrt{2\pi\sigma^2} \\ &= \kappa_{\mathrm{l}}\left(L(x) + L(y)\right), \end{aligned}$$

where $\kappa_{\mathrm{l}} > 2$. When it comes to discreteness, let $p_N(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{z^2}{2\sigma^2}}$, be an analytic extension of $p_N(x)$ to the complex plane, where $z = x + jy$. The magnitude of $p_N(z)$ is

$$|p_N(z)| = \frac{1}{\sqrt{2\pi\sigma^2}}\left|e^{-\frac{z^2}{2\sigma^2}}\right| = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2-y^2}{2\sigma^2}},$$

and is decreasing in $x = \Re(z)$. Therefore, $T_{\mathrm{u}}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2-\eta^2}{2\sigma^2}} = e^{\frac{\eta^2}{2\sigma^2}} p_N(x)$.

Checking for the strengthened-C8, the integral $-\int_{-\infty}^{+\infty} T_{\mathrm{u}}(x) \ln T_{\mathrm{l}}(x)\, dx = e^{\frac{\eta^2}{2\sigma^2}} h(N)$ which is finite because the noise differential entropy $h(N)$ is finite.

The following theorem is a specialization of Theorems 8 and 9 for this specific Gaussian case:

Theorem 11 (Capacity Results: Gaussian Noise)**.** *Whenever $\mathcal{C}(|x|) = o(f(x)^2)$, the support of the capacity-achieving input of channel (3.3) when $N \sim \mathcal{N}(0, \sigma^2)$ is unbounded.*

Whenever $\mathcal{C}(|x|) = \omega(f(x)^2)$, the support of the capacity-achieving input of channel (3.3) when $N \sim \mathcal{N}(0, \sigma^2)$ is compact. Furthermore, the optimal input is discrete with finite number of mass points whenever $\mathcal{C}(\cdot)$ and $f(\cdot)$ satisfy the analyticity and connectedness conditions of Theorem 10.

Previous work: A possibly non-linear ($f(x) = x^n$, $n \in \mathbb{N}^*$) Gaussian channel under an even moment constraint ($\mathcal{C}\left(|x|\right) = x^{2k}$) was considered in [49] as a core channel model from which results on multiple non-linear channel models were derived. The authors applied a standard Hilbert space decomposition using Hermite polynomials as bases and proved that, for $n < 2k$, the capacity-achieving distribution has the following behavior:

- Whenever $n = k$
  - if n is odd, the optimal input F^* is absolutely continuous.
  - if n is even, F^* is discrete with no accumulation points.
- Whenever $n < k$, F^* is discrete with finite number of mass points.
- Whenever $k < n < 2k$, F^* is discrete with no accumulation points.

We point out that while the results stated in Theorem 11 do not cover the limiting case $n = k$ –which corresponds to the case $\mathcal{C}\left(|x|\right) = \theta\left(f^2(x)\right)$, the result for the case "$n < k$" is identical. Whenever $k < n$, Theorem 11 states that the support of F^* is not bounded; a conclusion that could not be reached in [49].

3.6.2 Polynomially-Tailed Distributions

In this section, we refer by "polynomially-tailed" noise distributions to all distributions satisfying

$$p_N(x) = \Theta\left(\frac{1}{|x|^{1+\alpha}}\right), \quad \text{for some } \alpha > 0,$$

which include among others: the Gamma, Pareto (one sided) and alpha-stable distributions.

Checking the conditions: In order to proceed, we use the "obvious" lower and upper bounds on $p_N(x)$ for large values of $|x|$ instead of $p_N(x)$ itself and we state the corresponding theorems accordingly. These bounds are of the form $\frac{\zeta_1}{|x|^{1+\alpha}}$ and

$\frac{\zeta_{\mathrm{u}}}{|x|^{1+\alpha}}$, for some ζ_{l} and $\zeta_{\mathrm{u}} > 0$. We prove now that condition C7 is satisfied; Let

$$L(x) = \ln\left[\frac{|x|^{1+\alpha}}{\zeta_{\mathrm{l}}}\right] = (1+\alpha)\ln|x| - \ln\zeta_{\mathrm{l}},$$

which implies that for large-enough $|x|$ and $|y|$,

$$\begin{aligned}
L(x+y) &= (1+\alpha)\ln|x+y| - \ln\zeta_{\mathrm{l}} \\
&\leq (1+\alpha)\ln\left[|x|+|y|\right] - \ln\zeta_{\mathrm{l}} \\
&\leq p(1+\alpha)\left[\ln|x| + \ln|y|\right] - \ln\zeta_{\mathrm{l}} \\
&= p\left[(1+\alpha)\ln|x| - \ln\zeta_{\mathrm{l}} + (1+\alpha)\ln|y| - \ln\zeta_{\mathrm{l}}\right] + (2p-1)\ln\zeta_{\mathrm{l}} \\
&\leq 2p\left[(1+\alpha)\ln|x| - \ln\zeta_{\mathrm{l}} + (1+\alpha)\ln|y| - \ln\zeta_{\mathrm{l}}\right] \\
&= 2p\left[L(x) + L(y)\right],
\end{aligned}$$

where $p > 1$. Consequently, the following holds:

Theorem 12 (Capacity Results: Impulsive Noise)**.** *Whenever* $\mathcal{C}(|x|) = \omega(\ln|f(x)|)$, *the support of the capacity-achieving input of channel (3.3) when N is polynomially-tailed is compact.*

For example, for a linear channel subjected to an additive polynomially-tailed noise, the optimal input has a bounded support for any cost function that is super logarithmic (i.e., $\omega(\ln|x|)$) such as the average power constraint. When it comes to discreteness and strengthened-C8, it depends on the analyticity property of the specific $p_N(\cdot)$ under consideration.

The remaining part of this Section is dedicated to two important types of polynomially decaying distributions, for which we prove that the discreteness results of Theorem 10 apply.

Non-Totally Skewed Alpha-Stable and their Mixtures

Checking the conditions: For non-totally skewed laws, both the right and the left tails are polynomially decaying as $\Theta\left(\frac{1}{|x|^{\alpha+1}}\right)$ (see [71, Th.1.12, p.14]), and Theorem 12 holds. Furthermore, whenever $\alpha \geq 1$ the alpha-stable variables are analytically extendable, to the whole complex plane when $\alpha > 1$ and to some horizontal strip when $\alpha = 1$ [75, theorem 2.3.1 p. 48 and remark 1 p. 49]. We check in what follows the strengthened-C8. We derive in Appendix D a novel bound on the rate of decay of the complex extension of the alpha-stable PDF when $\alpha \geq 1$: For small-enough $\eta > 0$, there exist $\kappa > 0$ and $n_0 > 0$ such that

$$|p_N(z)| \leq \frac{\kappa}{|\Re(z)|^{\alpha+1}}, \quad \forall\, z \in \mathcal{S}_\eta : |\Re(z)| \geq n_0. \tag{3.26}$$

This bound insures the validity of Theorem 10 whenever the conditions on $\mathcal{C}(\cdot)$ and $f(\cdot)$ are satisfied, and hence the following theorem is valid:

Theorem 13 (Capacity Results: Alpha-Stable Noise)**.** *Whenever* $\mathcal{C}\left(|x|\right) = \omega\left(\ln|f(x)|\right)$, *the support of the capacity-achieving input of channel (3.3) when* $N \sim \mathcal{S}(\alpha, \beta, \gamma, \delta)$ *is a non-totally skewed alpha-stable variable is compact.*

Whenever $\alpha \geq 1$, *the optimal input is discrete with finite number of mass points whenever* $\mathcal{C}(\cdot)$ *and* $f(\cdot)$ *satisfy the analyticity and connectedness conditions of Theorem 10.*

Note that by virtue of the fact that condition C7 defines a convex set, the results presented here for one alpha-stable variable are valid for any convex combinations of them.

The capacity results of Theorem 13 when applied to the additive linear channel where the noise is modeled as a non-totally skewed alpha-stable variable says that, under any "super-logarithmic" average cost constraint, the capacity achieving input is of bounded support. Furthermore, the optimal input has discrete statistics whenever the alpha-stable noise has $\alpha \geq 1$. This result covers, among other "super-logarithmic" input cost functions, the fractional r-th moment constraint, $\mathrm{E}\left[|X|^r\right] \leq A$, $A > 0$ and $r > 10$.

We use a specialized numerical *Matlab* package to search for the positions of the optimal points and their respective probabilities whenever the optimal input is discrete. In Figure 3-1, we plot the capacity of channel (3.3) whenever $f(x) = x$, $\mathcal{C}(|x|) = x^2$ and $N \sim \mathcal{S}(\alpha, 0, 1, 0)$ for $\alpha = 1, 1.2, 1.5$ and 1.8. The capacity curves clearly shows that as α gets bigger the capacity is higher. This is in accordance with the fact that the lower the value of α, the distribution becomes heavier.

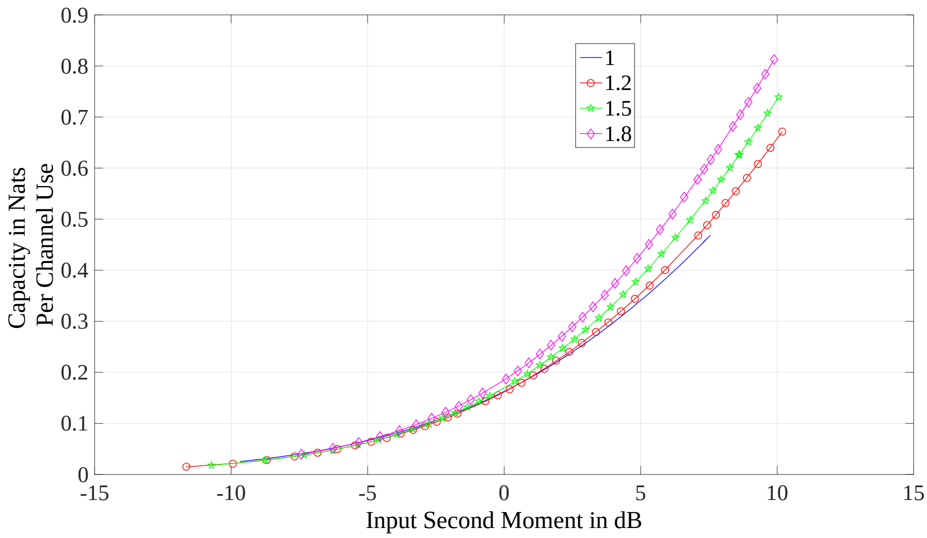

Figure 3-1: Capacity of the linear channel subject to symmetric "standard" alpha-stable noise $N \sim \mathcal{S}(\alpha, 0, 1, 0)$ for various values of the characteristic exponent α.

Composite noise: Gaussian + Alpha-Stable (Middleton Class B)

Recently, a compound noise model was adopted to capture potentially different sources of noise: a Gaussian model for the thermal noise and an alpha-stable model for the potential Multiple Access Interference (MAI), as is the case for ad-hoc self configuring networks with applications in CDMA networks [37], and in the general context of ultra wideband technologies [27]. Further information on the subject can be found in [4,26,28]. We note that this noise model is commonly referred to as the Middleton class B distribution [51, 52]. We consider hence the following additive noise $N = N_1 + N_2$, where

- $N_1 \sim \mathcal{S}(\alpha, \beta, \gamma, \delta)$, which represents the effect of the MAI, assumed a non totally-skewed alpha-stable RV.
- $N_2 \sim \mathcal{N}(\mu, \sigma^2)$ is a Gaussian RV that models the effect of thermal noise.

Checking the conditions: It has been proved in [88, Appendix I] that $p_N(x)$ is polynomially-tailed which implies that Theorem 12 holds for the compound noise model. In order to apply Theorem 10 for the channels impaired by the composite noise N, we use the fact that its PDF is analytically extendable on $\mathbb{C}$ (for all values of $0 < \alpha < 2$) and therefore on $\mathcal{S}_\eta$ [88, Appendix I], and check the strengthened-C8:

$$\begin{aligned} T_{\mathrm{u}}(x) = \sup_{|\Im(z)|<\eta} |p_N(z)| &\leq \sup_{|\Im(z)|<\eta} \frac{1}{\sqrt{2\pi\sigma^2}} \int \left| e^{-\frac{(z-t)^2}{2\sigma^2}} \right| p_{N_1}(t)\, dt \\ &\leq \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{\eta^2}{2\sigma^2}} \int e^{-\frac{(x-t)^2}{2\sigma^2}} p_{N_1}(t)\, dt = e^{\frac{\eta^2}{2\sigma^2}} p_N(x), \end{aligned}$$

which implies

$$-\int_{-\infty}^{+\infty} T_{\mathrm{u}}(x) \ln T_{\mathrm{l}}(x)\; dx \leq -e^{\frac{\eta^2}{\sigma^2}} \int_{-\infty}^{+\infty} p_N(x) \ln p_N(x)\, dx = e^{\frac{\eta^2}{\sigma^2}} h(N) < \infty.$$

The following theorem therefore holds:

Theorem 14 (Capacity Results: Composite Noise)**.** *Whenever $\mathcal{C}(|x|) = \omega(\ln|f(x)|)$, the support of the capacity-achieving input of channel (3.3) when $N = N_1 + N_2$ is compact. The optimal input is discrete with finite number of mass points whenever $\mathcal{C}(\cdot)$ and $f(\cdot)$ satisfy the analyticity and connectedness conditions of Theorem 10.*

Using *Matlab*, we evaluated the capacity of channel (3.3) under a 2^{nd}-moment constraint, when $f(x) = x$ and $N_2 \sim \mathcal{N}(0, 1)$. We considered two scenarios: one where the heavy-tail component is considered "moderate", $N_1 \sim \mathcal{S}(1, 0, 1, 0)$, and another where it is "mild": $N_1 \sim \mathcal{S}(1, 0, 0.1, 0)$.

In Figure 3-2, we plot the channel capacity in the moderate case as well as the achievable rates with a Gaussian input. We observe that, the relative loss in transmission rates is essentially constant (3 to 5%). When the heavy-tailed noise component becomes more accentuated, we expect this loss to be more pronounced. Indeed, in the mild case, the results show that a Gaussian input achieves rates which are very close to capacity. However, we emphasize that this statement does not imply that the channel can be approximated as a Gaussian channel. Indeed, the cumulative noise is heavy-tailed (with polynomial decay) and the capacity values are significantly less than $\frac{1}{2}\log\left(1 + \frac{\mathrm{E}[X^2]}{\sigma^2}\right)$: Even in the mild case, the capacity is found to be 0.298 nats/channel-use at 0.16 dB compared to 0.356 nats for the Gaussian channel model. In Figure 3-3 we plot the capacity of the linear channel where $N_1 \sim \mathcal{S}(\alpha, 0, 1, 0)$ for the values of $\alpha = 1$ and 1.5. The optimal input at 7.27dB was found to have 16 and 18 mass points respectively.

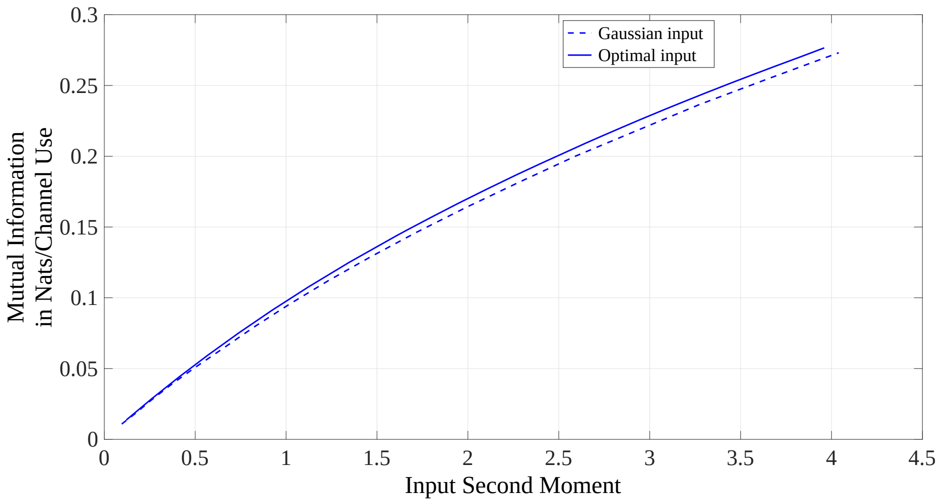

Figure 3-2: Capacity vs achievable rates using a Gaussian input for the composite noise: $N_1 \sim \mathcal{S}(1, 0, 1, 0)$ and $N_2 \sim \mathcal{N}(0, 1)$.

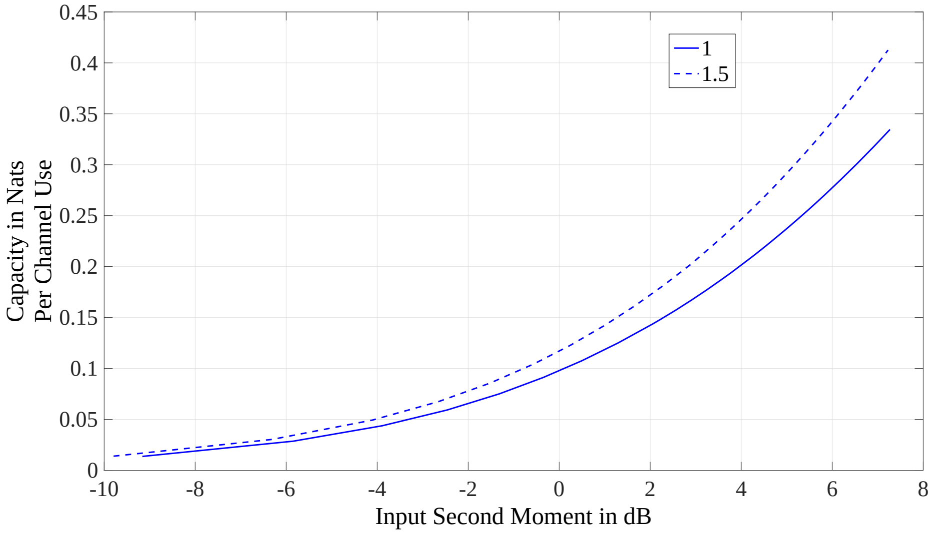

Figure 3-3: Capacity of the linear channel under the composite noise: a standard Gaussian & a standard alpha-stable for $\alpha = 1 \,\&\, 1.5$.

3.7 Related Publications

Finally, we note that the results of this chapter were presented as conference papers in [88, 89] and are currently being reviewed as a journal paper submitted to the IEEE Transactions on Information Theory since June 2015.

Chapter 4

Power Notions

4.1 Background

The second moment is a key parameter in communications theory. Whenever finite, it allows the definition of a Hilbert space of RVs in which it is considered as a power measure. Under this condition, the theory of communications is well established and understood: channel models, power definition, channel capacity, signaling schemes, optimal receivers, etc ... Specifically, considering the second moment as a measure of average power for the various signals was translated to a second moment constraint on the channel input in many information theoretic problems within the context of determining the channel capacity [54, 55, 57–64, 67, 68]. A generalization of the power notion and input constraints was made in [49] where even moment constraints were treated.

However, when it comes to infinite variance RVs the theory is far from being exhaustive. In addition to the theoretical interest, one is faced with practical scenarios in which signals with an infinite second moment should be dealt with. This is the case of additive channels that are subjected to impulsive noise, specifically the alpha-stable family. For these types of channels, the second moment is no longer considered to be a suitable power measure as all related quantities are infinite. This comes with added complexity of "loosing" the Hilbert space structure that a finite second moment guarantees. A survey of the literature shows that alternative measure of power were proposed:

- In [11], Shao and Nikias proposed the "dispersion" of a RV as a measure that plays a similar role to the variance. However, since no analytical expression is defined for the dispersion except for alpha-stable distributions, they propose the usage of the r-th absolute moment ($r < 2$) as an alternative which yields a non-linear signal processing theory.

  Additionally, the signal-to-dispersion ratio –defined as the ratio of the second moment of the input to the dispersion of the noise– is used as an alternative to the regular signal-to-noise ratio when the noise is alpha-stable. Using this ratio as a measure of the channel strength, some achievable rates were numerically evaluated for the alpha-stable channel [47].

  Being said, it would seem plausible to adopt a FOM $\mathsf{E}\left[|X|^r\right]$ for some $r > 0$ in the presence of alpha-stable distributions since they have have finite FLOM for $r < \alpha$ (See Section 2.2.1 Property 1).

- Based on logarithmic moments of the form $\mathsf{E}\left[\log|X|\right]$, an alternative notion of power was introduced by Gonzales [39] for the heavy tailed distributions which he called the Geometric Power (GP):

  $$\mathcal{S}_0(X) \hat{=} e^{\mathsf{E}[\log|X|]} \tag{4.1}$$

  He considered the logarithmic moments as a "universal framework" for dealing with algebraic tail processes that will overcome the shortcomings of the FLOM approach which he summarizes as follows:

  - Since r is usually restricted to the interval $(0;\alpha)$, constructing a valid FLOM requires the previous knowledge (or estimation) of α in order to choose an appropriate value of r.

– On the other hand, for any given $r > 0$, there will always be a "remaining" class of very impulsive processes (those with $\alpha \leq r$) for which the associated FLOM is not appropriate.

Also the discontinuity in the FLOM is yet another unpleasant feature. In fact, for a given $0 < r < 2$, two alpha stable distributions with characteristic exponents $\alpha_1 = r + \epsilon$ and $\alpha_2 = r - \epsilon$ (for some $\epsilon > 0$), will respectively have a finite and infinite r-th absolute moment though they would have similar statistical behavior. However, all stable distributions have a finite logarithmic moment [39].

The GP was used in formulating new impulsive signal processing techniques, nevertheless adopting it as power definition from an information theoretic perspective is faced with multiple complications. In fact, constraining the GP for the input signal is really a constraint on the logarithmic moment $\mathsf{E}[\log|X|]$ which can assume negative values and this will be in contradiction with the usual notion of positive average input cost/power. More importantly, for any discrete input X that has a mass point at zero, $\mathcal{S}_0(X)$ will be necessarily null even if other non-zero mass points are existent. This would yield a zero power for a non-zero signal. However the logarithmic nature of the constraint proposed by Gonzales seems to be a characteristic of the "right" form which is applicable to all heavy tailed distributions.

In this chapter, we further support the proposition that power measures with logarithmic tail behaviour are suitable in the presence of alpha-stable distributions however we do not restrict ourselves to definition (4.1) or equivalently to $\mathsf{E}[\log|X|]$. Instead, we work in the restriction that a strength measure P_X of a variable X satisfies the following:

R1- $\mathrm{P}_X \geq 0$, w.e. iff $X = 0$.

R2- $\mathrm{P}_{aX} = |a|\mathrm{P}_X$ for $a \in \mathbb{R}$

R3- P_X is a parameter that captures the width of the distribution (similar to the variance for finite variance variables or the dispersion for the alpha-stable variables...)

Though these restrictions do not contain some of the dispensable properties satisfied by the GP such as the multiplicativity and the triangular inequality properties [39], they are deemed sufficient to define a strength measure. We proceed next to argue the suitability of measuring the power through the average of logarithmically-tailed functions and later we propose a couple of newly defined power measures and we argue the reason why there are deemed suitable.

4.2 Suitable Power Measures: Connections to the Capacity Results of the Alpha-Stable Channel

Over continuous-alphabets channels, a common belief is that with "sufficient" power, one is capable of transmitting at arbitrarily large rates. Stated differently, if an input of infinite power is allowed, the channel capacity is infinite. This belief is perhaps inspired from the well-known AWGN and linear Gaussian channels for example. The results corroborated in Appendix B disproves this belief if one relates the notion of power to the second moment.

Let us recall the results provided in Chapter 3: they state that –for monotonically noise PDFs, there exists a threshold growth rate for the cost function which constitutes the transition between bounded and unbounded optimal inputs. Indeed, for an optimal input to be unbounded, a "necessary condition" for the cost function is to be at most $\Theta\left(\ln\left[\frac{1}{p_N(f(x))}\right]\right)$.

In a general setting, we next argue that this result provides insights on what is a suitable measure of signal strength. Though this question is not crucial when the additive noise has a finite second moment due to the natural power measure provided by the second moment, it seems of great importance when dealing with heavy tailed noise distributions having infinite second moments. We use the results of Chapter 3 and make the following reasoning:

- Let $\mathsf{E}\left[\mathcal{C}_0(|x|)\right]$ be a measure of the average signal strength where the strength function $\mathcal{C}_0(|x|)$ is positive, lower semi-continuous, non-decreasing in $|x|$ and let $p_N(x)$ be the noise PDF which is assumed to have a monotonic tail.

- Whenever $\mathcal{C}_0(|x|) = \omega\left(\ln\left[\frac{1}{p_N(f(x))}\right]\right)$, one can always find a cost function $\mathcal{C}(|x|)$ such that $\mathcal{C}(|x|)$ is both $\omega\left(\ln\left[\frac{1}{p_N(f(x))}\right]\right)$ and $o\left(\mathcal{C}_0(|x|)\right)$.

- Now, since $\mathcal{C}(|x|) = \omega\left(\ln\left[\frac{1}{p_N(f(x))}\right]\right)$, the channel capacity under an input constraint of the form $\mathsf{E}\left[\mathcal{C}(|x|)\right] \leq A$, $A > 0$ is achieved by a bounded input by virtue of Theorem 8. On the other hand, since $\mathcal{C}(|x|) = o\left(\mathcal{C}_0(|x|)\right)$, then there exists a distribution function satisfying the cost constraint with "signal strength" $\mathrm{E}\left[\mathcal{C}_0(|x|)\right]$ equal to ∞.

- Hence, in the input space of distribution functions, there exist distributions having possibly infinite strength while the capacity is achieved by a distribution which has a finite one since its support is bounded.

- This non-intuitive conclusion is only possible under the choice of a strength measure that is the average of a function of the form $\omega\left(\ln\left[\frac{1}{p_N(f(x))}\right]\right)$.

Following this line of thought, suitable signal strength functions should be at most $\Theta\left(\ln\left[\frac{1}{p_N(f(x))}\right]\right)$. Said differently, depending on the noise, functions that are of the form $\Theta\left(\ln\left[\frac{1}{p_N(f(x))}\right]\right)$ are more appropriate. This boils down to $\Theta(f^2(x))$ under the Gaussian noise and to $\Theta\left(\ln\left[f(x)\right]\right)$ for polynomially-tailed additive noise. The latter condition holds for the alpha-stable family being polynomially tailed. As an example, a suitable strength function for the linear additive alpha-stable channel is one of the form $\Theta\left(\ln x\right)$.

4.3 A Power Measure in the Fourier Domain

An important approach for searching for new power measures of signals with infinite second moments is to observe the Fourier domain. More specifically, the second moment of a RV X is under some mild conditions the second derivative of the characteristic function $\phi(\cdot)$ evaluated at 0. Hence, observing the behaviour of characteristic functions around the origin might give indicators on the average power of RVs. Though, in general, it is hard to always relate a local behaviour of the characteristic function to some strength function of the variable X, the Fourier domain approach seems appealing in the case of alpha-stable distributions and their domains of attractions since these variables have closed-form characteristic functions which is not the case for their density functions. We only consider in this section symmetric RVs:

Definition 4 (Power Measure in the Fourier Domain)**.** Let X be a real-valued RV and having a symmetric distribution function around the origin. Then, define

$$\mathrm{P}_{\tau}(X) = -\lim_{\omega \to 0} \frac{d}{d\omega}\left[|\omega|^{2-\tau}\frac{d\phi}{d\omega}(\omega)\right] \tag{4.2}$$

for some $0 < \tau \leq 2$ assuming that the limit is finite and that $\frac{d\phi}{d\omega}(\omega)$ exists in a neighbourhood of 0, not necessarily at 0.

We note that $\phi(\omega)$ is a real-valued, even function and hence $\frac{d\phi}{d\omega}(\omega)$, $\frac{d^2\phi(\omega)}{d\omega^2}$ are respectively odd and even in their domain of definition. For the limit to exist, we must have:

$$\begin{aligned}
\mathrm{P}_\tau(X) &= -\lim_{\omega\to 0^-}\frac{d}{d\omega}\left[|\omega|^{2-\tau}\frac{d\phi}{d\omega}(\omega)\right] = -\lim_{\omega\to 0^-}\frac{d}{d\omega}\left[(-\omega)^{2-\tau}\frac{d\phi}{d\omega}(\omega)\right]\\
&= -\lim_{\omega\to 0^-}\left[-(2-\tau)(-\omega)^{1-\tau}\frac{d\phi}{d\omega}(\omega) + (-\omega)^{2-\tau}\frac{d^2\phi}{d\omega^2}(\omega)\right]\\
&= -\lim_{\omega\to 0^+}\left[(2-\tau)\omega^{1-\tau}\frac{d\phi}{d\omega}(\omega) + \omega^{2-\tau}\frac{d^2\phi}{d\omega^2}(\omega)\right] \qquad (4.3)\\
&= -\lim_{\omega\to 0^+}\frac{d}{d\omega}\left[|\omega|^{2-\tau}\frac{d\phi}{d\omega}(\omega)\right]. \qquad (4.4)
\end{aligned}$$

We note that the equality in (4.3) is always satisfied as long as the limit exists and is finite since $\phi(\omega)$ is even. For this to be true, the following conditions are deemed sufficient for the existence of $\mathrm{P}_\tau(X)$:

a- $\phi(\omega)$ is $\mathcal{C}^2$ in a neighbourhood of 0, not necessarily at 0.

b- There exists a $0 < \tau \leq 2$ such that:

$$\lim_{\omega\to 0^+}\omega^{1-\tau}\frac{d\phi}{d\omega}(\omega) \quad \text{and} \quad \lim_{\omega\to 0^+}\omega^{2-\tau}\frac{d^2\phi}{d\omega^2}(\omega)$$

are both finite.

In addition, we assume

c- Whenever $\phi(\omega)$ is convex in a neighbourhood of 0, it should be such that $\frac{d^2\phi}{d\omega^2}(\omega) \leq |\omega|^{\epsilon-1}$ for some $\epsilon > 0$.

We note that condition c is rather a mild condition for technical considerations as it will appear later and is not related to the existence of P_τ.

4.3.1 A Parametrized Power Definition

Let X and Y be two independent RVs satisfying a, b and c. In this section, we list some of the important properties that are satisfied by $\mathrm{P}_\tau(\cdot)$ and are considered to some extent necessary in order to qualify it as power operator. We consider only the limit for $\omega > 0$.

Positiveness

We make the following remarks before proceeding into the proof that $\mathrm{P}_\tau(X) \geq 0$:

- $\phi(\omega)$ is a characteristic function, hence $\phi(\omega)$ is uniformly continuous and $\phi(\omega) < \phi(0)$ for all ω. Then there exists an open positive neighbourhood of 0 in which $\phi(\omega)$ is decreasing and $\frac{d\phi}{d\omega}(\omega) \leq 0$.
- As for $\lim_{\omega\to 0^+} \frac{d^2\phi}{d\omega^2}(\omega)$, it could assume both positive or negative values, and hence there exists an open positive neighbourhood of 0 for which $\phi(\omega)$ could be either concave or convex.

The proof is based on the above observations and is done separately for two different categories of $\phi(\omega)$:

$\phi(\omega)$ is concave in some open positive neighbourhood of 0 In this case, both $\frac{d\phi}{d\omega}(\omega)$ and $\frac{d^2\phi}{d\omega^2}(\omega)$ are non-positive when $\omega \to 0^+$. Equation (4.3) implies that $\mathrm{P}_\tau(X)$ is non-negative.

$\phi(\omega)$ is convex in some open positive neighbourhood of 0 This implies that necessarily $\lim_{\omega\to 0^+} \frac{d\phi}{d\omega}(\omega) < 0$. Now let $\epsilon > 0$ be such that condition c is satisfied, and consider the following function defined for $\omega \geq 0$:

$$G(\omega) = -\frac{1}{2-\tau}\omega^\epsilon$$

We note that $\lim_{\omega\to 0^+} \frac{d\phi}{d\omega}(\omega) < \lim_{\omega\to 0^+} G(\omega) = 0$. By continuity, there exists some open positive neighbourhood of 0 in which $0 < \frac{d\phi}{d\omega}(\omega) < G(\omega)$. Using property c, we get the following set of inequalities:

$$\frac{d^2\phi}{d\omega^2}(\omega) \leq \omega^{\epsilon-1} \leq -\frac{2-\tau}{\omega}\frac{d\phi}{d\omega}(\omega)$$

for ω in some open positive neighbourhood of 0. Multiplying by $\omega^{2-\tau}$ yields that $\mathrm{P}_\tau(X)$ is non-negative.

Scalability

Let κ be a real, non-zero value. Define $U = \kappa X$, we have:

$$\begin{aligned}
&\mathrm{P}_\tau(U)\\
&= -\lim_{\omega\to 0^+}\left[(2-\tau)\omega^{1-\tau}\frac{d\phi_U}{d\omega}(\omega) + \omega^{2-\tau}\frac{d^2\phi_U}{d\omega^2}(\omega)\right]\\
&= -\lim_{\omega\to 0^+}\left[\kappa(2-\tau)\omega^{1-\tau}\frac{d\phi_X}{d\omega}(\kappa\omega) + \kappa^2\omega^{2-\tau}\frac{d^2\phi_X}{d\omega^2}(\kappa\omega)\right]\\
&= -\lim_{|\kappa|\omega\to 0^+}\left[\mathrm{sgn}(\kappa)|\kappa|^\tau(2-\tau)\left(|\kappa|\omega\right)^{1-\tau}\frac{d\phi_X}{d\omega}(\kappa\omega) + |\kappa|^\tau\left(|\kappa|\omega\right)^{2-\tau}\frac{d^2\phi_X}{d\omega^2}(\kappa\omega)\right]\\
&= -|\kappa|^\tau\lim_{|\kappa|\omega\to 0^+}\left[(2-\tau)\left(|\kappa|\omega\right)^{1-\tau}\frac{d\phi_X}{d\omega}(|\kappa|\omega) + \left(|\kappa|\omega\right)^{2-\tau}\frac{d^2\phi_X}{d\omega^2}(|\kappa|\omega)\right] \qquad (4.5)\\
&= |\kappa|^\tau\mathrm{P}_\tau(X).
\end{aligned}$$

Equation (4.5) is due to the fact that $\frac{d\phi_X}{d\omega}(\omega)$ and $\frac{d^2\phi_X}{d\omega^2}(\omega)$ are respectively odd and even functions for all $\omega \neq 0$.

Additive Property

Let $Z = Y + X$. We have,

$$
\begin{aligned}
&\mathrm{P}_\tau(Z)\\
&= -\lim_{\omega\to 0^+}\left[(2-\tau)\omega^{1-\tau}\frac{d\phi_Z}{d\omega}(\omega) + \omega^{2-\tau}\frac{d^2\phi_Z}{d\omega^2}(\omega)\right]\\
&= -\lim_{\omega\to 0^+}\left[(2-\tau)\omega^{1-\tau}\frac{d\left(\phi_X(\omega)\phi_Y(\omega)\right)}{d\omega} + \omega^{2-\tau}\frac{d^2\left(\phi_X(\omega)\phi_Y(\omega)\right)}{d\omega^2}\right]\\
&= -\lim_{\omega\to 0^+}\left[(2-\tau)\omega^{1-\tau}\phi_X(\omega)\frac{d\phi_Y}{d\omega}(\omega) + (2-\tau)\omega^{1-\tau}\phi_Y(\omega)\frac{d\phi_X}{d\omega}(\omega)\right.\\
&\qquad\left. +\omega^{2-\tau}\phi_X(\omega)\frac{d^2\phi_Y}{d\omega^2}(\omega) + \omega^{2-\tau}\phi_Y(\omega)\frac{d^2\phi_X}{d\omega^2}(\omega) + 2\,\omega^{2-\tau}\frac{d\phi_X}{d\omega}(\omega)\frac{d\phi_Y}{d\omega}(\omega)\right]\\
&= -\lim_{\omega\to 0^+}\phi_X(0)\left[(2-\tau)\omega^{1-\tau}\frac{d\phi_Y}{d\omega}(\omega) + \omega^{2-\tau}\frac{d^2\phi_Y}{d\omega^2}(\omega)\right]\\
&\quad -\lim_{\omega\to 0^+}\phi_Y(0)\left[(2-\tau)\omega^{1-\tau}\frac{d\phi_X}{d\omega}(\omega) + \omega^{2-\tau}\frac{d^2\phi_X}{d\omega^2}(\omega)\right] - \lim_{\omega\to 0^+} 2\,\omega^{2-\tau}\frac{d\phi_X}{d\omega}(\omega)\frac{d\phi_Y}{d\omega}(\omega)\\
&= -\lim_{\omega\to 0^+}\left[(2-\tau)\omega^{1-\tau}\frac{d\phi_Y}{d\omega}(\omega) + \omega^{2-\tau}\frac{d^2\phi_Y}{d\omega^2}(\omega)\right]\\
&-\lim_{\omega\to 0^+} 2\,\omega^{\tau}\omega^{1-\tau}\frac{d\phi_X}{d\omega}(\omega)\omega^{1-\tau}\frac{d\phi_Y}{d\omega}(\omega) - \lim_{\omega\to 0^+}\left[(2-\tau)\omega^{1-\tau}\frac{d\phi_X}{d\omega}(\omega) + \omega^{2-\tau}\frac{d^2\phi_X}{d\omega^2}(\omega)\right]\\
&= \mathrm{P}_\tau(Y) + \mathrm{P}_\tau(X). \qquad (4.6)
\end{aligned}
$$

In order to write the last equation, we used the fact that

$$\lim_{\omega\to 0^+} 2\,\omega^{\tau}\omega^{1-\tau}\frac{d\phi_X}{d\omega}(\omega)\omega^{1-\tau}\frac{d\phi_Y}{d\omega}(\omega) = 0,$$

by virtue of the fact that $\lim_{\omega\to 0^+}\omega^{1-\tau}\frac{d\phi_X}{d\omega}(\omega)$ and $\lim_{\omega\to 0^+}\omega^{1-\tau}\frac{d\phi_Y}{d\omega}(\omega)$ are both finite and $\tau > 0$.

4.3.2 An Universal Power Definition

Considering equation (4.3), the following behaviour of $\mathrm{P}_\tau(X)$ can be deduced:

- When there exists a $0 < \tau \leq 2$ for which $\mathrm{P}_\tau(X)$ is finite non-zero:

- if $\tau \neq 2$, then $\mathrm{P}_t(X)$ assumes the value $+\infty$ for $\tau < t \leq 2$. On the other hand for $0 < t < \tau$, $\mathrm{P}_t(X) = 0$.
- if $\tau = 2$, then $\mathrm{P}_t(X) = 0$ for $0 < t \leq 2$.

- If no such τ exists, then there exits $0 < \hat{\tau} \leq 2$ such that $\mathrm{P}_t(X) = 0$ for $t \in (0, \hat{\tau}]$ and is infinite otherwise.

Due to these observations, we define the power as follows:

$$\mathrm{P}(X) = \mathrm{P}_{\tau_{\max}}(X), \tag{4.7}$$

where

$$\tau_{\max} = \mathrm{argmax}_{0<\tau\leq 2}\{\mathrm{P}_\tau < \infty\}.$$

In the remaining of this section, we check the properties of $\mathrm{P}(\cdot)$ that are inherited from $\mathrm{P}_\tau(\cdot)$. In fact let X, Y denote two independent RVs for which there exists respectively $0 < \tau_1 \leq 2$ and $0 < \tau_2 \leq 2$ such that $\mathrm{P}_{\tau_1}(X)$ and $\mathrm{P}_{\tau_2(Y)}$ are defined:

(i) Existence: Whenever there exists a $0 < \tau \leq 2$ for which $\mathrm{P}_\tau(X)$ is finite such $\tau_{\max}$ always exists and hence $\mathrm{P}(X)$ exists.

(ii) Positiveness: This a direct consequence of the positiveness of P_τ.

(iii) Scalability: By virtue of the remarks in the beginning of this section, $\mathrm{P}(X) = \mathrm{P}_{\tau_{\max}}(X)$ and hence $\mathrm{P}(\kappa X) = \kappa^{\tau_{\max}}\mathrm{P}(X)$.

(iv) Triangular Inequality: We assume WLOG that $\mathrm{P}(X) = \mathrm{P}_{\tau_1}(X)$ and $\mathrm{P}(Y) = \mathrm{P}_{\tau_2}(Y)$ and $\tau_1 \leq \tau_2$. Then, $\mathrm{P}_{\tau_1}(X+Y) = \mathrm{P}_{\tau_1}(X) + \mathrm{P}_{\tau_1}(Y)$ exists and is finite ($0 \leq \mathrm{P}_{\tau_1}(Y) \leq \mathrm{P}_{\tau_2}(Y)$ exists since $\tau_1 \leq \tau_2$). Furthermore it can be easily seen that P_τ does not exist if $\tau > \tau_1$. Hence

$$\mathrm{P}(X+Y) = \mathrm{P}_{\tau_1}(X+Y) = \mathrm{P}_{\tau_1}(X)+\mathrm{P}_{\tau_1}(Y) = \mathrm{P}(X)+\mathrm{P}_{\tau_1}(Y) \leq \mathrm{P}(X)+\mathrm{P}(Y),$$

with equality iff $\mathrm{P}(Y) = \mathrm{P}_{\tau_2}(Y) = \mathrm{P}_{\tau_1}(Y)$. This is feasible iff $\mathrm{P}(Y) = 0$ and/or $\tau_1 = \tau_2$.

Shortcomings: One final comment is that though it is guaranteed using (4.7) that whenever $X = 0$, $\mathrm{P}(X) = \mathrm{P}_2(X) = 0$, the contrary is not clear to be true. Namely, if we consider X such that $\mathrm{P}(X) = 0$, this would not necessarily imply that $X = 0$. This point will be addressed at a later stage to guarantee the preceding proposition. For the remaining of this section, we refer to $\tau_{\max}$ as the tuned parameter.

4.3.3 Evaluation of P_τ for the Stable Family and some Observations

1. When $\tau = 2$,

$$\mathrm{P}_2(F) = -\frac{d^2\phi}{d\omega^2}(\omega) = \int x^2 dF(x)$$

which is the standard power notion adopted for finite second moment RVs.

2. For symmetric stable distributions with exponent $0 < \alpha \leq 2$ with dispersion $\gamma > 0$, evaluating P_τ yields:

$$\begin{aligned}\mathrm{P}_\tau(X) &= -\lim_{\omega \to 0} \frac{d}{d\omega}\left[|\omega|^{2-\tau}\frac{d\phi}{d\omega}(\omega)\right] \\ &= -\lim_{\omega \to 0} \frac{d}{d\omega}\left[|\omega|^{2-\tau}\frac{de^{-\gamma^\alpha|\omega|^\alpha}}{d\omega}\right]. \end{aligned} \tag{4.8}$$

Taking the limit as $\omega \to 0^+$, we obtain

$$\begin{aligned}
&\lim_{\omega\to 0^+} \frac{d}{d\omega}\left[|\omega|^{2-\tau}\frac{de^{-\gamma^\alpha|\omega|^\alpha}}{d\omega}\right] \\
&= \lim_{\omega\to 0^+} \frac{d}{d\omega}\left[\omega^{2-\tau}\frac{de^{-\gamma^\alpha\omega^\alpha}}{d\omega}\right] \\
&= -\gamma^\alpha\alpha \lim_{\omega\to 0^+} \frac{d}{d\omega}\left[\omega^{1+\alpha-\tau}e^{-\gamma\omega^\alpha}\right] \\
&= -\gamma^\alpha\alpha \lim_{\omega\to 0^+} \omega^{\alpha-\tau}e^{-\gamma^\alpha\omega^\alpha}\left[(1+\alpha-\tau)-\gamma^\alpha\alpha\omega^\alpha\right].
\end{aligned}$$

Considering the above limits, 3 regimes are observed:

- If $\tau = \alpha$, then

$$\lim_{\omega\to 0^+} \frac{d}{d\omega}\left[|\omega|^{2-\tau}\frac{de^{-\gamma^\alpha|\omega|^\alpha}}{d\omega}\right] = \lim_{\omega\to 0^-} \frac{d}{d\omega}\left[|\omega|^{2-\tau}\frac{de^{-\gamma^\alpha|\omega|^\alpha}}{d\omega}\right] = -\alpha\gamma^\alpha,$$

  and

$$\mathrm{P}_\alpha\left(S(\alpha,\gamma)\right) = \alpha\gamma^\alpha$$

  For $\alpha = 2$, the power of a zero mean Gaussian with variance σ^2 is:

$$\mathrm{P}_2\left(\mathcal{N}(0,\sigma^2)\right) = 2\,\gamma^2 = 2\times\frac{\sigma^2}{2} = \sigma^2$$

- If $\tau < \alpha$, the two limits are both equal to 0. Hence, in this case:

$$\mathrm{P}_\tau\left(S(\alpha,\gamma)\right) = 0.$$

- If $\tau > \alpha$, both limits diverges to $-\infty$ and therefore $\mathrm{P}_\tau\left(S(\alpha,\gamma)\right)$ is $+\infty$.

As an interpretation of the above results, one could make the following observations:

- The parameter $0 < \tau \leq 2$ can be regarded as a "tune" parameter in order to capture the power of a RV. Within the class of stable distributions, this tune parameter is closely related to the characteristic exponent α. In fact, $\tau_{\max} = \alpha$ whenever $X \sim S(\alpha, \gamma)$ for $0 < \alpha \leq 2$ and $\gamma > 0$. As long as $\tau < \alpha$, P_τ fails to capture the power of $S(\alpha, \gamma)$ since it evaluates to 0 and τ is undertuned. At $\tau = \alpha$, P_τ is tuned and the power of $S(\alpha, \gamma)$ is proportional to the dispersion. Finally when $\tau > \alpha$, the power is infinite and τ is overtuned.

- For the Gaussian RV, there is no overtuned range for τ and the tuned value is $\tau_{\max} = 2$ which gives the standard notion of power as being the second moment. For all the remaining values, τ is undertuned. On the other hand the value $\tau = 2$ is considered to be overtuned for all the alpha-stable RVs. This is the basic fact that all the alpha-stable RVs have an infinite second moment.

- What if we want to generalize the mentioned behavior of the tuning parameter for a bigger class of RVs. Said differently, can one find a set of distribution functions for which there exists a value of $\tau_{\max}$?. Before we proceed, we make the following comments

  – First, when considering the space of symmetric RVs having a finite second moment, the conditions a, b and c are satisfied for any $0 < \tau \leq 2$ as long as $\phi(\omega)$ is $\mathcal{C}^2$ in a neighbourhood of 0. This is due to the fact that $\frac{d\phi}{d\omega}(\omega)|_{\omega=0} = 0$ and $\frac{d^2\phi(\omega)}{d\omega^2}|_{\omega=0} < \infty$. By virtue of the fact that P_τ is increasing in $0 < \tau \leq 2$, then $\tau_{\max} = 2$.

  – Second, when it comes to the infinite second moment space of distribution functions, since the tuned value of τ is equal to α for an alpha-stable RV, can it be generalized for a bigger space of RVs containing the alpha-stable ones that is parametrized by α? This suggests the space of density functions that belong to the domain of attraction of the stable law. This is discussed in the following section.

4.3.4 A Power Operator for Elements in $\mathbb{D}_{\alpha,0} = \mathbb{D}_\alpha$

We only consider in this section distribution functions that are symmetric with respect to 0, more precisely $F(x)$ such that $F(-x) = 1 - F(x) + \Pr(X = x)$. Let $F(x) \in \mathbb{D}_\alpha$ be a distribution function within the domain of normal attraction (refer to Section 2.2.2 for the definition) of the symmetric alpha-stable type $S(\alpha, \gamma)$ of characteristic exponent $0 < \alpha \leq 2$ and γ is any positive constant.

We assume that $F(x)$ satisfies conditions a and c of Section 4.3 and we wish to check the feasibility of finding a value of $0 < \tau \leq 2$ (and eventually $\tau_{\max}$) for which condition b is satisfied and P_τ exists for all $F \in \mathbb{D}_\alpha$. The value of $\tau_{\max}$ is the tuned value which provides a new power measure that captures the power of these heavy tailed densities. We propose evaluating $\mathrm{P}_\tau(X)$, where $X \sim F(x)$, for $\tau = \alpha$. Recall that $\phi_X(\omega)$ is $\mathcal{C}^2$ in a neighbourhood of 0 and satisfies equation (2.12) with $\beta = 0$ and $\delta = 0$. We have:

$$
\begin{aligned}
&\mathrm{P}_\alpha(X) \\
&= - \lim_{\omega \to 0^+} \frac{d}{d\omega} \left[|\omega|^{2-\alpha} \frac{d\phi_X(\omega)}{d\omega} \right] = \lim_{\omega \to 0^+} \frac{d}{d\omega} \left[|\omega|^{2-\alpha} \frac{d e^{-\gamma^\alpha |\omega|^\alpha (1+o(1))}}{d\omega} \right] \qquad (4.9) \\
&= - \lim_{\omega \to 0^+} \frac{d}{d\omega} \left[\omega^{2-\alpha} \frac{d e^{-\gamma^\alpha \omega^\alpha (1+o(1))}}{d\omega} \right] \\
&= \gamma^\alpha \lim_{\omega \to 0^+} \frac{d}{d\omega} \left[\left(\alpha\omega + \left(o\left(1\right)\omega^\alpha\right)' \omega^{2-\alpha} \right) e^{-\gamma^\alpha \omega^\alpha (1+o(1))} \right] \\
&= \gamma^\alpha \lim_{\omega \to 0^+} \Big[\alpha + \left(o\left(1\right)\omega^\alpha\right)'' \omega^{2-\alpha} + (2-\alpha)\left(o\left(1\right)\omega^\alpha\right)' \omega^{1-\alpha} + \\
&\qquad\qquad \alpha\omega + \left(o\left(1\right)\omega^\alpha\right)' \omega^{2-\alpha} \Big] e^{-\gamma^\alpha \omega^\alpha (1+o(1))} \\
&= \alpha\gamma^\alpha < \infty, \qquad (4.10)
\end{aligned}
$$

where we used Theorem 7 stated in Chapter 2 in order to write equation (4.9). The fact that $\phi_X(\omega)$ is $\mathcal{C}^2$ in a neighbourhood of 0 implies the existence of the first and the second derivatives of the term $o(1)$ in a neighbourhood of 0. To write equation (4.10) we used the following properties:

Let $f(x)$ be a $\mathcal{C}^1$ function in a neighbourhood of 0, then

- if $f(x) = o(x^l)$, $l > 0$, then $f'(x) = o(x^{l-1})$ by virtue of l'Hospital's rule.
- if $f(x) = o(x^l)$ and $g(x) = o(x^m)$, $l\ m \in \mathbb{R}$, then $f(x)g(x) = o(x^{l+m)}$

Since $\mathrm{P}_\alpha(X)$ is finite and non-zero, then $\tau_{\max} = \alpha$ and $\mathrm{P}(X) = \mathrm{P}_\alpha(X) = \alpha\gamma$, $\forall\ F(x) \in \mathbb{D}_\alpha$.

4.3.5 Discussions and Insights

The result of the previous section implies that the newly defined power operator $\mathrm{P}(X)$ generalizes the well known second moment as being the power operator for the domain of normal attraction of the Gaussian random variable to the bigger space of domains of normal attraction of all stable laws. In this regard, we make the following discussion:

(i) We consider the ensemble of domains of normal attraction $\mathbb{D} = \cup_{0<\alpha\leq 2}\mathbb{D}_\alpha$ of all symmetric stable types (including the Gaussian) $\mathbb{D}_\alpha$, $0 < \alpha \leq 2$. This space is composed of uncountable disjoint sets that are characterized as follows:

  - Domains of normal attraction of alpha-stable type $0 < \alpha < 2$. These are composed of distributions with a polynomial tail behavior according to equation (2.10).
  - The domain of normal attraction of the Gaussian type which is the space of distributions having a finite variance.

(ii) We define the generalized power notion for a RV $X \sim F(x)$ satisfying conditions a, b and c according to equation (4.7) as $\mathrm{P}(X)$. Whenever $F \in \mathbb{D}$, using this universal definition, we obtain:

- Whenever $F \in \mathbb{D}_\alpha$ for some $0 < \alpha < 2$, then $\mathrm{P}(X) = \mathrm{P}_\alpha(X)$ according to Definition 4.
- When $\alpha = 2$, $\mathrm{P}(X) = \mathrm{P}_2(X) = \mathrm{E}\left[X^2\right]$.

(iii) Each $\mathbb{D}_\alpha$, $0 < \alpha < 2$, contains all the stable laws $S(\alpha, \gamma)$ for all $\gamma > 0$. According to our defintion of power and by equation (4.10), each $\mathbb{D}_\alpha$ is composed of disjoint groups which are identified by $S(\alpha, \gamma)$, $\gamma > 0$. Within $\mathbb{D}_\alpha$, the elements in each group have a power value of $\alpha\gamma^\alpha$. From this perspective, one can define the dispersion of a distribution $F(x) \in \mathbb{D}_\alpha$, $0 < \alpha < 2$ as being the power scaled by $\frac{1}{\alpha}$. Therefore each $S(\alpha, \gamma)$ attracts the elements in $\mathbb{D}_\alpha$ having the same power, hence the same dispersion $\gamma^\alpha > 0$. For the Gaussian case and $\mathbb{D}_2$, this boils down to variables having a given variance that are attracted to a Gaussian having the same variance.

(iv) Over the set $\mathbb{D}$, the only distribution that has a zero power is the one that corresponds to $X = 0$. Hence under this setup, one can guarantee that $\mathrm{P}(X) = 0$ iff $X = 0$.

(v) Let X_1, X_2 be two independent RVs in $\mathbb{D}$ with power P_1, P_2 respectively. WLOG, there exists $0 < \alpha_1 \leq \alpha_2 \leq 2$ such that $X_1 \in \mathbb{D}_{\alpha_1}$, $X_2 \in \mathbb{D}_{\alpha_2}$. The dispersion of X_1 and X_2 are respectively $\gamma_1^{\alpha_1} = \frac{\mathrm{P}_1}{\alpha_1}$ and $\gamma_2^{\alpha_2} = \frac{\mathrm{P}_2}{\alpha_2}$. If we denote $\phi_{X_1}(\omega)$, $\phi_{X_2}(\omega)$ the respective characteristic functions of X_1 and X_2, then according to equation (2.12), we have the following in a neighbourhood of 0:

$$\begin{aligned}\log \phi_{X_1}(\omega) &= -\gamma_1^{\alpha_1}|\omega|^{\alpha_1}[1 + o(1)] \\ \log \phi_{X_2}(\omega) &= -\gamma_2^{\alpha_2}|\omega|^{\alpha_2}[1 + o(1)]\end{aligned}$$

Now let, $Z = X_1 + X_2$, then by independence

$$\log \phi_Z(\omega) = -\gamma_1^{\alpha_1}|\omega|^{\alpha_1}[1 + o(1)] - \gamma_2^{\alpha_2}|\omega|^{\alpha_2}[1 + o(1)].$$

We distinguish between two cases:

– if $\alpha_1 < \alpha_2$, then:

$$\log \phi_Z(\omega) = -\gamma_1^{\alpha_1} |\omega|^{\alpha_1} [1 + o(1)]. \tag{4.11}$$

and $Z \in \mathbb{D}_{\alpha_1}$. Then,

$$\mathrm{P}(Z) = \mathrm{P}_{\alpha_1}(Z) = \mathrm{P}_{\alpha_1}(X_1) + \mathrm{P}_{\alpha_1}(X_2) = \mathrm{P}_{\alpha_1}(X_1) = \mathrm{P}(X_1) = \mathrm{P}_1,$$

where the second equality is due to the additive property of $\mathrm{P}_\tau(\cdot)$ and the third equality since $\alpha_1 < \alpha_2$. The same result could be inferred directly from equation (4.11), where it is clear that $Z \in \mathbb{D}_{\alpha_1}$ with dispersion γ_1. Hence, the power of Z is P_1.

– Otherwise, when $\alpha_1 = \alpha_2$:

$$\log \phi_Z(\omega) = -(\gamma_1^{\alpha_1} + \gamma_2^{\alpha_1}) |\omega|^{\alpha_1} [1 + o(1)]. \tag{4.12}$$

Then, the power of Z is $\mathrm{P}_1 + \mathrm{P}_2$.

(vi) The fact that $\mathrm{P}(\kappa X) = |\kappa|^{\tau_{\max}} \mathrm{P}(X)$ makes the power of the scaled variable κX dependent not only on the scale κ but also on $\tau_{\max}$ which is not consistent with R2. We normalize $\mathrm{P}(X)$ as:

$$\mathrm{P}_{\mathrm{nor}}(X) = [\mathrm{P}(X)]^{\frac{1}{\tau_{\max}}}$$

$0 < \tau_{\max} \leq 2$. Under this definition of power, we obtain $\mathrm{P}_{\mathrm{nor}}(\kappa X) = |\kappa| \mathrm{P}_{\mathrm{nor}}(X)$. We note that the triangular inequality is conserved only when $1 \leq \tau_{\max} \leq 2$. In fact, let X, Y be two independent RVs satisfying a, b and

c. Then

$$\begin{aligned} P_{\text{nor}}(X+Y) = [P(X+Y)]^{\frac{1}{\tau_{\max}}} &\leq [(P(X)+P(Y))]^{\frac{1}{\tau_{\max}}} \\ &\leq [P(X)]^{\frac{1}{\tau_{\max}}} + [P(Y)]^{\frac{1}{\tau_{\max}}} \\ &= P_{\text{nor}}(X) + P_{\text{nor}}(Y), \end{aligned}$$

by virtue of the fact that $\frac{1}{\tau_{\max}} \leq 1$ for $1 \leq \tau_{\max} \leq 2$ and $P(X)$, $P(Y) \geq 0$. We refer to P_{nor} as the normalized power. Whenever $X \sim F(X) \in \mathbb{D}_\alpha$, we have:

$$P_{\text{nor}}(X) = [P(X)]^{\frac{1}{\alpha}}.$$

(vii) It is easy to check that $\mathbb{D}$ is closed under convolution. In fact, let $0 < \alpha_1 \leq \alpha_2 \leq 2$ and denote by $\mathbb{D}_{\alpha_1} * \mathbb{D}_{\alpha_2}$ the set whose elements are the convolution of the elements in $\mathbb{D}_{\alpha_1}$ with those $\mathbb{D}_{\alpha_2}$. Then, by equations (4.11) and (4.12), $\mathbb{D}_{\alpha_1} * \mathbb{D}_{\alpha_2} = \mathbb{D}_{\alpha_1}$.

(viii) We note that the new power operator cannot be used to evaluate power for variables that are in the full domain of attraction of stable laws, since outside the domain of normal attraction one can not guarantee the existence of $P(\cdot)$. This is in line with the fact that the domain of attraction of the Gaussian type contains infinite second moment variables, the domain of normal attraction being the subset composed of variables having a finite one.

(iX) We could have defined from the start $\mathbb{T}$ as the full space where we guarantee that $P(\cdot)$ is defined i.e. the space of variables satisfying a, b and c. Then, we write $\mathbb{T}$ as the union of the disjoint spaces $\mathbb{T}_\alpha$ each corresponding to a $\tau_{\max} = \alpha$, $0 < \alpha \leq 2$. We have that $\mathbb{D}_\alpha \subset \mathbb{T}_\alpha$ for all $0 < \alpha \leq 2$ and therefore $\mathbb{D} \subset \mathbb{T}$. The only drawback of the definition of $P(\cdot)$ over $\mathbb{T}$ is that we cannot guarantee that $P(X) = 0 \Longrightarrow X = 0$. A problem that is solved when considering only the set $\mathbb{D}$.

(X) Finally, this new approach of defining a generalized measure of the power of RVs by considering the local behaviour of characteristic functions around zero is an explicit definition of the dispersion of RVs. In fact, it evaluates to a scaled dispersion for the alpha-stable variables and within $\mathbb{D}_\alpha$. However, from an information theoretic perspective, this definition is faced with a serious drawback that can be summarized by the fact that a bounded $\mathrm{P}(\cdot)$ does not imply a bounded differential entropy. As an example, consider $X \sim \mathcal{S}(\alpha,\gamma)$, $0 < \alpha < 2$ and $Y \sim \mathcal{N}(0,\sigma^2)$ then according to to the discussion in point (v),

$$\mathrm{P}(X+Y) = \mathrm{P}(X) = \alpha\gamma^\alpha,$$

and

$$h(X+Y) \geq h(Y) = \frac{1}{2}\ln(2\pi e\sigma^2),$$

where we used the EPI to write lower bound. Letting $\sigma^2 \to +\infty$ then $h(X+Y) \to +\infty$ while $\mathrm{P}(X+Y)$ is maintained constant.

4.4 A Relative Power Measure

As a notion of average power, the second moment is the answer to a widely known result in communications theory; it is the input cost constraint under which a Gaussian input achieves the capacity of the AWGN channel. In order to come up with a notion of average power in the presence of alpha-stable distributions, the Gaussian channel result immediately suggests finding the input cost constraint for which the capacity of the additive independent alpha-stable noise channel is achieved by an alpha-stable input? We address this concept by restricting our quest to the class of SαS distributions as they present a natural extension to the Gaussian model and we proceed in two steps: First we solve this problem for the Cauchy channel ($\alpha = 1$, $\beta = 0$) as the Cauchy density has a closed-form expression which yields explicit expressions for the cost function and the power measure. At a second step, we make generic definitions and statements by following similar steps as in the Cauchy channel. Finally, we use the new power measure in the context of studying the capacity of stable channels. The new power is numerically evaluated for different types of probability distributions that do not necessarily have a finite second moment.

4.4.1 A Base Case: The Cauchy Channel

We consider in this section the Additive Independent Cauchy Noise (AICN) channel, where the noise is modeled as Cauchy distributed. Then, we propose a logarithmic input constraint under which a Cauchy input is proven to be optimal and achieves capacity. The input constraint is parametrized by a scalar k which will be interpreted as a power candidate for a substantially large set of RVs not necessarily having finite second moments. We draw a parallelism between this setup and that of the Gaussian channel under the second moment constraint treated by Shannon [54]. In fact, a Cauchy input yields a Cauchy output over this channel and achieves a capacity value of "$\log(1 + \text{SNR})$".

We define the AICN channel as:

$$Y = X + N, \tag{4.13}$$

where X, Y are the input and output respectively and N is an independent Cauchy RV[1], $N \sim \mathcal{C}(n_o; \gamma)$ of location parameter $n_o \in \mathbb{R}$ and dispersion $\gamma > 0$.

[1] The PDF of a Cauchy RV, $X \sim \mathcal{C}(\delta; \gamma)$ is given by:

$$p_X(x) = \frac{1}{\pi\gamma} \frac{1}{1 + \left[\frac{x-\delta}{\gamma}\right]^2}$$

Equivalently, the channel transition PDF is:

$$p_{Y|X}(y|x) = \frac{1}{\pi\gamma}\frac{1}{1 + \left[\frac{y-x-n_o}{\gamma}\right]^2}.$$

Since $I(X;Y) = I(X;Y-n_o)$ and $(N-n_o) \sim \mathcal{C}(0;\gamma)$, we assume WLOG in what follows that $N \sim \mathcal{C}(0;\gamma)$.

It is well known that a Cauchy distribution $X \sim \mathcal{C}(0;k)$ maximizes the entropy among all RVs X that satisfy $\mathsf{E}\left[\ln(1+\left[\frac{X}{k}\right]^2)\right] = \ln(4)$ [90, Sec.3.1.3, p.51] with an entropy value of $\ln(4\pi k)$. However, the space of RVs for which a Cauchy input achieves the capacity of the Cauchy channel is still to be determined. In the remainder of this section, we state and prove a theorem that answers this question.

Theorem 15 (Capacity of the Cauchy Channel)**.** *When the input distribution functions $F(\cdot)$ are subject to the constraint*

$$\mathsf{E}\left[\ln\left(\left[\frac{A+\gamma}{A}\right]^2 + \left[\frac{X}{A}\right]^2\right)\right] \leq \ln 4, \tag{4.14}$$

for a given $A \geq \gamma$, the capacity of channel (4.13) is

$$C = \log\left(\frac{A}{\gamma}\right), \tag{4.15}$$

and is achieved by $X^* \sim \mathcal{C}(0;A-\gamma)$[2]*.*

Before presenting the proof of the theorem, we make the following observations:

[2] *We define $\mathcal{C}(0;0)$ as the Dirac delta function*

(i) First, we point out that the constraint defined by (4.14) is equivalent to:

$$\exists\, k \in [\gamma, A],\ \mathsf{E}\left[\ln\left(\left[\frac{k+\gamma}{k}\right]^2 + \left[\frac{X}{k}\right]^2\right)\right] = \ln 4. \tag{4.16}$$

Indeed, the function

$$g(k) \mathrel{\hat{=}} \mathsf{E}\left[g(k;X)\right] \ \mathrel{\hat{=}} \mathsf{E}\left[\ln\left(\left[\frac{k+\gamma}{k}\right]^2 + \left[\frac{X}{k}\right]^2\right)\right],$$

is greater or equal to $\ln 4$ at $k = \gamma$: $g(\gamma) \geq \ln 4$. It is also decreasing and continuous in k. Monotonicity follows from the fact that the function

$$g(k;x) \mathrel{\hat{=}} \ln\left(\left[\frac{k+\gamma}{k}\right]^2 + \left[\frac{x}{k}\right]^2\right)$$

is decreasing in k. Continuity is due to the positiveness, monotonicity and continuity of $g(k;x)$ in k by the results of the Monotone Convergence Theorem (MCT).

(ii) Second, note that whenever $A < \gamma$, $\mathsf{E}\left[\ln\left(\left[\frac{A+\gamma}{A}\right]^2 + \left[\frac{X}{A}\right]^2\right)\right]$ is necessarily greater than $\ln 4$, the set of feasible inputs is the empty set and the problem is ill-defined.

(iii) Among all distribution functions $F(\cdot)$ having a finite logarithmic moment of the form:

$$\int_{|x|\geq\epsilon} \ln|x|\, dF(x) < +\infty, \tag{4.17}$$

for some $\epsilon > 0$, such a $k \geq 0$ always exists.

This can be established using the facts that $g(\gamma) \geq \ln 4$ and that whenever (4.17) is finite, $g(k) = \mathsf{E}\left[g(k;X)\right]$ is finite as well for all $k \geq \gamma$ and is decreasing to 0 as $k \to +\infty$.

(iv) The parameter k –whenever it exists– defined by the constraint (4.16) satisfies the following properties:

1- k is greater or equal to γ.

2- $k = \gamma$ if and only if $X = 0$.

3- k is increasing with the absolute value of a scale parameter. Indeed let $U = cX$ be a scaled version of RV X and assume that U satisfies equation (4.16) with parameter $k > \gamma$: $\mathsf{E}\left[\ln\left(\left[\frac{k+\gamma}{k}\right]^2 + \left[\frac{U}{k}\right]^2\right)\right] = \ln 4$. The fact that $\mathsf{E}\left[\ln\left(\left[\frac{k+\gamma}{k}\right]^2 + \left[\frac{cX}{k}\right]^2\right)\right]$ is decreasing in k, increasing in $|c|$ and continuous in both yields the required result.

4- As it will shown later, condition (4.16) will imply:

$$\mathsf{E}_Y\left[\ln\left(1 + \left[\frac{Y}{k}\right]^2\right)\right] = \ln 4$$

These properties indicate that one may view k as a measure of "power" of the output variable Y. When X is zero, this parameter is a measure of the power of N and is equal to γ. The result of Theorem 15 is that the capacity of the channel is the logarithm of the maximum received SNR.

Additionally, the quantity $\mathrm{P}_X > 0$ such that:

$$\mathsf{E}_X\left[\ln\left(1 + \left[\frac{X}{\mathrm{P}_X}\right]^2\right)\right] = \ln 4 \tag{4.18}$$

may be viewed as a measure of power of the input variable X. It satisfies properties R1, R2 and R3 stated in Section 4.1 and that are common to various known notions of power such that the variance or the dispersion. Additionally, the parameter P_X has two main advantages over the previously cited notions of power:

– P_X is defined for a bigger space than the variance. Namely, the space defined by (4.16) contains infinite variance RVs.

– As opposed to the dispersion of a RV (as proposed in [11]) which does not possess a closed form analytical expression, the parameter P_X can be always evaluated when it exists.

(v) One could define the power of the output to be "the square of the parameter k", in which case the capacity of the channel would be the usual half log of the received SNR.

We now proceed to proving Theorem 15.

Proof. Based on the previous observations, the input space is that of RV with distribution functions $F(\cdot)$ satisfying:

$$\mathsf{E}_X\left[\ln\left(\left[\frac{k+\gamma}{k}\right]^2+\left[\frac{X}{k}\right]^2\right)\right]=\ln 4,$$

for some $\gamma \leq k \leq A$, $A > \gamma$. The case $k = \gamma$ can be omitted since it corresponds to the Dirac delta distribution that yields a null mutual information. For channel (4.13) and since X and N are independent, using iterated expectations,

$$\mathsf{E}_Y\left[\ln\left(1+\left[\frac{Y}{k}\right]^2\right)\right]=\mathsf{E}_X\left[\mathsf{E}_N\left[\ln\left(1+\left[\frac{X+N}{k}\right]^2\right)\middle|X\right]\right]. \tag{4.19}$$

The inner expectation evaluates to

$$\mathsf{E}_N\left[\ln\left(1+\left[\frac{X+N}{k}\right]^2\right)\middle|X=x\right]=\frac{1}{\pi\gamma}\int_{-\infty}^{\infty}\ln\left(1+\left[\frac{x+n}{k}\right]^2\right)\frac{1}{1+\left[\frac{n}{\gamma}\right]^2}\,dn.$$

We prove in Appendix E that the function,

$$\begin{aligned} f(x;\xi) &\mathrel{\hat{=}} \frac{\xi}{\pi}\int_{-\infty}^{\infty}\ln(1+u^2)\,\frac{1}{1+(\xi u-x)^2}\,du, \qquad (4.20)\\ &= \ln\left(\left[\frac{\xi+1}{\xi}\right]^2+\left[\frac{x}{\xi}\right]^2\right). \end{aligned}$$

Using the change of variable $u = \frac{x+n}{k}$ equation (4.19) becomes,

$$\begin{aligned} \mathsf{E}_Y\left[\ln\left(1+\left[\frac{Y}{k}\right]^2\right)\right] &= \mathsf{E}_X\left[f\left(\frac{X}{\gamma};\frac{k}{\gamma}\right)\right] \\ &= \mathsf{E}_X\left[\ln\left(\left[\frac{k+\gamma}{k}\right]^2+\left[\frac{X}{k}\right]^2\right)\right] = \ln 4, \end{aligned}$$

which shows that when (4.16) is satisfied, the output entropy is maximized whenever $Y \sim \mathcal{C}(0;k)$ which is possible if and only if $X \sim \mathcal{C}(0;k-\gamma)$. It remains to check whether $X \sim \mathcal{C}(0;k-\gamma)$ satisfies the constraint (4.16). In fact, if $X \sim \mathcal{C}(0;k-\gamma)$,

$$p_X(x) = \frac{1}{\pi(k-\gamma)}\frac{1}{1+\left[\frac{x}{k-\gamma}\right]^2}$$

and

$$\begin{aligned} &\mathsf{E}_X\left[\ln\left(\left[\frac{k+\gamma}{k}\right]^2+\left[\frac{X}{k}\right]^2\right)\right] \\ &= \ln\left[\frac{k+\gamma}{k}\right]^2 + \mathsf{E}_X\left[\ln\left(1+\left[\frac{X}{k+\gamma}\right]^2\right)\right] \\ &= \ln\left[\frac{k+\gamma}{k}\right]^2 + \frac{k+\gamma}{\pi[k-\gamma]}\int_{-\infty}^{\infty}\ln(1+u^2)\frac{1}{1+\left[\frac{k+\gamma}{k-\gamma}u\right]^2}\,du \qquad (4.21) \\ &= \ln\left[\frac{k+\gamma}{k}\right]^2 + f\left(0;\frac{k+\gamma}{k-\gamma}\right) \\ &= \ln\left[\frac{k+\gamma}{k}\right]^2 + \ln\left(\frac{\frac{k+\gamma}{k-\gamma}+1}{\frac{k+\gamma}{k-\gamma}}\right)^2 = \ln 4, \end{aligned}$$

where we used the change of variable $u = \frac{x}{k+\gamma}$ to write (4.21). Hence, for a given fixed $k \geq \gamma$, the mutual information of (4.13) is maximized by $X^* \sim \mathcal{C}(0;k-\gamma)$ under the constraint (4.16) for which the output is $Y \sim \mathcal{C}(0;k)$. The value of the mutual information is given by:

$$\begin{aligned} I(X;Y) &= h(Y) - h(Y|X) = h\left(\mathcal{C}(0;k)\right) - h\left(\mathcal{C}(0;\gamma)\right) \\ &= \ln 4\pi k - \ln 4\pi\gamma = \ln\left(\frac{k}{\gamma}\right), \end{aligned}$$

where we used the fact that $h\left(\mathcal{C}(0;\gamma)\right) = \ln 4\pi\gamma$ for $\gamma > 0$. Since the expression is increasing with k, the optimal input $X^* \sim \mathcal{C}(0; A - \gamma)$ is capacity-achieving whenever $\gamma \leq k \leq A$. The capacity value is therefore

$$C = \ln\left(\frac{A}{\gamma}\right) = \ln\left(1 + \frac{P}{\gamma}\right).$$

□

We present in the remaining part of this section some numerical evaluations for some parameters and quantities defined above. In Figure 4-1, we plot the parameter $P = k$ –where k is defined in equation (4.16) for $\gamma = 1$, versus the scale parameter c for multiple symmetric stable densities with characteristic exponent α considered as inputs to channel (4.13). It can be seen that P increases with the scale as already proven.

Also, we note that P is decreasing with α for a fixed scale. This can be explained by the fact that the tails of the probability density function (PDF) for the stable family becomes heavier as α gets smaller, hence the higher power. Figure 4-2 shows the mutual information between the input and the output of channel (4.13) when $N \sim \mathcal{C}(0;1)$ and when the input is symmetric alpha-stable. The highest rate is achieved by a Cauchy input ($\alpha = 1$) which is the capacity by the results of Theorem 15. As α deviates from 1 whether higher or lower, the transmission rates decrease.

Looking back to equation (4.18), it defines a new measure to compute the signal strength. This definition is intimately related to the presence of a Cauchy variable as is the second moment in the presence of a Gaussian variable. We

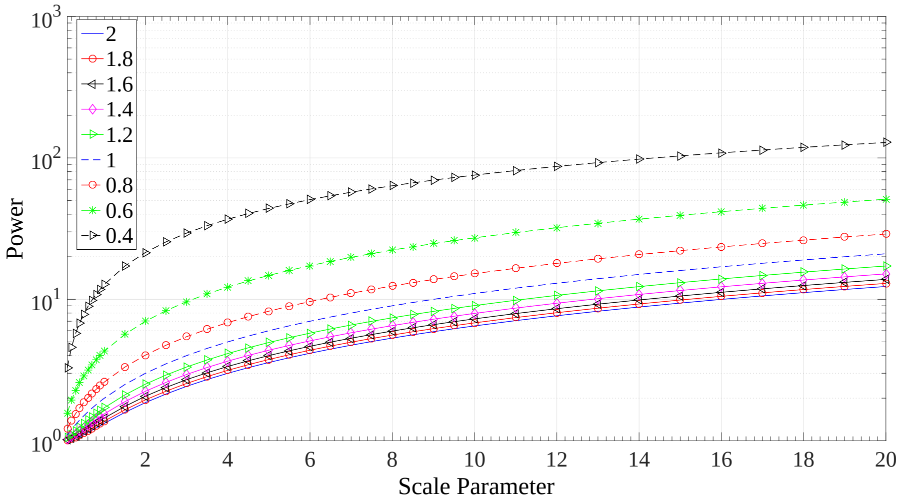

Figure 4-1: P versus scale for symmetric alpha-stable densities for different values of α. The scale parameter ranges from 0.1 to 20.

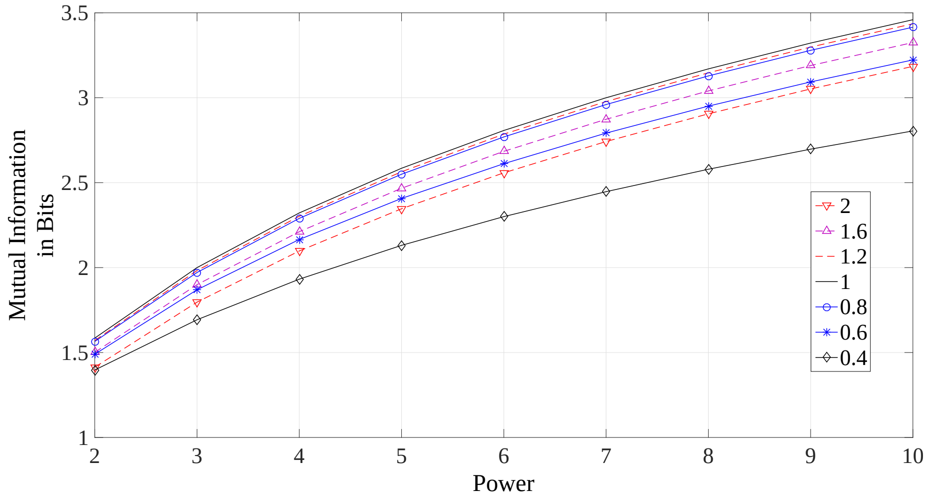

Figure 4-2: Mutual information in bits per channel use versus the parameter k for different alpha-stable inputs.

generalize the two definitions into a generic one which can be used to evaluate the signal's power with respect to a standard stable variable.

4.4.2 Location and Power Parameters in the Presence of Stable Variables

Let $\tilde{Z} \sim \mathcal{S}(\alpha, \left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha}})$ be a standard symmetric stable RV $(0 < \alpha \leq 2)$. For the rest of this chapter, we only consider RVs X such that $\mathsf{E}\left[\ln\left(1+|X|\right)\right]$ is finite.

Definition 5 (Location Parameter)**.** The location parameter of a RV X is the real scalar L_X such that:

$$\mathrm{L}_X = \mathrm{argmin}_{\nu\in\mathbb{R}} - \mathrm{E}\left[\ln p_{\tilde{Z}}(X-\nu)\right]. \tag{4.22}$$

Gonzales adopted the same methodology to define the zero-order location based on logarithmic moments [39].

Definition 6 (Power Parameter)**.** Except when $X = 0$, the power of a RV X is the non-negative scalar P_X such that:

$$-\mathrm{E}\left[\ln p_{\tilde{Z}}\left(\frac{X}{\mathrm{P}_X}\right)\right] = h(\tilde{Z}), \tag{4.23}$$

where $h(\tilde{Z})$ is the differential entropy of $\tilde{Z}$. Furthermore, $X = 0$ implies that $\mathrm{P}_X = 0$.

A better way to think about P_X is to consider it as a "relative power" with respect to $\tilde{Z}$ which can be considered as a reference variable whose power is equal to unity. Equations (4.22) and (4.23) can be evaluated in two special cases:

- When $\alpha = 2$ and $\tilde{Z} \sim \mathcal{N}(0;1)$ is a standard Gaussian RV with zero mean and unit variance,

$$\mathrm{L}_X = \mathrm{argmin}_{\nu\in\mathbb{R}}\mathrm{E}\left[(X-\nu)^2\right] = \mathrm{E}\left[X\right], \tag{4.24}$$

  and

$$\mathrm{P}_X = \left(\mathrm{E}\left[X^2\right]\right)^{\frac{1}{2}}. \tag{4.25}$$

- When $\alpha = 1$ and $N \sim \mathcal{C}(0;1)$ is a standard Cauchy RV with shift parameter 0 and unit dispersion,

$$\mathrm{L}_X = \mathrm{argmin}_{\nu\in\mathbb{R}} \mathrm{E}\left[\ln\left(1 + (X-\nu)^2\right)\right], \tag{4.26}$$

and P_X is such that:

$$\mathrm{E}\left[\ln\left(1 + \left(\frac{X}{\mathrm{P}_X}\right)^2\right)\right] = \ln 4, \tag{4.27}$$

where we used the fact that $h(\tilde{z}) = h\left(\mathcal{C}(0;1)\right) = \ln(4\pi)$.

The quantities L_X and P_X as defined in (4.22) and (4.23) are endowed respectively with some location and power properties.

Properties of $\mathbf{L}_X$:

(i) The location parameter is linear in the additive term, namely for any real number b:

$$\mathrm{L}_{X+b} = \mathrm{L}_X + b.$$

This directly follows from equation (4.22).

(ii) Whenever the distribution function of X is symmetric [3] with respect to μ_X, then $\mathrm{L}_X = \mu_X$. In fact, for any $\nu \geq 0$, taking the derivative of $-\mathrm{E}\left[\ln p_{\tilde{Z}}(X-\nu)\right]$ with respect to ν and applying the derivation inside the expectation operator gives:

$$\mathrm{E}\left[\frac{p'_{\tilde{Z}}(X-\nu)}{p_{\tilde{Z}}(X-\nu)}\right] = \int_{\mathbb{R}} \frac{p'_{\tilde{Z}}(x)}{p_{\tilde{Z}}(x)}\, dF_X(x+\nu) = 0, \tag{4.28}$$

iff $\nu = \mu_X$ by virtue of the fact that $p_{\tilde{Z}}(\cdot)$ is even, $p'_{\tilde{Z}}(\cdot)$ is odd and $p_X(x)$

[3] A distribution function $F(x)$ is said to be symmetric with respect to $\mu \in \mathbb{R}$ if and only if $F(\mu + x) = 1 - F(\mu - x) + \Pr(X = \mu)$ for all $x \in \mathbb{R}$.

is symmetric with respect to μ_X. This implies that:

$$\mathrm{L}_X = \mathrm{argmin}_{\nu \in \mathbb{R}} - \mathrm{E}\left[\ln p_{\tilde{Z}}(X - \nu)\right] = \mu_X. \tag{4.29}$$

The interchange between the derivative and the exception operator is justified by DCT since $\frac{p'_{\tilde{Z}}}{p_{\tilde{Z}}}(\cdot)$ is bounded (see Theorem 29 in Section D.3 for the tail behaviour of $p^{(n)}_{\tilde{Z}}(\cdot)$ the n-th derivative of $p_{\tilde{Z}}(\cdot)$). A direct consequence is that:

$$\mathrm{L}_{aX} = \mathrm{argmin}_{\nu \in \mathbb{R}} - \mathrm{E}\left[\ln p_{\tilde{Z}}(aX - \nu)\right] = a\mathrm{L}_X,$$

for any $a \in \mathbb{R}$.

(iii) Let X and Y be two independent RVs with symmetric PDFs having location parameters (points of symmetry) $\mu_X = \mathrm{L}_X$ and $\mu_Y = \mathrm{L}_Y$. Then,

$$\mathrm{L}_{X+Y} = \mathrm{L}_X + \mathrm{L}_Y. \tag{4.30}$$

This results from the fact that $p_Z(\cdot)$ the PDF of $Z = X + Y$ is symmetric with respect to $\mu_Z = \mu_X + \mu_Y$ and the fact that $\mathrm{L}_Z = \mu_Z$ by property (ii).

Properties of $\mathbf{P}_X$:

(iv) We start by showing that P_X satisfies property R1, i.e. $\mathrm{P}_X \geq 0$ with equality if and only if $X = 0$. According to Definition 6, it only suffices to show that:

$$\mathrm{P}_X = 0 \implies X = 0. \tag{4.31}$$

In order to prove equation (4.31), we assume $X \neq 0$ and we start by considering the LHS of equation (4.23) as a function of P:

$$g(\mathrm{P}) \,\hat{=}\, - \mathrm{E}\left[\ln p_{\tilde{Z}}\left(\frac{X}{\mathrm{P}}\right)\right] = -\mathrm{E}\left[\ln p_{\tilde{Z}}\left(\frac{|X|}{\mathrm{P}}\right)\right].$$

We prove $g\left(\mathrm{P}\right)$ to be continuous on $\mathbb{R}^{+*}$: let $\mathrm{P}_0 > 0$, then

$$\begin{aligned}
-\lim_{\mathrm{P}\to\mathrm{P}_0} \mathrm{E}\left[\ln p_{\tilde{Z}}\left(\frac{X}{\mathrm{P}}\right)\right] &= -\lim_{\mathrm{P}\to\mathrm{P}_0}\int \ln p_{\tilde{Z}}\left(\frac{x}{\mathrm{P}}\right)\, dF(x) \\
&= -\int \lim_{\mathrm{P}\to\mathrm{P}_0} \ln p_{\tilde{Z}}\left(\frac{x}{\mathrm{P}}\right)\, dF(x) \\
&= -\int \ln p_{\tilde{Z}}\left(\frac{x}{\mathrm{P}_0}\right)\, dF(x),
\end{aligned}$$

where $F(x)$ is the distribution function of X and in order to write the last equation we used the fact that $p_{\tilde{z}}(\cdot)$ is continuous on $\mathbb{R}$. The interchange in the order between the limit and the integral signs is justified by DCT by virtue of:

- In a neighbourhood of P_0, there exists a $\tilde{\mathrm{P}}$ such that for all $x \in \mathbb{R}$

$$\left|\ln p_{\tilde{Z}}\left(\frac{x}{\mathrm{P}}\right)\right| \leq \left|\ln p_{\tilde{Z}}\left(\frac{x}{\tilde{\mathrm{P}}}\right)\right|. \tag{4.32}$$

- Equation (4.32) implies that:

$$\mathrm{E}\left|\ln p_{\tilde{Z}}\left(\frac{X}{\mathrm{P}}\right)\right| \leq \mathrm{E}\left|\ln p_{\tilde{Z}}\left(\frac{X}{\tilde{\mathrm{P}}}\right)\right|,$$

  which is finite whenever $X \in \mathcal{L}$ if $\tilde{Z}$ is alpha-stable and whenever X has a finite variance if $\tilde{Z}$ is Gaussian.

We notice that $g\left(\mathrm{P}\right)$ is non-increasing in $\mathrm{P} > 0$ by virtue of the fact that $p_{\tilde{Z}}(\cdot)$ is symmetric and decreasing on the positive semi-axis. We evaluate next the limit values of $g(\mathrm{P})$ at 0 and $+\infty$.

- <u>The limit at zero:</u> since $X \neq 0$ there exits a $\delta > 0$ s.t. $\Pr\left(|X| \geq \delta\right) > 0$

and

$$g\left(\mathrm{P}\right) = -\int_{|x|\leq\delta} \ln p_{\tilde{Z}}\left(\frac{x}{\mathrm{P}}\right)\, dF_X(x) - \int_{|x|\geq\delta} \ln p_{\tilde{Z}}\left(\frac{x}{\mathrm{P}}\right)\, dF_X(x)$$
$$\geq -\Pr\left(|X|\leq\delta\right)\ln p_{\tilde{Z}}(0) - \Pr\left(|X|\geq\delta\right)\ln p_{\tilde{Z}}\left(\frac{\delta}{\mathrm{P}}\right), \quad (4.33)$$

where in order to write equation (4.33) we used the fact that $p_{\tilde{Z}}(\cdot)$ is decreasing on $\mathbb{R}^+$. Since $p_{\tilde{Z}}(x) \to 0$ as $|x| \to \infty$, then:

$$\lim_{\mathrm{P}\to 0} g\left(\mathrm{P}\right) = +\infty. \quad (4.34)$$

– The limit at infinity:

$$\lim_{\mathrm{P}\to+\infty} g(\mathrm{P}) = \lim_{\mathrm{P}\to+\infty} -\int_{\mathbb{R}} \ln p_{\tilde{Z}}\left(\frac{x}{\mathrm{P}}\right)\, dF_X(x)$$
$$= -\int_{\mathbb{R}} \lim_{\mathrm{P}\to+\infty} \ln p_{\tilde{Z}}\left(\frac{x}{\mathrm{P}}\right)\, dF_X(x)$$
$$= -\ln p_{\tilde{Z}}(0) \leq h\left(\tilde{Z}\right),$$

where the last inequality is true since $\max_{x\in\mathbb{R}} p_{\tilde{Z}}(x) = p_{\tilde{Z}}(0)$. The interchange between the limit and the integral sign is due to DCT. Indeed since $g\left(\mathrm{P}\right)$ is non-increasing, then for $\mathrm{P} > \mathrm{P}_t$:

$$-\int_{\mathbb{R}} \ln p_{\tilde{Z}}\left(\frac{x}{\mathrm{P}}\right)\, dF_X(x) \leq -\int_{\mathbb{R}} \ln p_{\tilde{Z}}\left(\frac{x}{\mathrm{P}_t}\right)\, dF_X(x),$$

which is integrable since $\mathsf{E}\left[\ln\left(1+|X|\right)\right]$ is finite and $\ln p_{\tilde{Z}}(|x|) = \Theta\left(\ln|x|\right)$.

Using the continuity of $g(\mathrm{P})$ (the LHS of equation (4.23)) and the fact that it is non-increasing from $+\infty$ to $-\ln p_{\tilde{Z}}(0)$ ($\leq h(\tilde{Z})$), there exists a non-zero P_X such that equation (4.23) is satisfied which proves equation (4.31).

(v) We prove now that P_X satisfies property R2. In fact, for any $a \in \mathbb{R}$,

$$\mathrm{P}_{aX} = |a|\,\mathrm{P}_X.$$

This directly follows from equation (4.23) and the fact that $p_{\tilde{Z}}$ is even.

(vi) Let X and Y be two independent RVs such that their distribution functions are symmetric with respect to zero and define $Z = X + Y$. Then $\mathrm{L}_X = \mathrm{L}_Y = 0$, and

$$\mathrm{P}_Z \geq \max\{\mathrm{P}_X; \mathrm{P}_Y\}. \tag{4.35}$$

The proof goes as follows:

$$\begin{aligned}
-\mathrm{E}\left[\ln p_{\tilde{Z}}\left(\frac{Z}{\mathrm{P}_Y}\right)\right] &= \mathrm{E}_X\left[-\mathrm{E}_Y\left[\ln p_{\tilde{Z}}\left(\frac{x+Y}{\mathrm{P}_Y}\right)\Big|X\right]\right] && (4.36)\\
&\geq \mathrm{E}_X\left[-\mathrm{E}_Y\left[\ln p_{\tilde{Z}}\left(\frac{Y}{\mathrm{P}_Y} - \mathrm{L}_{\frac{Y}{\mathrm{P}_Y}}\right)\right]\right] && (4.37)\\
&= \mathrm{E}_X\left[-\mathrm{E}_Y\left[\ln p_{\tilde{Z}}\left(\frac{Y}{\mathrm{P}_Y}\right)\right]\right] && (4.38)\\
&= h(\tilde{Z}), && (4.39)
\end{aligned}$$

where equation (4.39) is due to the definition of P_Y. Equation (4.36) is due to the fact that X and Y are independent and equation (4.37) is due to the definition of the location parameter. Equation (4.38) is justified by the fact that $\mathrm{L}_{\frac{Y}{\mathrm{P}_Y}} = \frac{\mathrm{L}_Y}{\mathrm{P}_Y} = 0$. Equation (4.39) implies that $\mathrm{P}_Z \geq \mathrm{P}_Y$ since the function $-\mathrm{E}\left[\ln p_{\tilde{Z}}\left(\frac{Z}{\mathrm{P}}\right)\right]$ is non-increasing in $\mathrm{P} \geq 0$. Similarly, we prove that $\mathrm{P}_Z \geq \mathrm{P}_Y$.

(vii) Let X and Y be two independent RVs such that Y has an even PDF that is non-increasing on $\mathbb{R}^+$, then P_{cX+Y} is increasing with $|c|$, $c \in \mathbb{R}$. In fact, it has been already proven that $-\mathrm{E}\left[\ln p_{\tilde{Z}}\left(\frac{cX+Y}{\mathrm{P}}\right)\right]$ is non-increasing in $\mathrm{P} > 0$. Next, we show that $-\mathrm{E}\left[\ln p_{\tilde{Z}}\left(\frac{cX+Y}{\mathrm{P}}\right)\right]$ is non-decreasing in $|c|$. To this end, we write

$$-\mathrm{E}\left[\ln p_{\tilde{Z}}\left(\frac{cX+Y}{\mathrm{P}}\right)\right] = \mathrm{E}_X\left[-\mathrm{E}_Y\left[\ln p_{\tilde{Z}}\left(\frac{cx+Y}{\mathrm{P}}\right)\Big|X\right]\right], \qquad (4.40)$$

and it is enough to show that $-\mathrm{E}_Y\left[\ln p_{\tilde{Z}}\left(\frac{cx+Y}{\mathrm{P}}\right)\right]$ is non-decreasing in $|c|$. We have

$$\begin{aligned}
&-\mathrm{E}_Y\left[\ln p_{\tilde{Z}}\left(\frac{cx+Y}{\mathrm{P}}\right)\right] \\
&= -\int_{-\infty}^{+\infty} p_Y(y)\ln p_{\tilde{Z}}\left(\frac{cx+y}{\mathrm{P}}\right)\,dy \qquad (4.41)\\
&= -\int_{-\infty}^{+\infty} p_Y(u-cx)\ln p_{\tilde{Z}}\left(\frac{u}{\mathrm{P}}\right)\,du \\
&= -\int_{0}^{+\infty} p_Y(u-cx)\ln p_{\tilde{Z}}\left(\frac{u}{\mathrm{P}}\right)\,du - \int_{0}^{+\infty} p_Y(u+cx)\ln p_{\tilde{Z}}\left(\frac{u}{\mathrm{P}}\right)\,du. \qquad (4.42)
\end{aligned}$$

Equation (4.42) shows the symmetry of the expectation with respect to c and x. Therefore, one can restrict the proof to the case when c and x are non-negative. Hence, for c and x non-negative, taking the derivative of equation (4.41) with respect to c and interchanging the limit and the derivative sign as done in (ii) yields

$$-\frac{x}{\mathrm{P}}\int_{-\infty}^{+\infty} p_Y(y)\frac{p'_{\tilde{Z}}\left(\frac{cx+y}{\mathrm{P}}\right)}{p_{\tilde{Z}}\left(\frac{cx+y}{\mathrm{P}}\right)}\,dy = -\frac{x}{\mathrm{P}}\mathsf{E}\left[\frac{p'_{\tilde{Z}}\left(\frac{cx+Y}{\mathrm{P}}\right)}{p_{\tilde{Z}}\left(\frac{cx+Y}{\mathrm{P}}\right)}\right] \geq 0,$$

which is true by virtue of the fact that $p_Y(y)$ is even non-increasing on $\mathbb{R}^+$, $\frac{p'_{\tilde{Z}}}{p_{\tilde{Z}}}(\cdot)$ is an odd function that is non-positive on $\mathbb{R}^+$ and both c and x are non-negative. This implies that $-\mathrm{E}_Y\left[\ln p_{\tilde{Z}}\left(\frac{cx+Y}{\mathrm{P}}\right)\right]$ and $-\mathrm{E}\left[\ln p_{\tilde{Z}}\left(\frac{cX+Y}{\mathrm{P}}\right)\right]$ are non-decreasing in $|c|$. The fact that $-\mathrm{E}\left[\ln p_{\tilde{Z}}\left(\frac{cX+Y}{\mathrm{P}}\right)\right]$ is non-increasing in P and non-decreasing in $|c|$ yields the required result.

(viii) Whenever $X \sim \mathcal{S}(\alpha, \gamma_X)$ is a symmetric stable variable, $\mathrm{P}_X = \frac{\gamma_X}{\gamma_{\tilde{Z}}} =$

$(\alpha)^{\frac{1}{\alpha}}\gamma_X$. Indeed, $X \sim \mathcal{S}(\alpha, \gamma_X)$ has the same distribution as $\frac{\gamma_X}{\gamma_{\tilde{Z}}}\tilde{Z}$ and

$$p_X(x) = \frac{\gamma_{\tilde{Z}}}{\gamma_X} p_{\tilde{Z}}\left(\frac{\gamma_{\tilde{Z}}}{\gamma_X}x\right).$$

Therefore, for $\mathrm{P}_X = \frac{\gamma_X}{\gamma_{\tilde{Z}}}$,

$$\begin{aligned}
&- \mathrm{E}\left[\ln p_{\tilde{Z}}\left(\frac{X}{\mathrm{P}_X}\right)\right] \\
&= -\frac{\gamma_{\tilde{Z}}}{\gamma_X}\int p_{\tilde{Z}}\left(\frac{\gamma_{\tilde{Z}}}{\gamma_X}x\right)\ln p_{\tilde{Z}}\left(\frac{x}{\mathrm{P}_X}\right)\,dx \\
&= h(\tilde{Z}).
\end{aligned}$$

Though the power definition as stated in equation (4.23) is implicit and dependent on the density function of symmetric stable variables which does not have closed form expressions except in the special cases of the Cauchy and the Gaussian distributions, the computation of the power P_X of a certain RV X is rather simple using numerical computations. In fact, the stable densities can be computed numerically as inverse Fourier transforms or by using *Matlab* packages that compute these densities such as the "*Stable*" package provided by prof. John P. Nolan. We use here the latter and we develop a *Matlab* code that computes the power according to definition (4.23). We plot in Figure 4-3, the power of several probability laws (Gaussian, uniform, Laplace, Cauchy and alpha-stable ($\alpha = 0.6$) with respect to a multitude of alpha-stable distributions with the characteristic exponent α ranging from 0.4 to 1.8. For example, consider $\tilde{Z}$ with characteristic exponent $\alpha = 1.2$. The power of a Gaussian variable $\mathcal{N}(0, 2)$ with respect to $\tilde{Z}$ is equal to 0.7869. Hence, using the scalability property (v), the power of a Gaussian variable with zero mean and variance σ^2 whenever $\tilde{Z} \sim \mathcal{S}\left(1.2, (1.2)^{-\frac{1}{1.2}}\right)$ is equal to $0.7869\frac{\sigma}{\sqrt{2}} = 0.5564\,\sigma$. Note that as already known, the power the Gaussian variable $X \sim \mathcal{N}(0, \sigma^2)$ with respect to $\tilde{Z} \sim \mathcal{N}(0, 1)$ is equal to σ. Another example is when $X \sim \mathcal{U}[-a, +a]$ a uniform RV with zero mean and variance equal to $\frac{a^2}{3}$: with respect to $\tilde{Z}$ with $\alpha = 0.8$, it has a power of $0.3036\frac{a}{\sqrt{3}} = 0.1753\,a$ whereas with respect to the Gaussian law the power is equal to the standard deviation $\frac{a}{\sqrt{3}} = 0.5774\,a$.

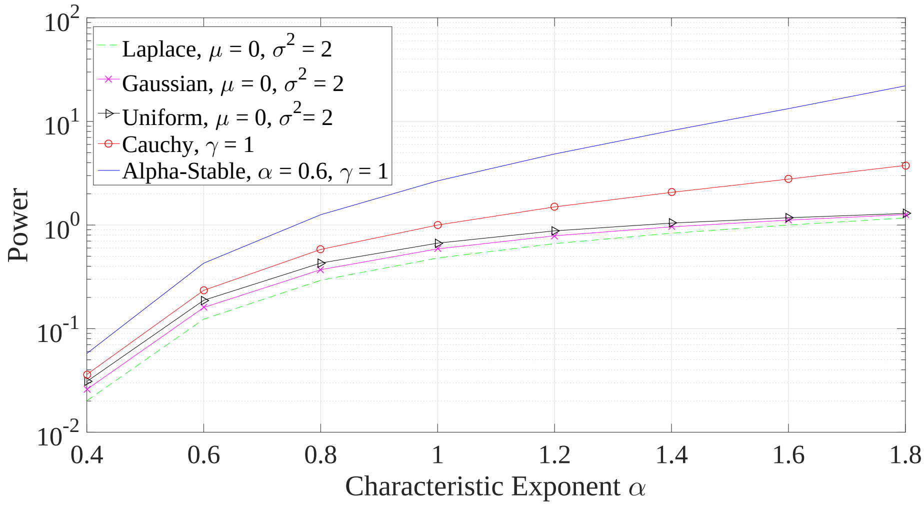

Figure 4-3: Evaluation of the power of some probability laws with respect to $\tilde{Z} \sim \mathcal{S}\left(\alpha, (\alpha)^{-\frac{1}{\alpha}}\right)$ for different values of α.

4.4.3 Applications

Stable Maximizing Entropy

Having defined a generic power definition when considering stable noise environments, a first question is to find the solution of the entropy maximization problem subject to a constraint on the newly defined power. Namely, define:

$$\mathcal{P} = \left\{F \in \mathcal{F} : -\int \ln p_{\tilde{Z}}\left(\frac{X}{\mathrm{P}}\right)\, dF(x) = h(\tilde{Z}), \mathrm{P} > 0\right\}. \tag{4.43}$$

According to [56, Section 12.1], among all distribution functions $F \in \mathcal{P}$, the one that maximizes differential entropy has the following PDF:

$$p^*(x) = e^{\lambda_0 + \lambda_1 \ln p_{\tilde{Z}}\left(\frac{X}{\mathrm{P}}\right)},$$

where λ_0 and λ_1 are chosen so that $p^*(x) \in \mathcal{P}$. A direct solution to the problem is:

$$F^* = \text{argmax}_{F \in \mathcal{P}} h(F), \tag{4.44}$$

where F^* is the distribution function of a symmetric stable variable distributed according to $\mathrm{P}\tilde{Z} \sim \mathcal{S}(\alpha, (\frac{1}{\alpha})^{\frac{1}{\alpha}}\mathrm{P})$ whereby property (5-) we verify that $F^* \in \mathcal{P}$. The value of the maximum is:

$$h(F^*) = h(\tilde{Z}) + \ln \mathrm{P} \tag{4.45}$$

As a direct generalization, one can define:

$$\mathcal{P}_A = \{F \in \mathcal{P} : \mathrm{P} \leq A, A > 0\}, \tag{4.46}$$

for which

$$F_A^* = \text{argmax}_{F \in \mathcal{P}_A} h(F),$$

where F^* is the distribution function of a symmetric stable variable distributed according to $\mathrm{A}\tilde{Z}$.

Communicating over Stable Channels

Consider the additive linear channel:

$$Y = X + N, \tag{4.47}$$

where Y is the channel output, X is the input and $N \sim \mathcal{S}(\alpha, \gamma_N)$ is stable additive noise which is independent of X. We ask the following question: what constraint is to be imposed on the input such that a stable input achieves the capacity of channel (4.47). Under this scenario, and knowing that a stable input of channel (4.47) generates a stable output, a sufficient condition is that the output space induced by the channel is the space where a stable variable maximizes entropy, specifically a space of the form (4.46). To this end, we consider the output space $\mathcal{P}_{[\mathrm{P}_N,A]}$ of probability distributions:

$$\mathcal{P}_{[\mathrm{P}_N,A]} = \{F \in \mathcal{P} : \mathrm{P}_N \leq \mathrm{P} \leq A, A \geq \mathrm{P}_N\}, \tag{4.48}$$

where $\mathrm{P}_N = (\alpha)^{\frac{1}{\alpha}}\gamma_N$. Under condition (4.48), a stable output $Y^* \sim \mathcal{S}(\alpha, \left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha}} A)$ maximizes the output entropy and achieves the channel capacity C:

$$\begin{aligned} C = h(Y^*) - h(N) &= \ln(A) + h(\tilde{Z}) - \ln(\mathrm{P}_N) - h(\tilde{Z}) \\ &= \ln\left(\frac{A}{\mathrm{P}_N}\right) = \ln\left(\mathrm{SNR}_{\mathrm{output}}\right), \end{aligned}$$

where we use the fact that $h(N) = \ln(\mathrm{P}_N) + h(\tilde{Z})$ since $\gamma_N = \mathrm{P}_N\gamma_{\tilde{Z}}$. The optimal input X^* which yields Y^* is also distributed according to a stable variable with parameter γ_{X^*}:

$$\gamma_{X^*}^\alpha = \gamma_{Y^*}^\alpha - \gamma_N^\alpha = \frac{1}{\alpha}(A^\alpha - \mathrm{P}_N^\alpha),$$

which gives by property (5-)

$$\mathrm{P}_{X^*}^\alpha = \alpha\gamma_{X^*}^\alpha = A^\alpha - \mathrm{P}_N^\alpha.$$

Finally, it remains to determine the input cost constraint that yields the output space $\mathcal{P}_{[\mathrm{P}_N,A]}$. The output condition (4.48) is explicitly stated as the space of all RVs Y such that there exists a $\mathrm{P} > 0$, $\mathrm{P}_N \leq \mathrm{P} \leq A$ for some fixed $A > 0$ and

$$\begin{aligned} -\mathrm{E}\left[\ln p_{\tilde{Z}}\left(\frac{Y}{P}\right)\right] &= h(\tilde{Z}) \\ \mathrm{E}_X\left\{-\mathrm{E}_N\left[\ln p_{\tilde{Z}}\left(\frac{X+N}{P}\right)\middle| X = x\right]\right\} &= h(\tilde{Z}), \end{aligned} \tag{4.49}$$

(4.50)

where we used the iterated expectations to write the second equation by virtue of the fact that X and N are independent. Equation (4.49) implies that the input cost function $\mathcal{C}(\cdot)$ is:

$$\mathcal{C}(x) = -\mathrm{E}_N\left[\ln p_{\tilde{Z}}\left(\frac{x+N}{P}\right)\right], \tag{4.51}$$

and the input cost constraint can be stated as follows: there exists a $\mathrm{P} > 0$, $\mathrm{P}_N \leq \mathrm{P} \leq A$ for some fixed $A > 0$ such that:

$$\mathrm{E}\left[\mathcal{C}(X)\right] = h(\tilde{Z}), \tag{4.52}$$

where $\mathcal{C}(\cdot)$ is defined in equation (4.51). The cost function and the cost constraint can be written in a different form. In fact, considering equation (4.51)

$$\begin{aligned}
&\mathcal{C}(x) \\
&= -\int p_{\mathrm{P}_N\tilde{Z}}(n)\ln p_{\tilde{Z}}\left(\frac{x+n}{P}\right) dn \\
&= -\int p_{\mathrm{P}_N\tilde{Z}}(v-x)\ln\left(\frac{1}{\mathrm{P}}p_{\tilde{Z}}\left(\frac{v}{\mathrm{P}}\right)\right) dv - \ln \mathrm{P} \\
&= -\int p_{\mathrm{P}_N\tilde{Z}}(v-x)\ln p_{\mathrm{P}\tilde{Z}}(v)\; dv - \ln \mathrm{P} \\
&= D\left(p_{\mathrm{P}_N\tilde{Z}}(v-x)||p_{\mathrm{P}\tilde{Z}}(v)\right) + h(\mathrm{P}_N\tilde{Z}) - \ln \mathrm{P} \\
&= D\left(p_{\mathrm{P}_N\tilde{Z}}(v-x)||p_{\mathrm{P}\tilde{Z}}(v)\right) + h(\tilde{Z}) + \ln\frac{\mathrm{P}_N}{\mathrm{P}},
\end{aligned} \tag{4.53}$$

where $D(p||q)$ is the Kullback-Leibler divergence between two PDFs p and q. Using equation (4.53), the input cost constraint can be rewritten in a different form:

$$\mathrm{E}\left[D\left(p_{\mathrm{P}_N\tilde{Z}}(v-X)||p_{\mathrm{P}\tilde{Z}}(v)\right)\right] = \ln\frac{\mathrm{P}}{\mathrm{P}_N}.$$

We note that the capacity problem of the stable channel (4.47) under the input cost constraint (4.52) is a generalization to the well known AWGN channel under the average power constraint introduced by Shannon [54] and the additive independent Cauchy channel under a logarithmic constraint presented in Section 4.4.1.

Finally, The generic cost function $\mathcal{C}(x)$ presented in this section is $\Theta(x^2)$ when $\alpha = 2$ and indeed is $\Theta\left(\ln|x|\right)$ otherwise. In fact, if one considers equation (4.51) when $0 < \alpha < 2$, by virtue of the fact that $\ln p_{\tilde{Z}}(x) = \Theta(\ln|x|)$ it can be shown that $\mathcal{C}(x) = \Theta\left(\ln|x|\right)$ by using the same methodology as done in Section 3.3.

4.4.4 Extensions and Insights

According to equation (4.23), the power measure P_X is related to a choice of $\tilde{Z}$ or equivalently a choice of $0 < \alpha \leq 2$ and P_X as previously mentioned can be looked at as the relative power of X with respect to that of $\tilde{Z}$. Naturally one would ask the following: In a specific scenario, what value of alpha is more suitable? An answer to this question is given when considering, for example, an additive noise channel $Y = X + N$. In fact, in most communications' applications, the quantity of interest for a system engineer is the received signal or the output Y as it generally represents the quantity that will undergo further processing in order to retrieve the useful information. In addition, the noise variable N imposed by the channel represents another important variable since relevant quantities and performance measures are computed function of the relative power between the output signal and the noise, a quantity that is commonly referred to as the output SNR. Moreover, the output Y is shaped by the noise N, hence it has "similar" characteristics to those of N (for example, an infinite variance N will always induce an infinite variance Y). This is to say, that in the context of an additive stable noise channel, it would seem natural to measure the power of the different signals with respect to a reference stable variable whose power evaluates to unity. Hence the choice of α and then $\tilde{Z}$ becomes straightforward depending on the stable noise characteristic exponent α.

A natural extension is to generalize the adoption of P_X or a specific $\tilde{Z}$, to cases where the noise is not necessarily stable however falls in the domain of normal attraction $\mathbb{D}_\alpha$ of the stable variables. Depending on the value of α, the choice of $\tilde{Z}$ and therefore P_X is adopted. This includes any noise variable having a finite second moment since it belongs to $\mathbb{D}_2$ and P_X is equal to the second moment in this case. Also, it includes all impulsive noise variables whose tail behaviour is $\Theta\left(\frac{1}{|x|^{1+\alpha}}\right)$, $0 < \alpha < 2$ where in this case P_X is evaluated according to the corresponding value of α using equation (4.23).

4.5 Related Publications

At last, we acknowledge that the results of Section 4.4.1 were presented as a conference paper [91].

Chapter 5

Generalized Information Theoretic Inequalities

5.1 Background

Information inequalities have been investigated since the foundation of information theory. Two such important ones are due to Shannon [54]:

- The first one is an upper bound on the (differential) entropy of RVs having a finite second moment by virtue of the fact that Gaussian distributions maximize entropy under a second moment constraint: for any RVs X and Z having respectively finite variances σ_X^2 and σ_Z^2, we have

$$h(X + Z) \leq \frac{1}{2} \ln 2\pi e \left(\sigma_X^2 + \sigma_Z^2\right). \tag{5.1}$$

- The second one is a lower bound on the entropy of independent sums of RVs and commonly known as the Entropy Power Inequality (EPI). The EPI states that given two real independent RVs X, Z such that $h(X)$, $h(Z)$ and

$h(X+Z)$ exist, then (Corollary 3, [53])

$$N(X+Z) \geq N(X) + N(Z), \tag{5.2}$$

where $N(X)$ is the *entropy power* of X and is equal to

$$N(X) = \frac{1}{2\pi e} e^{2h(X)}.$$

While Shannon proposed equation (5.2) and proved it locally around the normal distribution, Stam [76] was the first to prove this result in general followed by Blachman [77] in what is considered to be a simplified proof. The proof was done via the usage of two information identities:

1- The Fisher Information Inequality (FII): Let X and Z be two independent RVs such that the respective *Fisher informations* $J(X)$ and $J(Z)$ exist. Then

$$\frac{1}{J(X+Z)} \geq \frac{1}{J(X)} + \frac{1}{J(Z)}. \tag{5.3}$$

The Fisher information $J(Y)$ of a RV Y having a PDF $p(y)$ is defined as:

$$J(Y) = \int_{-\infty}^{+\infty} \frac{1}{p(y)} p'^2(y)\, dy,$$

whenever the derivative and the integral exit.

2- The de Bruijn's identity: For any $\epsilon > 0$,

$$\frac{d}{d\epsilon} h(X + \sqrt{\epsilon} Z) = \frac{\sigma^2}{2} J(X + \sqrt{\epsilon} Z), \tag{5.4}$$

where Z is a Gaussian RV with mean zero and variance σ^2 independent of X. Rioul proved that the de Bruijn's identity holds at $\epsilon = 0^+$ for any finite-variance RV Z (Proposition 7, p. 39, [82]).

The remarkable similarity between equations (5.2) and (5.3) was pointed out in Stam's paper [76] who in addition, related the entropy power and the Fisher information by an "uncertainty principle-type" relation:

$$N(X)J(X) \geq 1, \tag{5.5}$$

which is commonly known as the Isoperimetric Inequality for Entropies (IIE) [92, Theorem 16]. Interestingly, equality holds in equation (5.5) whenever X is Gaussian distributed and in equations (5.1)–(5.3) whenever X and Z are independent Gaussian. As it can be noticed, the previously cited inequalities revolve around Gaussian variables and some of them are related to variables with finite variances (equation (5.1) for example).

In this chapter, we generalize these information theoretic inequalities that are based on the Gaussian setting to generic ones in the stable setting ($0 < \alpha \leq 2$) and which coincides with the regular identities in the Gaussian setup ($\alpha = 2$). Furthermore, we find new identities that were previously unknown even in the Gaussian case. We use convolutions along small perturbations to upper bound some relevant information theoretic quantities as done in [93] where some moment constraints were imposed on X which is not the case here.

We start by considering an alternative formulation of Fisher information that may be more relevant than $J(X)$ when dealing with RVs corrupted by additive noise of infinite second moment; In essence, our starting point is one where –in a similar fashion to the Gaussian case– we enforce a generalized de-Bruijn identity to hold: motivated by the fact that the derivative of the differential entropy with respect to small variations in the direction of a Gaussian variable is a scaled $J(\cdot)$, we propose in this work to define a new notion of Fisher information that we call "Fisher information of order α" as a derivative of differential entropy in the direction of infinitesimal perturbations along stable variables. Next, we derive an integral expression for the new quantity that is a generalization of the well-known expression of the Fisher information. Recently, in a non-published work [94], Toscani proposed a definition of a "fractional Fisher information" in a relative manner with respect to alpha-stable variables and showed that the relative fractional information quantity has rather interesting properties and satisfies a Fisher information inequality type of identities. The approach adopted by Toscani as well the proposed expressions, the proved identities and their significance are different than those proposed in this dissertation though similar tools such as fractional derivatives and Riesz potentials figure in both definitions of the new Fisher informations. Both works were developed independently.

The new expression of the "Fisher information of order α", when restricted to the range $1 < \alpha \leq 2$, is found to satisfy a data processing inequality and a Generalized FII (GFII). Then, we use the GFII and the generalized de-Bruijn to provide an upper bound on the differential entropy of the independent sum of two RVs where one of them is stable distributed providing, among others, new implications in the Gaussian setting.

5.2 Fisher Information of Order α: A Generalized Information Measure

Definition 7 (Fisher information of order α)**.** Let X be a finite differential entropy RV and N an independent standard symmetric stable variable, $N \sim \mathcal{S}(\alpha, 1)$, $0 < \alpha \leq 2$. We define the "Fisher information of order α" $J_\alpha(X)$ as follows:

$$J_\alpha(X) = \lim_{t \to 0^+} \frac{h\left(X + \sqrt[\alpha]{t}\, N\right) - h(X)}{t} \tag{5.6}$$

whenever the limit exists.

For a d-dimensional random vector $\mathbf{X} = (X_1, \cdots, X_d)$, $J_\alpha(\mathbf{X})$ is defined as in (5.6) where $\mathbf{N} = (N_1, \cdots, N_d)$ is a

- standard sub-Gaussian SαS vector whenever $\alpha \neq 2$ (refer to Appendix G).

- Gaussian vector of Independent and Identically Distributed (IID) components with mean zero and variance 2 when $\alpha = 2$.

Few observations may be readily made about $J_\alpha(\mathbf{X})$:

1- By definition, $J_\alpha(\mathbf{X})$ represents the rate of variation of $h(\mathbf{X})$ under a small disturbance in the direction of a standard sub-Gaussian SαS vector. It represents the limit of positive quantities and therefore $J_\alpha(\mathbf{X}) \geq 0$.

2- When the stable noise $\mathbf{N}$ is Gaussian, i.e. $\alpha = 2$ and $J_2(\mathbf{X})$ coincides with the usual notion of Fisher information.

3- Let $\mathbf{c} \in \mathbb{R}^d$, then $J_\alpha(\mathbf{X} + \mathbf{c}) = J_\alpha(\mathbf{X})$. This follows directly from the definition and from the translation invariant property of the differential entropy.

4- If $\mathbf{X} \sim \boldsymbol{\mathcal{S}}(\alpha, \gamma)$ then $J_\alpha(\mathbf{X}) = \frac{d}{\alpha}\frac{1}{\gamma^\alpha}$ nats. Indeed, let $\mathbf{X} \sim \boldsymbol{\mathcal{S}}(\alpha, \gamma)$ then $\mathbf{X} + \sqrt[\alpha]{\epsilon}\mathbf{N} \sim \boldsymbol{\mathcal{S}}\left(\alpha, \sqrt[\alpha]{\gamma^\alpha + \epsilon}\right)$ and

$$\begin{aligned}
J_\alpha(\mathbf{X}) &= \lim_{\epsilon \to 0} \frac{h\left(\mathbf{X} + \sqrt[\alpha]{\epsilon}\mathbf{N}\right) - h(\mathbf{X})}{\epsilon} \\
&= \lim_{\epsilon \to 0} \frac{h\left(\sqrt[\alpha]{\gamma^\alpha + \epsilon}\mathbf{N}\right) - h(\gamma\mathbf{N})}{\epsilon} \\
&= \lim_{\epsilon \to 0} \frac{h\left(\mathbf{N}\right) + d\ln\left(\sqrt[\alpha]{1 + \frac{\epsilon}{\gamma^\alpha}}\right) - h(\mathbf{N})}{\epsilon} \\
&= \frac{d}{\alpha}\frac{1}{\gamma^\alpha} \text{ nats.}
\end{aligned}$$

This result comes in accordance with the fact that $J_2(\mathbf{X}) = J(\mathbf{X}) = \frac{d}{\sigma^2}$ whenever $\mathbf{X} \sim \boldsymbol{\mathcal{N}}(0; \sigma^2)$ is Gaussian. This is true since in this case $\alpha = 2$ and for a Gaussian variable $\gamma^2 = \frac{\sigma^2}{2}$.

5- Let $a > 0$, then $J_\alpha(a\mathbf{X}) = \frac{1}{a^\alpha} J_\alpha(\mathbf{X})$. In fact,

$$\begin{aligned} J_\alpha(a\mathbf{X}) &= \lim_{\epsilon \to 0} \frac{h\left(a\mathbf{X} + \sqrt[\alpha]{\epsilon}\mathbf{N}\right) - h(a\mathbf{X})}{\epsilon} \\ &= \lim_{\epsilon \to 0} \frac{h\left(\mathbf{X} + \sqrt[\alpha]{\frac{\epsilon}{a^\alpha}}\mathbf{N}\right) + d\ln a - h(\mathbf{X}) - d\ln a}{\epsilon} \\ &= \frac{1}{a^\alpha} \lim_{\epsilon \to 0} \frac{h\left(\mathbf{X} + \sqrt[\alpha]{\frac{\epsilon}{a^\alpha}}\mathbf{N}\right) - h(\mathbf{X})}{\frac{\epsilon}{a^\alpha}} \\ &= \frac{1}{a^\alpha} J_\alpha(\mathbf{X}). \end{aligned}$$

6- Let $\mathbf{Z}$ be a random vector independent of $\mathbf{X}$. Then

$$\begin{aligned} J_\alpha(\mathbf{X} + \mathbf{Z}) &= \lim_{t \to 0} \frac{h\left(\mathbf{X} + \mathbf{Z} + \sqrt[\alpha]{t}\mathbf{N}\right) - h(\mathbf{X} + \mathbf{Z})}{t} \\ &= \lim_{t \to 0} \frac{I(\mathbf{X} + \mathbf{Z} + \sqrt[\alpha]{t}\mathbf{N}; \mathbf{N})}{t} \\ &\leq \lim_{t \to 0} \frac{I(\mathbf{X} + \sqrt[\alpha]{t}\mathbf{N}; \mathbf{N})}{t} = J_\alpha(\mathbf{X}), \end{aligned}$$

where the inequality is due to the fact that $\mathbf{N} - \mathbf{X} + \sqrt[\alpha]{t}\mathbf{N} - \mathbf{X} + \mathbf{Z} + \sqrt[\alpha]{t}\mathbf{N}$ is a Markov chain.

7- $J_\alpha(\cdot)$ is sub-additive for independent random vectors: Let $\mathbf{X} = (X_1, \cdots, X_d)$ be a collection of d independent RVs having Fisher informations $\{J_\alpha(X_i)\}_{i=1}^d$, then $J_\alpha(\mathbf{X}) = J_\alpha(X_1, \cdots, X_d) \leq \sum_{i=1}^d J_\alpha(X_i)$, because $h(Z_1, \cdots, Z_d) \leq \sum_{i=1}^d h(Z_i)$ with equality whenever $\{Z_i\}_{i=1}^d$ are independent. We already know that $J_2(\cdot)$ is additive and it will be later shown in this chapter that $J_\alpha(\cdot)$ is in fact additive.

Due to the above, one may consider $J_\alpha(\mathbf{X})$, $0 < \alpha \leq 2$ as a measure of information. A single random vector $\mathbf{X}$ might hence have different information measures which represent from an estimation theory perspective a reasonable fact since the statistics of the additive noise $\mathbf{N}$ affects the estimation of $\mathbf{X}$ based on the observation of $\mathbf{X} + \mathbf{N}$. From this perspective, the original Fisher information would seem suitable when the adopted noise model is Gaussian or when we are restricting the RV to have a finite second moment. We find in what follows an expression of $J_\alpha(\mathbf{X})$.

5.3 Main Results

We limit our study to random vectors $\mathbf{X} \in \mathcal{V}$ defined as:

$$\mathcal{V} = \Big\{\text{Absolutely continuous RVs } \mathbf{U} : p_{\mathbf{U}}(\mathbf{u}) > 0, \\ h(\mathbf{U}) \text{ is finite } \& \int \ln\left(1 + \|\mathbf{U}\|\right) p_{\mathbf{U}}(\mathbf{u})\, d\mathbf{u} \text{ is finite } \Big\}.$$

5.3.1 Concavity of Differential Entropy

Let $\mathbf{U}$ be an infinitely divisible random vector with characteristic function $\phi_{\mathbf{U}}(\boldsymbol{\omega})$ For each real $t \geq 0$, denote by $F_t(\cdot)$ the unique probability distribution (Theorem 2.3.9, p. 65, [95]) with characteristic function:

$$\phi_t(\boldsymbol{\omega}) = e^{t \ln \phi_{\mathbf{U}}(\boldsymbol{\omega})}, \tag{5.7}$$

where $\ln(\cdot)$ is the principal branch of the logarithm. For the rest of this section, we denote by $\mathbf{U}_t$ a random vector with characteristic function $\phi_t(\boldsymbol{\omega})$ as defined in equation (5.7). Note that $\mathbf{U}_0$ is deterministically equal to 0 (*i.e.*, distributed according to the Dirac delta distribution) and $\mathbf{U}_1$ is distributed according to $\mathbf{U}$. The family of probability distributions $\{F_t(\cdot)\}_{t\geq 0}$ forms a continuous convolution semi-group in the space of probability measures on $\mathbb{R}^d$ (see Definition 2.3.8 and Theorem 2.3.9, [95]) and hence one can write:

$$\mathbf{U}_s + \mathbf{U}_t = \mathbf{U}_{s+t} \qquad \forall s, t \geq 0,$$

where $\mathbf{U}_s$ and $\mathbf{U}_t$ are independent.

Lemma 8 (Concavity of Differential Entropy)**.** *Let* $\mathbf{U}$ *be an infinitely divisible random vector and* $\{\mathbf{U}_t\}_{t\geq 0}$ *an associated family of random vectors distributed according to equation (5.7) and independent of* $\mathbf{X}$*. The differential entropy* $h(\mathbf{X}+\mathbf{U}_t)$ *is a concave function in* $t \geq 0$*.*

In the case of a stable-distributed $\mathbf{U}$, the family $\{\mathbf{U}_t\}_{t\geq 0}$ has the same distribution as $\left\{\sqrt[\alpha]{t}\mathbf{U}\right\}_t$, $0 < \alpha \leq 2$. When $\alpha = 2$, it is already known that the entropy (and actually even the entropy power) of $(\mathbf{X} + \sqrt{t}\mathbf{U})$ is concave in t ((Section VII, p. 51, [82]) and [96]).

Proof. We start by noting that $h(\mathbf{X} + \mathbf{U}_t)$ is non-decreasing in t. For $0 \leq s < t$,

$$h(\mathbf{X} + \mathbf{U}_t) = h(\mathbf{X} + \mathbf{U}_s + \mathbf{U}_{t-s}) \geq h(\mathbf{X} + \mathbf{U}_s),$$

where $\mathbf{U}_t$, $\mathbf{U}_s$ and $\mathbf{U}_{t-s}$ are three independent instances of random vectors in the family $\{\mathbf{U}_t\}_{t\geq 0}$. Next we show that $h(\mathbf{X} + \mathbf{U}_t)$ is midpoint concave: Let $\mathbf{U}_t$, $\mathbf{U}_s$, $\mathbf{U}_{(t+s)/2}$ and $\mathbf{U}_{(t-s)/2}$ be independent random vectors in the family $\{\mathbf{U}_t\}_{t\geq 0}$. For $0 \leq s < t$,

$$\begin{aligned} h(\mathbf{X} + \mathbf{U}_t) - h(\mathbf{X} + \mathbf{U}_{(t+s)/2}) &= h(\mathbf{X} + \mathbf{U}_{(t+s)/2} + \mathbf{U}_{(t-s)/2}) - h(\mathbf{X} + \mathbf{U}_{(t+s)/2}) \\ &= I(\mathbf{X} + \mathbf{U}_{(t+s)/2} + \mathbf{U}_{(t-s)/2}; \mathbf{U}_{(t-s)/2}) \qquad (5.8) \\ &\leq I(\mathbf{X} + \mathbf{U}_s + \mathbf{U}_{(t-s)/2}; \mathbf{U}_{(t-s)/2}) \qquad (5.9) \\ &= h(\mathbf{X} + \mathbf{U}_{(t+s)/2}) - h(\mathbf{X} + \mathbf{U}_s) \end{aligned}$$

where equation (5.8) is the definition of the mutual information and equation (5.9) is the application of the data processing inequality to the Markov chain $\mathbf{U}_{(t-s)/2} - (\mathbf{X} + \mathbf{U}_s + \mathbf{U}_{(t-s)/2}) - (\mathbf{X} + \mathbf{U}_{(t+s)/2} + \mathbf{U}_{(t-s)/2})$. Therefore,

$$h(\mathbf{X} + \mathbf{U}_{(t+s)/2}) \geq \frac{1}{2}\big[h(\mathbf{X} + \mathbf{U}_t) + h(\mathbf{X} + \mathbf{U}_s)\big],$$

and the function is midpoint concave for $t \geq 0$. Since the function is non-decreasing, it is Lebesgue measurable and midpoint concavity guarantees its concavity. $\square$

An interesting implication of Lemma 8 is that $h(\mathbf{X}+\mathbf{U}_t)$ as a function of t is below any of its tangents. Particularly,

$$h(\mathbf{X}+\mathbf{U}_t) \leq h(\mathbf{X}) + t\frac{dh(\mathbf{X}+\mathbf{U}_t)}{dt}\Big|_{t=0}. \tag{5.10}$$

5.3.2 An expression of $J_\alpha(\cdot)$

Lemma 9 (An Expression of the Fisher Information of Order α)**.** *For* $0<\alpha\leq 2$ *and* $\gamma\in\mathbb{R}$*, let* $\mathbf{N}\sim\boldsymbol{\mathcal{S}}(\alpha,\gamma)$ *be a SαS vector (as defined in Appendix G). Consider an independent* $\mathbf{X}\in\mathcal{V}$ *with characteristic function* $\phi_{\mathbf{X}}(\boldsymbol{\omega})$ *such that*

$$\left\{\ln p_{\mathbf{X}+\sqrt[\alpha]{t}\mathbf{N}}(\mathbf{x})\mathcal{F}^{-1}\left[\|\boldsymbol{\omega}\|^\alpha\phi_{\mathbf{X}+\sqrt[\alpha]{t}\mathbf{N}}(-\boldsymbol{\omega})\right](\mathbf{x})\right\}_{t\in[0,\epsilon)}$$

are assumed to be uniformly bounded in t *by an integrable function of* $\mathbf{x}$*. Its Fisher information of order* α *is*

$$J_\alpha(\mathbf{X}) = \int \ln p_{\mathbf{X}}(\mathbf{x})\,\mathcal{F}^{-1}\left[\|\boldsymbol{\omega}\|^\alpha\phi_{\mathbf{X}}(-\boldsymbol{\omega})\right](\mathbf{x})\,d\mathbf{x}.^{1} \tag{5.11}$$

Proof. We first note that $h(\mathbf{X}+\sqrt[\alpha]{t}\mathbf{N})$ exists and is finite since $\sqrt[\alpha]{t}\,\mathbf{N}\sim\boldsymbol{\mathcal{S}}(\alpha,t\gamma)$ has a bounded PDF and $\mathbf{X}\in\mathcal{V}$ [82, Proposition 1]. Also $h(\mathbf{X}+\sqrt[\alpha]{t}\mathbf{N})$ is concave in t by the result of Lemma 8. Therefore it is everywhere left and right differentiable and a.e differentiable. Hence $\frac{d}{dt}h(\mathbf{X}+\sqrt[\alpha]{t}\mathbf{N})$ exists a.e. in t and $\frac{d}{dt}h(\mathbf{X}+\sqrt[\alpha]{t}\mathbf{N})\Big|_{t=0^+}$ exists. Now, let $t\geq\eta\geq 0$ and denote $\mathbf{X}_t=\mathbf{X}+\sqrt[\alpha]{t}\mathbf{N}$ with characteristic function

$$\begin{aligned}\phi_{\mathbf{X}_t}(\boldsymbol{\omega}) &= \phi_{\mathbf{X}}(\boldsymbol{\omega})\,e^{-t\gamma^\alpha\|\boldsymbol{\omega}\|^\alpha} = \phi_{\mathbf{X}_\eta}(\boldsymbol{\omega})\,e^{-(t-\eta)\gamma^\alpha\|\boldsymbol{\omega}\|^\alpha}\\ &= \phi_{\mathbf{X}_\eta}(\boldsymbol{\omega}) - (t-\eta)\gamma^\alpha\|\boldsymbol{\omega}\|^\alpha\phi_{\mathbf{X}_\eta}(\boldsymbol{\omega}) + o(t-\eta),\end{aligned}$$

[1] $\mathcal{F}^{-1}(\cdot)$ *denotes the inverse distributional Fourier transform.*

By the linearity of the inverse Fourier transform, we obtain

$$p_{\mathbf{X}_t}(\mathbf{x}) = p_{\mathbf{X}_\eta}(\mathbf{x}) - (t-\eta)\gamma^\alpha \mathcal{F}^{-1}\big[\|\boldsymbol{\omega}\|^\alpha \phi_{\mathbf{X}_\eta}(-\boldsymbol{\omega})\big](\mathbf{x}) + o(t-\eta), \tag{5.12}$$

which is valid since the inverse distributional Fourier transform $\mathcal{F}^{-1}\big[\|\boldsymbol{\omega}\|^{n\alpha}\phi_{\mathbf{X}_\eta}(\boldsymbol{\omega})\big]$ exists for all $m \geq 1$ because $\|\boldsymbol{\omega}\|^{m\alpha}\phi_{\mathbf{X}_\eta}(\boldsymbol{\omega})$ is a tempered function by virtue of the fact that $\phi_{\mathbf{X}_\eta}(\boldsymbol{\omega})$ is an L^1-characteristic function and hence is in $\mathcal{L}^\infty(\mathbb{R}^d)$. Equation (5.12) implies that

$$\left.\frac{d\,p_{\mathbf{X}_\tau}(\mathbf{x})}{d\tau}\right|_{\tau=\eta} = -\gamma^\alpha \mathcal{F}^{-1}\big[\|\boldsymbol{\omega}\|^\alpha \phi_{\mathbf{X}_\eta}(-\boldsymbol{\omega})\big](\mathbf{x}),$$

and by the MVT: for some $0 \leq b(t) \leq t$,

$$\begin{aligned}
\frac{h(\mathbf{X}_t) - h(\mathbf{X})}{t} &= -\int_{\mathbb{R}^d} \frac{p_{\mathbf{X}_t}(\mathbf{x})\,\ln p_{\mathbf{X}_t}(\mathbf{x}) - p_{\mathbf{X}}(\mathbf{x})\,\ln p_{\mathbf{X}}(\mathbf{x})}{t}\,d\mathbf{x} \\
&= -\int_{\mathbb{R}^d} \Big[1 + \ln p_{\mathbf{X}_{b(t)}}(\mathbf{x})\Big] \left.\frac{d\,p_{\mathbf{X}_\tau}(\mathbf{x})}{d\tau}\right|_{\tau=b(t)} d\mathbf{x} \\
&= \gamma^\alpha \int_{\mathbb{R}^d} \Big[1 + \ln p_{\mathbf{X}_{b(t)}}(\mathbf{x})\Big]\,\mathcal{F}^{-1}\big[\|\boldsymbol{\omega}\|^\alpha \phi_{\mathbf{X}_{b(t)}}(-\boldsymbol{\omega})\big](\mathbf{x}) \\
&= \gamma^\alpha \int_{\mathbb{R}^d} \ln p_{\mathbf{X}_{b(t)}}(\mathbf{x})\mathcal{F}^{-1}\big[\|\boldsymbol{\omega}\|^\alpha \phi_{\mathbf{X}_{b(t)}}(-\boldsymbol{\omega})\big](\mathbf{x})\,d\mathbf{x},
\end{aligned}$$

which is true since

$$\int \mathcal{F}^{-1}\big[\|\boldsymbol{\omega}\|^\alpha \phi_{\mathbf{X}_{b(t)}}(-\boldsymbol{\omega})\big](\mathbf{x})\,d\mathbf{x} = \int \delta(\boldsymbol{\omega})\,\|\boldsymbol{\omega}\|^\alpha \phi_{\mathbf{X}_{b(t)}}(-\boldsymbol{\omega})\,d\boldsymbol{\omega} = 0.$$

The imposed conditions insure that Lebesgue's DCT holds and the limit may be passed inside the integral and

$$J_\alpha(\mathbf{X}) = \int_{\mathbb{R}^d} \ln p_{\mathbf{X}}(\mathbf{x})\mathcal{F}^{-1}\big[\|\boldsymbol{\omega}\|^\alpha \phi_{\mathbf{X}}(-\boldsymbol{\omega})\big](\mathbf{x})\,d\mathbf{x},$$

provided that $[\|\boldsymbol{\omega}\|^\alpha \phi_{\mathbf{X}}(-\boldsymbol{\omega})] \in \mathrm{L}^1(\mathbb{R}^d)$. ☐

We note that, whenever $\alpha = 2$, equation (5.11) gives the regular expression of the Fisher information. In fact, in the scalar case

$$\begin{aligned} J(X) = J_2(X) &= \int \ln p_X(x)\, \mathcal{F}^{-1}\big[|\omega|^2 \phi_X(-\omega)\big](x)\, dx \\ &= -\int \ln p_X(x)\, \frac{d^2}{dx^2}\, p_X(x)\, dx, \end{aligned}$$

where the last equality is valid as long as $\ln p_X(x)\, \frac{d}{dx}\, p_X(x)|_{-\infty}^{+\infty}$ vanishes. In the d-dimensional case, $J_2(\mathbf{X})$ is also consistent with the regular definition of the Fisher information being the trace of the Fisher information matrix. The sufficient condition listed in the statement of the lemma, is a technical condition involving "fractional" derivatives of the PDF $p_{\mathbf{X}}(\mathbf{x})$. Whenever $\alpha = 2$, this condition boils down to similar type of conditions imposed by Kullback [97, pages 26-27] to prove the well-known result relating the second derivative of the divergence to the Fisher information: a result that implies de Bruijn's identity at zero (see [82]). Similar to the work done by Barron [98] regarding the existence and finiteness of $J_2(\cdot)$, we prove in Appendix F, in the scalar case, that when X is replaced with $X_\eta = X + \sqrt[\alpha]{\eta} N'$ for some $\eta > 0$ where $N' \sim \mathcal{S}(\alpha, \gamma)$, the condition stated in the lemma holds true for any $X \in \mathcal{L} = \big\{ \text{RVs } U : \int \ln(1 + \|U\|)\, dF_U(u) \text{ is finite } \big\}$ and

$$\frac{d}{dt}\, h(X_\eta + \sqrt[\alpha]{t} N)\Big|_{t=0^+} = \gamma^\alpha J_\alpha(X_\eta), \tag{5.13}$$

for $N^{'}$ and N IID. Since $\sqrt[\alpha]{\eta}\, N' + \sqrt[\alpha]{t}\, N$ is distributed according to $\sqrt[\alpha]{\eta + t}\, N$, then equation (5.13) is equivalent to a generalized de Bruijn's identity of the form:

$$\frac{\partial}{\partial \eta}\, h(X_\eta) = \gamma^\alpha J_\alpha(X_\eta), \tag{5.14}$$

for $\eta > 0$ and $X \in \mathcal{L}$ and where $J_\alpha(X_\eta)$ is given by equation (5.11). Note that whenever the conditions in the lemma are satisfied the generalized de Bruijn's identity holds for the vector case. Finally, the generalized de Bruijn's identity holds at 0^+ as well whenever the conditions in the lemma are satisfied.

To compute $J_\alpha(\cdot)$, we implement the fast Fourier transform theorem using *Matlab* by following a similar strategy as in [50]. We plot in Figure 5-1 the evaluation of $J_\alpha(\cdot)$ for a collection of alpha-stable variables $X \sim \mathcal{S}\left(r, (r)^{-\frac{1}{r}}\right)$ parametrised by the characteristic exponent r. It is observed that as the value of r increases, $J_\alpha(X)$ increases. Furthermore for fixed r, $J_\alpha(X)$ decreases with α.

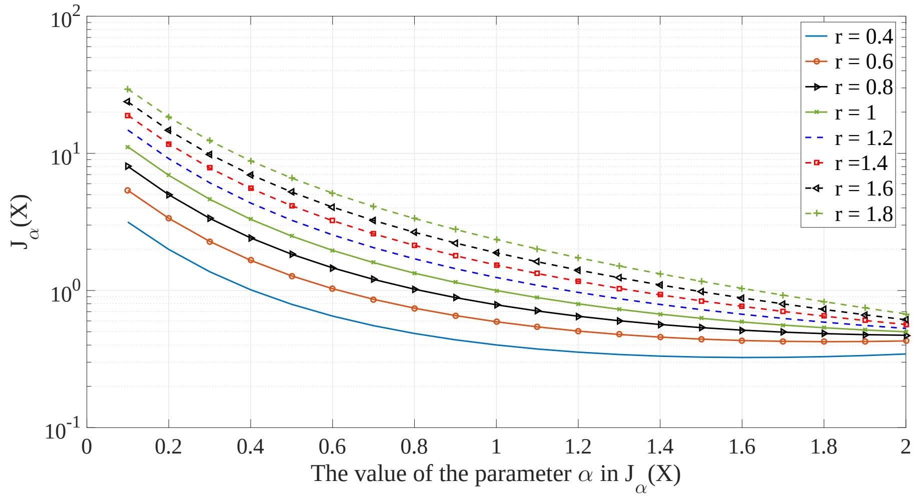

Figure 5-1: Evaluation of $J_\alpha(X)$ for $X \sim \mathcal{S}\left(r, (r)^{-\frac{1}{r}}\right)$ for different values of α and r .

5.3.3 A Generalized Fisher Information Inequality

The Fisher Information Inequality (FII) is an important identity that relates the Fisher information of the sum of independent RVs to those of the individual variables. It was first proven by Stam [76] and then by Blachman [77]. Both authors deduced the Entropy Power Inequality (EPI) from the FII via de Bruijn's identity. Stam relied on a data processing inequality of the Fisher information in the proof of the FII, a methodology that was later used by Zamir [99] in a more elaborate fashion. Finally, Rioul [82] derived a mutual information inequality, an identity that implies the EPI and by the means of de Bruijn's identity implies the FII.

Data processing inequality for J_α, $1 < \alpha \leq 2$ The data processing inequality asserts that gains could not be achieved when processing information. In terms of mutual information, if the RVs X–Y–Z form a Markov chain, this boils down to saying that [56, p.34 Theorem 2.8.1]:

$$I(Z; X) \leq I(Y; X),$$

with equality if X–Z–Y is also a Markov chain. In [99], Zamir proved an equivalent inequality for the Fisher information in a variable Y about a parameter θ. We follow similar steps and extend the data processing inequality to J_α; An inequality which we will use next to prove the Generalized Fisher Information Inequality (GFII).

Theorem 16 (Data Processing Inequality for the Fisher Information of Order α)**.** *Let θ be a fixed scalar parameter and let $\mathbf{Y}_\theta = \mathbf{Y} + \theta\mathbf{1}$ and $\mathbf{Z}_\theta = \mathbf{Z} + \theta\mathbf{1}$ be two vectors of possibly different dimensions. If θ–$\mathbf{Y}_\theta$–$\mathbf{Z}_\theta$, i.e., the conditional distribution of $\mathbf{Z}_\theta$ given $\mathbf{Y}_\theta$ is independent of θ, then whenever $J_\alpha(\mathbf{Y}_\theta; \theta|\mathbf{Z}_\theta) \geq 0$ we have*

$$J_\alpha(\mathbf{Z}_\theta; \theta) \leq J_\alpha(\mathbf{Y}_\theta; \theta), \quad 1 < \alpha \leq 2 \tag{5.15}$$

where

$$J_\alpha(\mathbf{Y}_\theta; \theta) \quad \hat{=} \quad -E\left[I_{2-\alpha}\left(\frac{d^2}{d\theta^2}\ln p_{\mathbf{Y}_\theta}\right)(\mathbf{Y}_\theta)\right], \tag{5.16}$$

and where

$$J_\alpha(\mathbf{Y}_\theta; \theta|\mathbf{Z}_\theta) = E_{\mathbf{Z}_\theta}\left[J_\alpha\left(\mathbf{Y}_\theta; \theta|\mathbf{Z}_\theta\right)\right]. \tag{5.17}$$

The operator $I_{2-\alpha}(\cdot)$ is the Riesz potential of order $(2-\alpha)$. Moreover, whenever *the* $\{\mathbf{Y}_j\}_j$ *are independent,*

$$J_\alpha(\mathbf{Y}_\theta;\theta) = J_\alpha(\mathbf{Y}). \tag{5.18}$$

Proof. For a definition of the Riesz potential, the reader is referred to Appendix G. We also note that the condition $J_\alpha(\mathbf{Y}_\theta;\theta|\mathbf{Z}_\theta) \geq 0$ is needed since there are no formal guarantees that $J_\alpha(\mathbf{X}_\theta;\theta)$ as defined in equation (5.16) is always non-negative as it is the case for $J_\alpha(\mathbf{X})$. The non-negativity of $J_\alpha(\mathbf{X}_\theta;\theta)$ is guaranteed, for example, whenever θ is a translation parameter and when the $\{X_i\}$'s are independent as it will be shown in the proof of this theorem. Another case when non-negativity is guaranteed is encountered in the coming proof of the GFII. We start by proving equation (5.18). Assume that the Y_j's are independent,

$$\begin{aligned}
J_\alpha(\mathbf{Y}) &= \int \ln p_{\mathbf{Y}}(\mathbf{y})\, \mathcal{F}^{-1}\left[\|\boldsymbol{\omega}\|^{\alpha}\phi_{\mathbf{Y}}(-\boldsymbol{\omega})\right](\mathbf{y})\, d\mathbf{y} \\
&= -\int \ln p_{\mathbf{Y}}(\mathbf{y})\ \triangle \left(\mathcal{F}^{-1}\left[\|\boldsymbol{\omega}\|^{\alpha-2}\phi_{\mathbf{Y}}(-\boldsymbol{\omega})\right](\mathbf{y})\right) d\mathbf{y} && (5.19)\\
&= -\int \triangle\left(\ln p_{\mathbf{Y}}(\mathbf{y})\right)\ I_{2-\alpha}(p_{\mathbf{Y}})(\mathbf{y})\, d\mathbf{y} && (5.20)\\
&= -\int \left[\sum_j \frac{d^2}{dy_j^2} \ln p_{Y_j}(y_j)\right]\ I_{2-\alpha}(p_{\mathbf{Y}})(\mathbf{y})\, d\mathbf{y} && (5.21)\\
&= -\int \left[\sum_j \frac{d^2}{d\theta^2} \ln p_{Y_j}(y_j-\theta)\right]\ I_{2-\alpha}(p_{\mathbf{Y}})(\mathbf{y}-\theta\mathbf{1})\, d\mathbf{y} \\
&= -\int I_{2-\alpha}\left(\frac{d^2}{d\theta^2}\ln p_{\mathbf{Y}_\theta}\right)(\mathbf{y})\, p_{\mathbf{Y}_\theta}(\mathbf{y})\, d\mathbf{y} && (5.22)\\
&= J_\alpha(\mathbf{Y}_\theta;\theta), && (5.23)
\end{aligned}$$

where $\triangle$ denotes the Laplacian operator. Equation (5.19) is due to basic properties of the Fourier transform since $I_{2-\alpha}(p_{\mathbf{Y}})(\mathbf{y}) = \mathcal{F}^{-1}\left[\|\boldsymbol{\omega}\|^{\alpha-2}\phi_{\mathbf{Y}}(-\boldsymbol{\omega})\right](\mathbf{y})$ decays to 0 at "∞". In order to write equation (5.20), we use Green's first identity [100] in the following form: Let $\bigtriangledown$ denotes the gradient operator and $\times$

denotes the dot product. If $\Psi(\cdot)$ and $\Phi(\cdot)$ are real valued functions on $\mathbb{R}^n$, then

$$\begin{aligned}\int_{\mathbb{R}^n} \Psi(\mathbf{y}) \bigtriangleup \Phi(\mathbf{y})\, d\mathbf{y} &= -\int_{\mathbb{R}^n} \bigtriangledown\Psi(\mathbf{y}) \times \bigtriangledown\Phi(\mathbf{y})\, d\mathbf{y} \\ &+ \lim_{R\to+\infty} \int_{\|\mathbf{y}\|=R} \Psi(\mathbf{y}) \bigtriangledown \Phi(\mathbf{y}) \times \mathbf{n} dS(\mathbf{y}),\end{aligned}$$

where $\mathbf{n}$ is the outward vector orthogonal to the surface of the sphere $\|\mathbf{y}\| = R$ in $\mathbb{R}^n$. As long as:

$$\begin{aligned}&\lim_{R\to+\infty} \int_{\|\mathbf{y}\|=R} \ln p_{\mathbf{Y}}(\mathbf{y}) \bigtriangledown I_{2-\alpha}(p_{\mathbf{Y}})(\mathbf{y}) \times \mathbf{n} dS(\mathbf{y}) = 0 \\ &\quad \text{and} \\ &\lim_{R\to+\infty} \int_{\|\mathbf{y}\|=R} I_{2-\alpha}(p_{\mathbf{Y}})(\mathbf{y}) \bigtriangledown \ln p_{\mathbf{Y}}(\mathbf{y}) \times \mathbf{n} dS(\mathbf{y}) = 0,\end{aligned}$$

applying twice Green's theorem justifies equation (5.20). Equation (5.21) is true by virtue of the independence of the Y_j's and equation (5.22) holds true whenever $\left|\frac{d^2}{d\theta^2} \ln p_{\mathbf{Y}_\theta}(\mathbf{y})\right| I_{2-\alpha}(p_{\mathbf{Y}_\theta})(\mathbf{y})$ is integrable (see Appendix G).

Now, consider

$$J_\alpha(\mathbf{Y}_\theta, \mathbf{Z}_\theta; \theta) = -\mathsf{E}_{\mathbf{Y},\mathbf{Z}} \left[I_{2-\alpha} \left(\frac{d^2}{d\theta^2} (\ln p_{\mathbf{Y}_\theta,\mathbf{Z}_\theta}) \right) (\mathbf{Y}_\theta, \mathbf{Z}_\theta) \right].$$

We have

$$\ln p_{\mathbf{Y}_\theta,\mathbf{Z}_\theta}(\mathbf{y}, \mathbf{z}; \theta) = \ln p_{\mathbf{Z}_\theta}(\mathbf{z}; \theta) + \ln p_{\mathbf{Y}_\theta|\mathbf{Z}_\theta}(\mathbf{y}; \theta|\mathbf{z}),$$

which yields

$$\begin{aligned}J_\alpha(\mathbf{Y}_\theta, \mathbf{Z}_\theta; \theta) &= J_\alpha(\mathbf{Z}_\theta; \theta) + J_\alpha(\mathbf{Y}_\theta; \theta|\mathbf{Z}_\theta) \quad (5.24) \\ &\geq J_\alpha(\mathbf{Z}_\theta; \theta). \quad (5.25)\end{aligned}$$

Equation (5.24) is due to the linearity property of the derivative, the Riesz potential [101] and the expectation operator. Equation (5.25) is justified $J_\alpha(\mathbf{Y}_\theta; \theta|\mathbf{Z}_\theta) \geq$

0 by assumption. Equality holds iff $J_\alpha(\mathbf{Y}_\theta;\theta|\mathbf{Z}_\theta) = 0$ which is true if θ–$\mathbf{Z}_\theta$–$\mathbf{Y}_\theta$ forms a Markov chain. On the other hand, one can write

$$J_\alpha(\mathbf{Y}_\theta, \mathbf{Z}_\theta;\theta) = J_\alpha(\mathbf{Y}_\theta;\theta), \tag{5.26}$$

since $\ln p_{\mathbf{Z}_\theta|\mathbf{Y}_\theta}(\cdot|\mathbf{y})$ is independent of θ by virtue of the fact that $\mathbf{Z}_\theta$ is conditionally independent of θ given $\mathbf{Y}_\theta$. Equations (5.25) and (5.26) give the required result.

□

Before proceeding with the proof of the GFII, we make a comment on the additivity of $J_\alpha(\mathbf{Y})$ when $\mathbf{Y}$ has independent components. As mentioned in property 7 of $J_\alpha(\cdot)$, the additivity does hold. This follows directly from eqaution (5.21). In fact,

$$\begin{aligned}
J_\alpha(\mathbf{Y}) &= -\int \left[\sum_j \frac{d^2}{dy_j^2}\ln p_{Y_j}(y_j)\right] I_{2-\alpha}(p_{\mathbf{Y}})(\mathbf{y})\, d\mathbf{y} \\
&= -\int I_{2-\alpha}\left(\sum_j \frac{d^2}{dy_j^2}\ln p_{Y_j}\right)(y_j)\, p_{\mathbf{Y}}(\mathbf{y})\, d\mathbf{y} \\
&= -\sum_j \int I_{2-\alpha}\left(\frac{d^2}{dy_j^2}\ln p_{Y_j}\right)(y_j)\, p_{\mathbf{Y}}(\mathbf{y})\, d\mathbf{y} \\
&= -\sum_j \int I_{2-\alpha}\left(\frac{d^2}{dy_j^2}\ln p_{Y_j}\right)(y_j)\, p_{Y_j}(y_j)\, dy_j \qquad (5.27) \\
&= \sum_j \int \ln p_{Y_j}(y_j) I_{2-\alpha}\left(\frac{d^2}{dy_j^2} p_{Y_j}\right)(y_j)\, dy_j \\
&= \sum_j J_\alpha(Y_j),
\end{aligned}$$

where equation (5.27) is due to the independence of the Y_j's.

Proof of the GFII

Theorem 17 (Generalized Fisher Information Inequality(GFII))**.** *Let* $1 < \alpha \leq 2$

and let Y_1 and Y_2 be two independent RVs, then

$$J_\alpha^{\frac{1}{1-\alpha}}(Y_1+Y_2) \geq J_\alpha^{\frac{1}{1-\alpha}}(Y_1) + J_\alpha^{\frac{1}{1-\alpha}}(Y_2). \tag{5.28}$$

We note that whenever $\alpha = 2$, equation (5.28) boils down to the well-known "classical" FII.

Proof. For the matter of the proof, we make use of the data processing inequality established in Theorem 16. In fact, let ω_1 and $\omega_2 \in \mathbb{R}^{+*}$ be two positive numbers such that $\omega_1 + \omega_2 = 1$. Also let $\epsilon > 0$ and N_1, N_2 be two independent RVs distributed according to $\mathcal{S}(\alpha, 1)$. Then for any $\theta \in \mathbb{R}$ we have

$$\theta - \left(\frac{Y_1}{\omega_1} + \theta + \sqrt[\alpha]{\epsilon}N_1, \frac{Y_2}{\omega_2} + \theta + \sqrt[\alpha]{\epsilon}N_2\right) - \left(Y_1 + Y_2 + \theta + \sqrt[\alpha]{\epsilon}N\right) \tag{5.29}$$

form a Markov chain where $N = \omega_1 N_1 + \omega_2 N_2$ behaves statistically according to $\mathcal{S}_\alpha\left(\alpha, \sqrt[\alpha]{\omega_1^\alpha + \omega_2^\alpha}\right)$. Define $Y_{1\theta} = \frac{Y_1}{\omega_1} + \theta + \sqrt[\alpha]{\epsilon}N_1$, $Y_{2\theta} = \frac{Y_2}{\omega_2} + \theta + \sqrt[\alpha]{\epsilon}N_2$ and $Z_\theta = \omega_1 Y_{1\theta} + \omega_2 Y_{2\theta}$, then by virtue of Theorem 16, we obtain

$$J_\alpha\left(Z_\theta;\theta\right) \quad \leq \quad J_\alpha\left(\left(Y_{1\theta}, Y_{2\theta}\right);\theta\right), \tag{5.30}$$

under the condition that

$$J_\alpha\left(\left(Y_{1\theta}, Y_{2\theta}\right);\theta\middle|Z_\theta\right) \hat{=} \mathsf{E}_{\mathbf{Z}_\theta}\left[J_\alpha\left(\left(Y_{1\theta}, Y_{2\theta}\right);\theta|Z_\theta\right)\right] \geq 0, \tag{5.31}$$

which we prove next. In fact, similarly to equation (5.24), one can write:

$$J_\alpha\left(\left(Y_{1\theta}, Y_{2\theta}\right);\theta\middle|Z_\theta = z\right) \quad = \quad J_\alpha\left(Y_{1\theta};\theta\middle|Z_\theta = z\right) + J_\alpha\left(Y_{2\theta};\theta\middle|\left(Y_{1\theta}, Z_\theta = z\right)\right),$$

where

$$J_\alpha\left(Y_{2\theta};\theta\middle|\left(Y_{1\theta}, Z_\theta = z\right)\right) = \mathrm{E}_{Y_{1\theta}}\left[J_\alpha\left(Y_{2\theta};\theta\middle|\left(Y_{1\theta}, Z_\theta = z\right)\right)\right] = 0, \tag{5.32}$$

since $J_\alpha\left(Y_{2\theta};\theta\middle|(Y_{1\theta}=y_1, Z_\theta=z)\right) = 0$ for every y_1 by virtue of the fact that $p_{Y_{2\theta}|(Y_{1\theta},Z_\theta)}(\cdot)$ is independent of θ. Considering the first term in the RHS of equation (5.32), and by similar arguments that were made in the proof of equation (5.23), one can write:

$$J_\alpha\left(Y_{1\theta};\theta\middle|Z_\theta=z\right) = J_\alpha\left(\frac{Y_1}{\omega_1}+\sqrt[\alpha]{\epsilon}N_1\middle|Z_\theta=z\right) \tag{5.33}$$

which is non-negative by definition (property 1). Then equation (5.31) is satisfied and equation (5.30) holds true. Since θ is a translation parameter and by the fact that $\frac{Y_1}{\omega_1}+\sqrt[\alpha]{\epsilon}N_1$ and $\frac{Y_1}{\omega_2}+\sqrt[\alpha]{\epsilon}N_2$ are independent, equation (5.18) implies that equation (5.30) is equivalent to:

$$\begin{aligned} J_\alpha\left(Y_1+Y_2+\sqrt[\alpha]{\epsilon}N\right) &\leq J_\alpha\left(\frac{Y_1}{\omega_1}+\sqrt[\alpha]{\epsilon}N_1, \frac{Y_2}{\omega_2}+\sqrt[\alpha]{\epsilon}N_2\right) \\ &= J_\alpha\left(\frac{Y_1}{\omega_1}+\sqrt[\alpha]{\epsilon}N_1\right)+J_\alpha\left(\frac{Y_2}{\omega_2}+\sqrt[\alpha]{\epsilon}N_2\right), \end{aligned} \tag{5.34}$$

where we use property 7 of $J_\alpha(\cdot)$ to write equation (5.34) by virtue of the fact that $\left(\frac{Y_1}{\omega_1}+\sqrt[\alpha]{\epsilon}N_1\right)$ and $\left(\frac{Y_2}{\omega_2}+\sqrt[\alpha]{\epsilon}N_2\right)$ are independent. Note that while finding an expression of $J_\alpha(\cdot)$ a condition of uniform boundedness is imposed on some quantity of interest (see the statement of lemma 9) which implies the continuity of $J_\alpha(X+\sqrt[\alpha]{\epsilon}S)$ in $\epsilon \geq 0$ whenever S a symmetric alpha-stable RV. Therefore considering equation (5.34) and taking the limit as $\epsilon \to 0$ yields

$$\begin{aligned} J_\alpha\left(Y_1+Y_2\right) &\leq J_\alpha\left(\frac{Y_1}{\omega_1}\right)+J_\alpha\left(\frac{Y_2}{\omega_2}\right) \\ &\leq \omega_1^\alpha J_\alpha(Y_1)+\omega_2^\alpha J_\alpha(Y_2). \end{aligned} \tag{5.35}$$

where equation (5.35) is due to property 5 of $J_\alpha(\cdot)$. Equation (5.35) holds true for any ω_1 and ω_2 satisfying the conditions of the theorem. The tightest inequality for (5.35) holds for ω_1^* and ω_2^* such that:

$$\begin{aligned}
\omega_1^* &= \operatorname{argmin}_{0\leq\omega_1\leq 1}\{\omega_1^\alpha J_\alpha(Y_1) + (1-\omega_1)^\alpha J_\alpha(Y_2)\} \\
&= \frac{J_\alpha^{\frac{1}{\alpha-1}}(Y_2)}{J_\alpha^{\frac{1}{\alpha-1}}(Y_1) + J_\alpha^{\frac{1}{\alpha-1}}(Y_2)} \\
\omega_2^* = 1 - \omega_1^* &= \frac{J_\alpha^{\frac{1}{\alpha-1}}(Y_1)}{J_\alpha^{\frac{1}{\alpha-1}}(Y_1) + J_\alpha^{\frac{1}{\alpha-1}}(Y_2)},
\end{aligned}$$

for which equation (5.35) gives

$$J_\alpha\left(Y_1 + Y_2\right) \leq \frac{J_\alpha(Y_1)J_\alpha(Y_2)}{\left[J_\alpha^{\frac{1}{\alpha-1}}(Y_1) + J_\alpha^{\frac{1}{\alpha-1}}(Y_2)\right]^{\alpha-1}},$$

which completes the proof of the theorem. □

5.3.4 Upperbounds on the Differential Entropy of Sums Having a Stable Component

An important category of information inequalities consists of finding upper bounds on the entropy of independent sums. When it comes to *discrete* entropy, a bound on the entropy of the sum exists [56]:

$$H(X+Z) \leq H(X) + H(Z). \tag{5.36}$$

In addition, several identities involving discrete entropy of sums were shown in [102, 103] using the Plünnecke-Ruzsa sumset theory and its analogy to Shannon entropy. Except for equation (5.1), that holds for finite variance RVs, the differential entropy inequalities provided in some sense a lower bound on the entropy of sums of independent RVs. Equation (5.36) does not always hold for differential entropies, and unless the variance is finite, if we start with two RVs

X and Z having respectively finite differential entropies $h(X)$ and $h(Z)$, one does not have a clear idea on how much the growth of $h(X + Z)$ will be. The authors in [104] deferred this to the fact that discrete entropy has a functional submodularity property which is not the case for differential entropy. Nevertheless, the authors were able to derive various useful inequalities. Madiman [105] used basic information theoretic relations to prove the submodularity of the entropy of independent sums and found accordingly upper bounds on the discrete and differential entropy of sums. Though, in its general form, the problem of upper bounding the differential entropy of independent sums is not always possible [53, proposition 4], several results are known in particular settings. Cover *et al.* [106] solved the problem of maximizing the differential entropy of the sum of dependent RVs having the same marginal log-concave densities. In [107], Ordentlich found the maximizing probability distribution for the differential entropy of the independent sum of n finitely supported symmetric RVs. For "sufficiently convex" probability distributions, an interesting reverse EPI was proven to hold in (Theorem 1.1, p. 63, [108]). The primary objective of this section is to derive an upper bound on the differential entropy of the sum $X + N$ of two independent RVs

- X has a finite differential entropy $h(X)$.
- N is a symmetric stable variable.

The proof is based on the stability property with the application of the GFII and the generalized de Bruijn's identity. Besides the novelty of the bound itself, it has several implications and can be possibly used for variables X with infinite second moments. Even when the second moment of X is finite, in some cases our bound can be tighter than equation (5.1).

Theorem 18 (Upper bound on the Entropy of Sums having a Stable Component)**.** *Let $Z \sim \mathcal{S}(\alpha, \gamma)$, $1 < \alpha \leq 2$, and let X be independent of Z with finite*

$h(X)$. Then

$$h(X+Z) - h(X) \leq \gamma^\alpha J_\alpha(X)\, {}_2F_1\left(\alpha-1, \alpha-1; \alpha; -(\alpha\gamma^\alpha J_\alpha(X))^{\frac{1}{\alpha-1}}\right),$$

where ${}_2F_1(a,b;c;z)$ is the analytic continuation of the Gauss hypergeometric function on the complex plane with a cut along the real axis from 1 to $+\infty$

For more details on hypergeometric functions, the reader may refer to Appendix G. Theorem 18 provides an upperbound on the entropy of the sum of two variables when one them is stable. As a special case, when $\alpha = 2$, it gives a previously unknown upper bound for Gaussian noise channels. Furthermore, Theorem 18 gives an analytical bound on the change in the transmission rates of the linear stable channel function of an input scaling operation. Let $a > 0$, then

$$\begin{aligned} & h(aX+Z) \\ & \leq h(aX) + \gamma^\alpha J_\alpha(aX)\, {}_2F_1\left(\alpha-1, \alpha-1; \alpha; -(\alpha\gamma^\alpha J_\alpha(aX))^{\frac{1}{\alpha-1}}\right), \\ & = h(X) + \ln a + \left(\frac{\gamma}{a}\right)^\alpha J_\alpha(X)\, {}_2F_1\left(\alpha-1, \alpha-1; \alpha; -(\alpha\left(\frac{\gamma}{a}\right)^\alpha J_\alpha(X))^{\frac{1}{\alpha-1}}\right), \end{aligned}$$

where we used the fact that $h(aX) = h(X) + \ln a$ and $J(aX) = \frac{1}{a^\alpha} J(X)$. Subtracting $h(Z)$ from both sides of the equation gives

$$I(aX+Z;X) - I(X+Z;X) \tag{5.37}$$

$$\leq \ln a + \left(\frac{\gamma}{a}\right)^\alpha J_\alpha(X)\, {}_2F_1\left(\alpha-1, \alpha-1; \alpha; -(\alpha\left(\frac{\gamma}{a}\right)^\alpha J_\alpha(X))^{\frac{1}{\alpha-1}}\right). \tag{5.38}$$

By virtue of the fact that ${}_2F_1\left(\alpha-1, \alpha-1; \alpha; 0\right) = 1$,

$$\lim_{a\to+\infty} \left(\frac{\gamma}{a}\right)^\alpha J_\alpha(X)\, {}_2F_1\left(\alpha-1, \alpha-1; \alpha; -(\alpha\left(\frac{\gamma}{a}\right)^\alpha J_\alpha(X))^{\frac{1}{\alpha-1}}\right) = 0,$$

and it can be seen that the variation in the transmissions rates is bounded by a logarithmically growing function for large values of a. This is a known behavior of the optimal transmission rates that are achieved by Gaussian inputs.

Proof. Using the extended de Bruijn's identiy (equation (5.14)), we write:

$$
\begin{aligned}
h(X+Z) - h(X) &= \int_0^1 \gamma^\alpha J_\alpha(X + \sqrt[\alpha]{\eta} Z)\, d\eta \\
&\leq \gamma^\alpha \int_0^1 \frac{J_\alpha(X) J_\alpha(\sqrt[\alpha]{\eta} Z)}{\left(J_\alpha^{\frac{1}{\alpha-1}}(X) + J_\alpha^{\frac{1}{\alpha-1}}(\sqrt[\alpha]{\eta} Z)\right)^{\alpha-1}}\, d\eta \qquad (5.39) \\
&= \gamma^\alpha \int_0^1 \frac{J_\alpha(X) \frac{1}{\alpha\gamma^\alpha \eta}}{\left(J_\alpha^{\frac{1}{\alpha-1}}(X) + (\frac{1}{\alpha\gamma^\alpha\eta})^{\frac{1}{\alpha-1}}\right)^{\alpha-1}}\, d\eta \qquad (5.40) \\
&= (\alpha-1)\gamma^\alpha J_\alpha(X) \int_0^1 \frac{u^{\alpha-2}}{\left((\alpha\gamma^\alpha J_\alpha(X))^{\frac{1}{\alpha-1}} u + 1\right)^{\alpha-1}} du \\
&= \gamma^\alpha J_\alpha(X)\, {}_2F_1\left(\alpha-1, \alpha-1; \alpha; -(\alpha\gamma^\alpha J_\alpha(X))^{\frac{1}{\alpha-1}}\right). \qquad (5.41)
\end{aligned}
$$

where we use the GFII in order to write equation (5.39) and properties 4 and 5 of $J_\alpha(\cdot)$ to validate equation (5.40). □

Equation (5.41) can be looked at as an upperbound on the entropy of the sum of two variables when one them is stable. We note that by using the identity:

$$
\ln(1+t) = t\, {}_2F_1(1, 1; 2; -t),
$$

equation (5.41) when evaluated for $Z \sim \mathcal{N}(0; \sigma^2)$ and $\alpha = 2$ boils down to the following:

Theorem 19 (Upper bound on the Entropy of Sums having a Gaussian Component)**.** *Let $Z \sim \mathcal{N}(0, \sigma^2)$ and X be an independent RV such that $h(X)$ and $J(X)$ are finite. The differential entropy of $X + Z$ is upper bounded by:*

$$
h(X+Z) \leq h(X) + \frac{1}{2} \ln\left(1 + \sigma^2 J(X)\right), \qquad (5.42)
$$

and equality holds if and only if both X and Z are Gaussian distributed.

The rest of this section is dedicated to the direct implications of identity equation (5.42) which are fourfold:

1- While the usefulness of this upper bound is clear for RVs X having an infinite second moment for which equation (5.1) fails, it can in some cases, present a tighter upper bound than the one provided by Shannon for finite second moment variables X. This is the case, for example, when $Z \sim \mathcal{N}(\mu_1, \sigma^2)$ and X is a RV having the following PDF:

$$p_X(x) = \begin{cases} f(x+a) & -1-a \leq x \leq 1-a \\ f(x-a) & -1+a \leq x \leq 1+a, \end{cases}$$

for some $a > 0$ and where

$$f(x) = \begin{cases} \frac{3}{4}(1+x)^2 & -1 \leq x \leq 0 \\ \\ \frac{3}{4}(1-x)^2 & 0 < x \leq 1 \\ \\ 0 & \text{otherwise.} \end{cases}$$

The involved quantities related to X are easily computed and they evaluate to the following: $\mathrm{E}[X] = 0$, $\mathrm{E}[X^2] = a^2 + \frac{1}{10}$, $J(X) = 12$ and $h(X) = \ln\frac{4}{3} + \frac{2}{3}$, for which equation (5.42) becomes

$$h(X+Z) \leq h(X) + \frac{1}{2}\ln\left(1 + \sigma^2 J(X)\right) = \ln\frac{4}{3} + \frac{2}{3} + \frac{1}{2}\ln(1 + 12\,\sigma^2), \quad (5.43)$$

while equation (5.1) becomes

$$h(X+Z) \leq \frac{1}{2}\ln 2\pi e\left(\sigma_X^2 + \sigma^2\right) = \frac{1}{2}\ln 2\pi e + \frac{1}{2}\ln\left(a^2 + \frac{1}{10} + \sigma^2\right). \quad (5.44)$$

Comparing equations (5.43) and (5.44), it can be seen that our upper bound is independent of a whereas the Shannon bound increases to ∞ and gets looser as a increases. This is explained by the fact that our bound is location independent and depends only on the PDF of X whereas the Shannon bound is location dependent via the variance of X.

2- Theorem 19 gives a logarithmically growing analytical bound on the change in the transmission rates of the linear Gaussian channel function of an input scaling operation. This has been already mentioned when talking about the implications of Theorem 18. In the Gaussian case, the bound boils down to:

$$I(aX + Z; X) - I(X + Z; X) \leq \frac{1}{2} \ln \left(a^2 + \sigma^2 J(X) \right).$$

3- If the EPI is regarded as being a lower bound on the entropy of sums, equation (5.42) can be considered as its upper bound counterpart whenever one of the variables is Gaussian. In fact using both of these inequalities gives:

$$N(X) + N(Z) \leq N(Y) \leq N(X) + N(Z) \left[N(X) J(X) \right]. \tag{5.45}$$

It can be seen that the sandwich bound is efficient whenever the IIE in equation (5.5) evaluated for the variable X is close to its lower bound of 1.

4- Finally, in the context of communicating over a channel, it is well-known that, under a second moment constraint, the best way to "fight" Gaussian noise is to use Gaussian inputs. This follows from the fact that Gaussian variables maximize entropy under a second moment constraint. Conversely, when using a Gaussian input, the worst noise in terms of minimizing the transmission rates is also Gaussian. This is a direct result of the EPI and is also due to the fact that Gaussian distributions have the highest entropy and therefore are the worst noise to deal with. If one were to make a similar statement where instead of the second moment, the Fisher information is constrained, *i.e.*, if the input X is subject to a Fisher information constraint: $J(X) \leq A$ for some $A > 0$, then the input minimizing the mutual information of the additive white Gaussian channel is Gaussian distributed. This is a result of the EPI in equation (5.2) and the IIE in equation (5.5). They both reduce in this setting to

$$\arg \min_{X:J(X)\leq A} h(X+Z) \sim \mathcal{N}\left(0; \frac{1}{A}\right).$$

Reciprocally, for a Gaussian input, what is the noise that maximizes the mutual information subject to a Fisher information constraint? this problem can be formally stated as follows: If $X \sim \mathcal{N}(0;p)$, find

$$\arg \max_{Z:J(Z)\leq A} h(X+Z).$$

An intuitive answer would be Gaussian since it has the minimum entropy for a given Fisher information. Indeed, equation (5.42) provides the answer:

$$I(Y;X) \leq \frac{1}{2} \ln\left(1 + pJ(Z)\right),$$

is maximized whenever $Z \sim \mathcal{N}\left(0; \frac{1}{A}\right)$.

5.4 Related Publications

The results concerning the upper bound on the entropy of independent sums when one of the variables is Gaussian were published in [109].

Chapter 6

Estimation in Stable Noise Environments

6.1 Background

In the context of estimation, the use of the Mean Square Error (MSE) is tightly related to the assumption of finite variance RVs. One can even argue that it is related to a "potential Gaussian" setup. Well-known identities such as the Cramer-Rao bound which provides a lower bound on the mean square error of unbiased estimators in the from of the inverse of $J(X)$ are only valid in the finite variance setup. If the observed noisy variable is of infinite second moment, then the use of the Cramer-Rao bound in its classical form is to say the least problematic. We derive in this chapter a generalized Cramer-Rao bound, that relates the power (as defined in Section 4.4.2) of the estimation error to the generalized Fisher information $J_\alpha(\cdot)$. This can be achieved through a Generalized Isoperimetric Inequality for Entropies (GIIE) which we prove to hold.

6.2 A Generalized Isoperimetric Inequality for Entropies

Define $N_\alpha(X)$, $0 < \alpha \leq 2$, the entropy power of order α as

$$N_\alpha(X) = \frac{1}{e^{\alpha h(\tilde{Z})}} e^{\alpha h(X)}, \tag{6.1}$$

where $\tilde{Z} \sim \mathcal{S}\left(\alpha, (\frac{1}{\alpha})^{\frac{1}{\alpha}}\right)$.

Theorem 20 (Generalized Isoperimetric Inequality for Entropies (GIIE))**.** *Let X be a RV such that both $h(X)$ and $J_\alpha(X)$ exist, $1 < \alpha \leq 2$. Then*

$$N_\alpha(X) J_\alpha(X) \geq \kappa_\alpha, \tag{6.2}$$

where $\kappa_\alpha = e^{(\alpha-1)(\psi(\alpha)+\gamma_e)-1}$, γ_e is the Euler-Mascheroni constant and $\psi(\cdot)$ is the digamma function.

The evaluation of equation (6.2) for $\alpha = 2$ yields the well known IIE [92, Theorem 16]:

$$N(X)J(X) \geq 1. \tag{6.3}$$

The equality in equation (6.3) holds when X is Gaussian distributed. For general values of $1 < \alpha \leq 2$, whether the equality in equation (6.2) is achievable or not is still not answered.

Proof. In what follows, we make use of the results of Theorem 18 and use equation (5.41) to generalize the known IIE to one in terms of $J_\alpha(\cdot)$. In fact,

$$\alpha(h(X+Z) - h(X)) = t^{\alpha-1} {}_2F_1\left(\alpha-1, \alpha-1; \alpha; -t\right), \tag{6.4}$$

where $t = (\alpha\gamma^\alpha J_\alpha(X))^{\frac{1}{\alpha-1}}$. Since $t > 0$, using a transformation property of the

Gauss hypergeometric function presented in Appendix G, equation (6.4) gives:

$$\begin{aligned} \alpha(h(X+Z)-h(X)) &= t^{\alpha-1}(1+t)^{1-\alpha}\,{}_2F_1\left(\alpha-1,1;\alpha;\frac{t}{1+t}\right) \\ &= \left(\frac{t}{1+t}\right)^{\alpha-1}{}_2F_1\left(\alpha-1,1;\alpha;\frac{t}{1+t}\right). \end{aligned} \tag{6.5}$$

Using the series representation of the Gauss hypergeometric on the open unit disk, one can write:

$$\begin{aligned} {}_2F_1\left(\alpha-1,1;\alpha;\frac{t}{1+t}\right) &= \sum_{n=0}^{+\infty}\frac{(\alpha-1)_n(1)_n}{(\alpha)_n}\left(\frac{t}{1+t}\right)^n \\ &= \sum_{n=0}^{+\infty}\frac{\alpha-1}{n+\alpha-1}\left(\frac{t}{1+t}\right)^n, \end{aligned}$$

where $(A)_n = \frac{\Gamma(A+n)}{\Gamma(A)}$. The last equation is derived by making use of the following two properties of the gamma function:

$$\begin{aligned} \Gamma(A) &= (A-1)! \qquad A \in \mathbb{N}^* \\ \Gamma(A+1) &= A\Gamma(A) \qquad A \in \mathbb{R}^{+*}. \end{aligned}$$

Equation (6.5) is hence:

$$\alpha(h(X+Z)-h(X)) = (\alpha-1)\left(\frac{t}{1+t}\right)^{\alpha-1}\sum_{n=0}^{+\infty}\frac{1}{n+\alpha-1}\left(\frac{t}{1+t}\right)^n. \tag{6.6}$$

The LHS of equation (6.6) is lower bounded by:

$$\alpha(h(X+Z)-h(X)) \geq \alpha(h(Z)-h(X)) = \ln\frac{t^{\alpha-1}}{N_\alpha(X)J_\alpha(X)}, \tag{6.7}$$

where we used equation (6.1) and the fact that $t = (\alpha\gamma^\alpha J_\alpha(X))^{\frac{1}{\alpha-1}}$ in order to write the equality. Considering now the RHS of (6.6),

$$
\begin{aligned}
&(\alpha-1)\left(\frac{t}{1+t}\right)^{\alpha-1}\sum_{n=0}^{+\infty}\frac{1}{n+\alpha-1}\left(\frac{t}{1+t}\right)^{n}\\
&=(\alpha-1)\left(\frac{t}{1+t}\right)^{\alpha-1}\left[\frac{1}{\alpha-1}-\ln\left(1-\frac{t}{1+t}\right)\right.\\
&\qquad\left.-(\alpha-1)\sum_{n=1}^{+\infty}\frac{1}{n(\alpha+n-1)}\left(\frac{t}{1+t}\right)^{n}\right]\\
&=\left(\frac{t}{1+t}\right)^{\alpha-1}+(\alpha-1)\left(\frac{t}{1+t}\right)^{\alpha-1}\ln(1+t)\\
&\qquad-(\alpha-1)^2\sum_{n=1}^{+\infty}\frac{1}{n(\alpha+n-1)}\left(\frac{t}{1+t}\right)^{n+\alpha-1}.
\end{aligned}
$$

Hence equation (6.6) implies for any $t>0$:

$$
\begin{aligned}
&\ln N_\alpha(X)J_\alpha(X)-(\alpha-1)\ln t\\
&\geq-\left(\frac{t}{1+t}\right)^{\alpha-1}-(\alpha-1)\left(\frac{t}{1+t}\right)^{\alpha-1}\ln(1+t)\\
&\qquad+(\alpha-1)^2\sum_{n=1}^{+\infty}\frac{1}{n(\alpha+n-1)}\left(\frac{t}{1+t}\right)^{n+\alpha-1}.
\end{aligned}
\tag{6.8}
$$

Letting the scale $\gamma\to+\infty$, then $t\to+\infty$ and equation (6.8) gives

$$
\begin{aligned}
\ln N_\alpha(X)J_\alpha(X)&\geq(\alpha-1)^2\sum_{n=1}^{+\infty}\frac{1}{n(\alpha+n-1)}-1 \qquad (6.9)\\
&=(\alpha-1)\left(\psi(\alpha)+\gamma_{\mathrm{e}}\right)-1,
\end{aligned}
$$

which is the required result. In order to write equation (6.9), we used the fact that the series $\sum_{n=1}^{+\infty}\frac{1}{n(\alpha+n-1)}\left(\frac{t}{1+t}\right)^{n+\alpha-1}$ is absolutely convergent in order to interchange the order of the limits. Equation (6.2) is referred to as the generalized isoperimetric inequality for entropies. Finally, we note that whenever $\alpha=2$, $\psi(\alpha)=-\gamma_{\mathrm{e}}+1$, and the generalized isoperimetric inequality boils down to the standard one.

We plot in Figure 6-1 the evaluation of the LHS of equation (6.2) at the values of $\alpha = [1.2, 1.4, 1.6, 1.8]$ for alpha-stable RVs $\mathcal{S}\left(r, (r)^{-\frac{1}{r}}\right)$ for the values of $r = [0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8]$. The horizontal lines represent the RHS of equation (6.2) for the considered values of α. Note that stable variables do not achieve the lower bound of the GIIE (equation (6.2)) except when $\alpha = 2$ where Gaussian variables achieve the lower bound. In fact, Figure 6-2 shows the evaluation of the product $N_{1.8}(X)J_{1.8}(X)$ whenever $X = X_1 + X_2$ where $X_1 \sim \mathcal{S}\left(\alpha, (\alpha)^{-\frac{1}{\alpha}}\right)$ for $r = 1.8$ and $X_2 \sim \mathcal{N}(0, \sigma^2)$ for different value of σ. The minimum is achieved for $\sigma = 4$ and not when X is alpha-stable (which is the case when $\sigma = 0$). Note that the computed minimum in Figure 6-2 is by no means a global minimum.

Whether there exists RVs that achieve the minimum of $N_\alpha(X)J_\alpha(X)$ and whether the lower bound κ_α is tight or not are still to be determined. $\square$

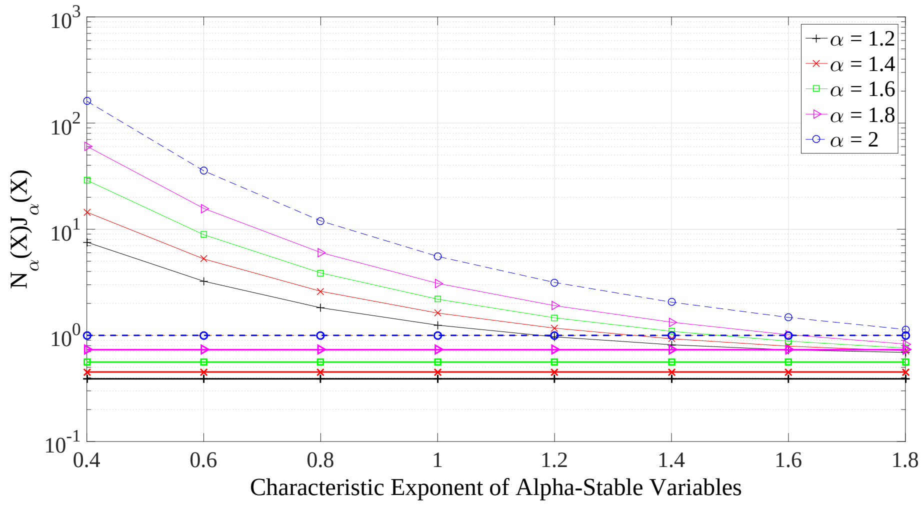

Figure 6-1: Evaluation of $N_\alpha(X)J_\alpha(X)$ and comparing it to κ_α for $X \sim \mathcal{S}\left(r, (r)^{-\frac{1}{r}}\right)$ for different values of α and r.

Figure 6-3 shows the relative tightness of the lower bound κ_α when the RHS of equation (6.2) is evaluated at alpha-stable variables with characteristic exponents

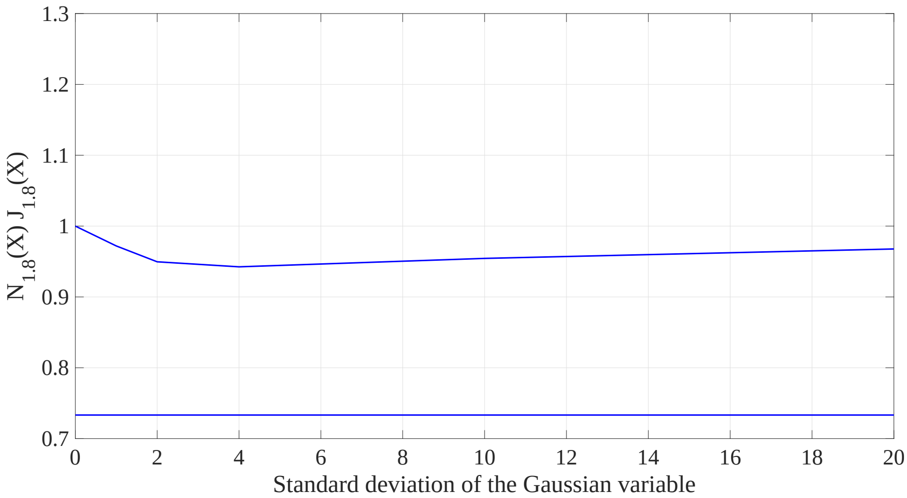

Figure 6-2: Evaluation of $N_{1.8}(X)J_{1.8}(X)$ and comparing it to $\kappa_{1.8} = 0.7333$ for $X = X_1 + X_2$ where $X_1 \sim \mathcal{S}\left(r, (r)^{-\frac{1}{r}}\right)$ for $r = 1.8$ and $X_2 \sim \mathcal{N}(0, \sigma^2)$ for different values of σ.

r ranging from 0.4 to 1.8. If we consider for example on the x-axis the value of $r = 0.8$ which corresponds to the alpha-stable variable $X \sim \mathcal{S}\left(r, (r)^{-\frac{1}{r}}\right)$, Figure 6-3 says that as α increases the relative tightness of $N_\alpha(X)J_\alpha(X)$ with respect to κ_α decreases.

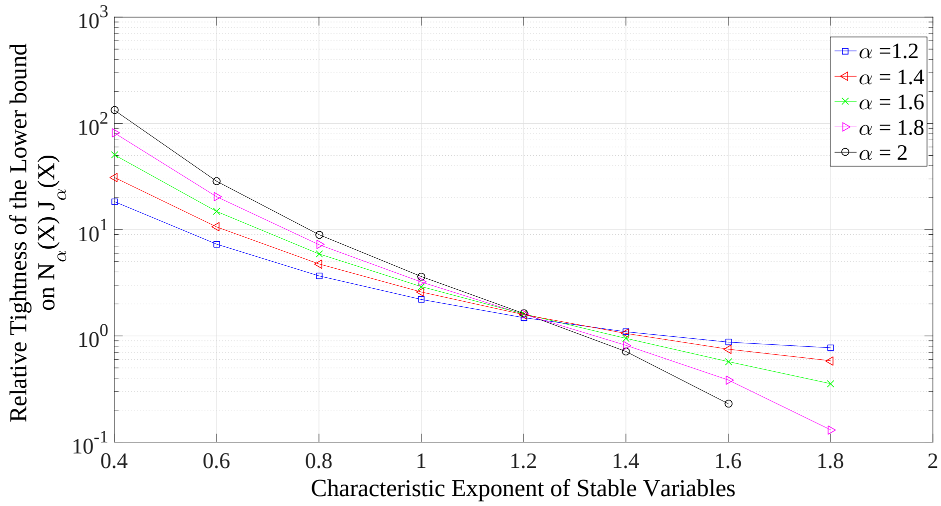

Figure 6-3: Relative tightness of κ_α for alpha-stable variables.

A direct implication of equation (6.2) is summarized in the following: let P_X denotes the power of the RV X according to equation (4.23). Then according to equation (4.44):

$$N_\alpha(X) \leq N_\alpha(Z),$$

where $Z \sim \mathrm{P}_X \tilde{Z}$ is an alpha-stable variable of characteristic exponent α and power equal to P_X. Equation (6.2) necessarily implies

$$J_\alpha(X) \geq \kappa_\alpha J_\alpha(Z) = \kappa_\alpha \frac{1}{(\mathrm{P}_X)^\alpha}, \tag{6.10}$$

where we used the fact that

$$N_\alpha(Z) J_\alpha(Z) = 1.$$

Equation (6.10) is a generalization of the well known fact that for any X with variance σ^2, $J_2(X) \geq J_2(Z) = \frac{1}{\sigma^2}$ where $Z \sim \mathcal{N}(\mu, \sigma^2)$.

6.3 Estimation in impulsive noise environments: A Generalized Cramer-Rao Bound

Suppose we want to estimate a non-random parameter $\theta \in \mathbb{R}$ based on a noisy observation X where the additive noise N is supposed to be of impulsive nature. Needless to say that in this case the Minimum Mean Square Error (MMSE) estimator is not sensible. We proceed by considering unbiased estimators where the location and power parameters are defined according to Section 4.4.2. More explicitly, let

$$X = \theta + N, \tag{6.11}$$

where N is a noise variable having both $h(N)$ and $J_\alpha(N)$ exist and finite. Define $\hat{\theta}(X)$ as an unbiased estimator of θ based on the observation of the RV X where by unbiased is meant that:

Definition (Unbiased Estimators)**.** Let $\hat{\theta}(X)$ be an estimator of a parameter θ based on the observation of a RV X. $\hat{\theta}(X)$ is said to be unbiased if

$$\mathrm{L}_{\hat{\theta}(X)} = \theta,$$

and

$$\mathrm{L}_{\hat{\theta}(X)-\theta} = 0, \tag{6.12}$$

where L_X is the location parameter defined in equation (4.22) and equation (6.12) is by virtue of property 1- of the location parameter.

A good indicator of the quality of an estimator $\hat{\theta}(X)$ is the power of the "error" $\left(\hat{\theta}(X) - \theta\right)$. We find next a lower bound on such metric which generalizes the previously known Cramer-Rao bound.

Theorem 21 (Generalized Cramer-Rao Bound)**.** *Let $\hat{\theta}(X)$ be an unbiased estimator of the parameter θ based on the observation X according to equation (6.11). Then*

$$P_{\hat{\theta}(X)-\theta} \geq \frac{\kappa_\alpha^{\frac{1}{\alpha}}}{J_\alpha^{\frac{1}{\alpha}}(N)}, \tag{6.13}$$

where $P_{\hat{\theta}(X)-\theta}$ is the power of the "error" $\left(\hat{\theta}(X) - \theta\right)$ according to equation (4.23).

Before starting the proof, we note that whenever $\alpha = 2$ the result of Theorem boils down to the classical Cramer-Rao bound:

$$\mathrm{E}\left[\left(\hat{\theta}(X) - \theta\right)^2\right] \geq \frac{1}{J(N)}. \tag{6.14}$$

Proof. Let $\mathrm{P_e} = \mathrm{P}_{\hat{\theta}(X)-\theta}$. By the results of Section 4.4.3, among all RVs that have a power equal to $\mathrm{P_e}$, the entropy maximizing variable Z is distributed according to $\mathcal{S}\left(\alpha, \left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha}} \mathrm{P_e}\right)$ and

$$h\left(\hat{\theta}(X) - \theta\right) \leq h(\tilde{Z}) + \ln \mathrm{P_e}, \tag{6.15}$$

where $\tilde{Z} \sim \mathcal{S}\left(\alpha, \left(\frac{1}{\alpha}\right)^{\frac{1}{\alpha}}\right)$. Equation (6.15) implies that:

$$N_\alpha\left(\hat{\theta}(X) - \theta\right) \leq \mathrm{P}_\mathrm{e}^\alpha. \tag{6.16}$$

On the other hand

$$J_\alpha\left(\hat{\theta}(X) - \theta\right) = J_\alpha\left(\hat{\theta}(X)\right) \leq J_\alpha(X) = J_\alpha(N), \tag{6.17}$$

where the first and the last equalities are due to property 3- of Section 5.2 and the inequality is due to the data processing inequality for $J_\alpha(\cdot)$ proven in Theorem 16. Applying the GIIE (6.2) to $\hat{\theta}(X) - \theta$, we obtain:

$$N_\alpha\left(\hat{\theta}(X) - \theta\right) J_\alpha\left(\hat{\theta}(X) - \theta\right) \geq \kappa_\alpha, \tag{6.18}$$

which gives along with equations (6.16) and (6.17),

$$J_\alpha(N)\,\mathrm{P}_\mathrm{e}^\alpha \geq \kappa_\alpha. \tag{6.19}$$

Equation (6.19) establishes a generalized novel Cramer-Rao lower bound on the power P_e of the error of all unbiased estimators $\theta\,(\hat{X})$:

$$\mathrm{P}_\mathrm{e} \geq \frac{\kappa_\alpha^{\frac{1}{\alpha}}}{J_\alpha^{\frac{1}{\alpha}}(N)},$$

which gives

$$\mathrm{P}_\mathrm{e} \geq (\alpha\kappa_\alpha)^{\frac{1}{\alpha}}\,\gamma_N,$$

when $N \sim \mathcal{S}(\alpha, \gamma_N)$, $1 < \alpha \leq 2$ is a stable variable by virtue of the fact that $J_\alpha(N) = \frac{1}{\alpha\gamma_N^\alpha}$ (see property 4- in Section 5.2). □

As an example, in the case where $N \sim \mathcal{S}(\alpha, \gamma_N)$, $1 < \alpha \leq 2$, the Maximum

Likelihood (ML) estimator $\hat{\theta}_{\mathrm{ML}}(X)$ is given by:

$$\hat{\theta}_{\mathrm{ML}}(X) = \mathrm{argmax}_{\theta} \ln p_N(X - \theta) = X$$

since N is symmetric. Hence $\hat{\theta}_{\mathrm{ML}}(X)$ is unbiased. The power of $\hat{\theta}_{\mathrm{ML}}(X) - \theta = N$ is $\mathrm{P_e} = \mathrm{P}_N = (\alpha)^{\frac{1}{\alpha}}\gamma_N = \frac{1}{J_{\alpha}^{\frac{1}{\alpha}}(N)}$ for which equation (6.13) holds true. Finally, we mention that equation (6.13) establishes a new metric to measure the average error strength and hence the estimator performance when the noisy measurements are affected by an additive noise of impulsive nature. The choice of a specific value of α is straightforward whenever the noise belongs to the α-parametrized domains of normal attraction of stable variables. The quality of the estimator $\hat{\theta}(X)$ is tied to the closeness of $\mathrm{P_e}$ to its lower bound, both of which are computable numerically as previously shown in Chapters 4 and 5 where the power and generalized Fisher information are computed for several types of probability laws. On a final note, we mention that it is not known whether equation (6.13) is tight is not known in general. The tightness is already know when $\alpha = 2$ (equation 6.14) for $\hat{\theta}(X) = X$ and N a Gaussian variable. We believe that answering the tightness question is equivalent to a similar question when it comes to the GFII (5.28).

Chapter 7

Conclusions

In any communication or measurement setup, the observed signal is a perturbed/ distorted version of the signal of interest. Whether the source of the noise comes from the equipments heating or an interferer, the effect of the perturbation is generally modeled in an additive manner. Generally, the role of a system designer is to build an efficient system that recovers at the receiver side the information present in a quantity of interest. In this thesis we studied various theoretical aspects of such problems when the noise is heavy tailed, a scenario in which alpha-stable distributions play a central role and find applications in diverse fields of engineering and some other disciplines. Whenever possible, such as in a communications setup, one may choose the "best" input signal – equivalently the "best" input statistics that satisfies the different constraints and carries the required information. This optimal input achieves the best tradeoff in the sense that it maximizes the performance while conserving a low average error. The classical approach to this fundamental problem in communications theory is done from a channel input perspective. Under this perspective and in the purpose of emulating real scenarios, input signals are supposed to abide by some power constraints. The Hilbert space structure of random processes having finite $\mathbb{L}^2$-norm lead, by using orthonormal expansions, to translating the $\mathbb{L}^2$-norm power constraint to a second moment constraint in the Hilbert space of RVs having finite second moments. Assuming that the additive noise would also have a finite second moment, this approach quantified the different metrics of the channel with respect to the input power measure irrespective of the noise model. As an example, the capacity of the linear additive Gaussian channel under an average power constraint is given by the famous formula "$C = \ln(1 + \text{SNR})$" where the "SNR" is the signal to noise ratio between the variance of the input to that of the Gaussian noise, hence relating the input power as defined for the input space to the noise power since the noise falls within the input space. Naturally, this approach breaks when the noise is not of the same "nature" as the input space. This is the case of impulsive noise models having infinite second moments which do not belong to the input space of finite power (second moment) RVs .

We adopted an alternative way to approach the problem by looking at the received signal. Naturally, the performance of any adopted strategy at the input is viewed by its effect at the output end. This is translated also to the involved mathematical quantities in the transmission/reconstruction problem since the maximization of the transmission rates involves a maximization of the output differential entropy. Therefore, in a communications theory setup, it seems reasonable to consider the additive channel while imposing a "quality" constraint on the output. By restricting the output space to satisfy certain power requirements, we are indirectly taking into consideration the nature of the noise in the formulation of the constraint which constructs an input space of variables of the same "nature" of the noise. This is in accordance with the fact that the system designer has no control over the noise model which is dictated by the channel and can assume the possibility of choosing, form an input space similar in nature to that of the noise, the input signal that best overcomes the noise effect. Back to the linear AWGN channel, the output approach gives exactly the same answer as the input approach: constraining the output average power implies a constraint on the input average power. Furthermore, in this case, the capacity is better viewed in terms of the output metrics "$C = \ln(\text{SNR})$", where in this case "SNR" is the output signal to noise ratio.

When dealing with an additive alpha-stable noise channel, the contaminated signal at the receiver level has already an infinite second moment whatever constraints one would impose on the input. From the "output approach" perspective, the system designer tries to evaluate the performance of the system by setting a measure that quantifies the strength of the received signal. Since the second moment is no longer an option, other tools are to be investigated. A standard way to define a measurement operator is to evaluate the average cost of the RV for a given cost function. As a first observation, the logarithmic tail behaviour of the cost function seems to be essential. It simply gives finite measurements for all heavy tailed distribution functions. This first observation is further supported by the generic channel capacity results of Chapter 3 via the answer to these two questions: is it normal for an infinite power signal not to carry more than a finite number of information bits? and is it reasonable that a finite power signal achieves better transmission rates than one whose power is infinite? If the answer is no, then by the results of Chapter 3, whenever impulsive noise channels are encountered logarithmic cost functions should be used as a way to measure the signals' power.

We propose in Section 4.4.2, an expression to evaluate the power of signals in symmetric alpha-stable noise environments. Though the value of the power is incorporated within the cost function, it represents an average of a logarithmically tailed cost function. Besides the logarithmic tail behaviour of the averaged function, the main argument for suggesting P_X as defined in equation 4.23, is to find a definition that is generic for the stable space of noise distributions, including the Gaussian since stable distributions are the most common noise models encountered by virtue of the GCLT. Equation (4.23) is chosen to become the standard deviation in the Gaussian case in order to unify the order of the power operator in such a way if the variable is linearly scaled then the power also scales linearly.

Definition (4.23) defines a space where the alpha-stable noise is the worst in terms of entropy/randomness which implies that the alpha-stable channel model is a worst-case scenario whenever there is an impulsive noise assumption. This fact mimics the role of the Gaussian variable among the finite variance space of RVs and generalize it to an equivalent role of stable variables among the space of RVs that have a finite power P_X.

This central role is reflected also in other related problems specially in estimation theory. A simple estimation problem of the location parameter of an alpha-stable variable is not well understood and performance measures of a given estimator are to be further investigated. Though the work of Gonzales [1] was in this direction, we believe that it was suitable for the Cauchy case and not generic to the whole family of symmetric stable distributions. Additionally, that work was developed with a "signal processing" aspect in mind. The generic quantities defined in this dissertation such as the generalized Fisher information, the generalized entropy power, the power P_X and other relevant quantities establish an "extension" of the Gaussian estimation theory to a stable estimation theory in general and may be viewed as complementary to the works found in the literature by answering the "fundamental-limits" questions. The generalized Cramer-Rao bound proven in Chapter 6 sets a novel lower bound on the power of the estimation error for any "unbiased" estimator of a location parameter that can be used to characterize the performance of estimators in impulsive noise environments. It naturally opens the door for the related problems of efficiency and ML estimator.

Appendix A

The Karush-Kuhn-Tucker Conditions

In this appendix we state the general requirements needed to be satisfied by the optimization set i.e. the set of the probability distribution functions that satisfy given constraints and the objective function i.e. the mutual information between the input and the output of a given channel in order to prove the existence and uniqueness of the optimal input. We start by the basic optimization problem and find equivalent necessary and sufficient conditions for the solution to satisfy. These conditions are known as the KKT conditions. For further explanations on this subject see [57].

A.1 Existence and Uniqueness

Theorem 22 (Extreme Value Principle)**.** *If I is a real-valued, weak continuous functional on a weak compact set $\Omega \subseteq \mathcal{F}$, then I* achieves *its maximum on Ω.*

If furthermore Ω is convex, and I is strictly concave, then the maximum

$$C = \max_{F \in \Omega} I(F)$$

is achieved by a unique F_o *in* Ω.

Proof.

- First statement [85]
- Second statement follows from strict concavity

□

A.2 Lagrangian Theorem for Constrained Optimization Problems

Theorem (Lagrangian Theorem). *[85] Let $\mathcal{F}$ be a vector space and Ω a convex subset of $\mathcal{F}$.*

Let I be a real-valued concave functional on Ω and g a convex mapping from Ω to $\mathbb{R}$. Assume the existence of a point $F_1 \in \Omega$ for which $g(F_1) < 0$.

Let

$$C = \sup_{\substack{F \in \Omega \\ g(F) \leq 0}} I(F)$$

and assume C is finite. Then there is an element $\lambda_o \geq 0$ such that

$$C = \sup_{F \in \Omega} \{I(F) - \lambda_o g(F)\}.$$

Furthermore, if the supremum is achieved in the original problem at F_o, it is achieved by F_o in the second and

$$\lambda_o g(F_o) = 0.$$

A.3 KKT conditions

Theorem (Optimization and Differentiability)**.** *Assume a weakly (Gateaux) differentiable functional I on a convex set Ω achieves its maximum.*

- *If I achieves its maximum at F_o then $I'_{F_o}(F) \leq 0$ for all $F \in \Omega$.*
- *If I is concave, then $I'_{F_o}(F) \leq 0$ for all $F \in \Omega$ implies that I achieves its maximum at F_0.*

Proof. See Smith [57] □

Theorem (Karush-Kuhn-Tucker (KKT) Conditions)**.** *The capacity C of a constrained channel*

$$C = \sup_{\int h(x)\, dF(x) \leq a} I(F)$$

is achieved by F_o if and only if there exists a $\gamma \geq 0$ such that

$$\gamma(h(x) - a) + C - \int p(y|x) \ln \left[\frac{p(y|x)}{p(y; F_o)} \right] dy \geq 0$$

for all x, with equality if x is a point of increase of F_o.

Proof. See Abou-Faycal, Trott and Shamai [59] □

Appendix B

Sufficient Conditions for Finiteness of Channel Capacity

In this appendix, we show that assumptions C1 to C6 imposed in Chapter 3 on channel (3.3) are sufficient conditions that ensure that the mutual information between the channel's input and output is finite –and hence well-defined– and we make use of the extreme value principle [85] to ensure that the capacity problem yields a finite and achievable solution. This could be achieved by enforcing two characteristics:

1- The input space $\mathcal{P}_A$ of feasible distribution functions is compact.

2- The mutual information between the input and the output of the channel is continuous in the input distribution function.

We emphasize that these two properties are intimately related to the channel model and the input constraints if any. Under rather mild technical conditions imposed on the noise PDF (conditions C5 and C6), we show that whenever the input cost function has a "super-logarithmic growth" (condition C4) the channel capacity is *finite* and *achievable*.

Establishing the continuity of the mutual information under any "super - logarithmic" input constraint is achieved using a novel result on the *convergence* *of differential entropies*. While numerous studies have tackled this subject (see for example [110, 111]), the conditions presented in Section B.1 are among the weakest that insure this convergence whenever PDFs converge point-wise.

B.1 Convergence of differential entropies

In this section we establish a sufficient condition for the convergence of differential entropies whenever there is point-wise convergence of the corresponding PDFs. More precisely, we prove a theorem that guarantees this convergence under some rather-mild sufficient conditions. In layman terms, this theorem states that whenever the PDFs satisfy a super-logarithmic type of moment, point-wise convergence will imply convergence of differential entropies. We emphasize that the new conditions are weaker than some of those derived by Godavarti et al. [111, Theorem 1]. Alternative conditions found in [111, Theorem 4] are not directly related to those presented hereafter.

Theorem 23 (Convergence of Differential Entropies)**.** *Let the sequence of PDFs on* $\mathbb{R}$*,* $\{p_m(y)\}_{m\geq 1}$ *and* $p(y)$ *satisfy the following conditions:*

A1- The PDFs $\{p_m(y)\}_m$ *and* $p(y)$ *are uniformly upperbounded:*

$$\exists\, M \in (0,\infty) \; s.t. \sup_{y\in\mathbb{R}, m\geq 1} \Big\{ p_m(y), p(y) \Big\} \leq M. \tag{B.1}$$

A2- There exists a non-negative and non-decreasing function $l : [0,\infty) \to [0,\infty)$*, such that* $l(y) = \omega\left(\ln(y)\right)$ *(i.e.* $\forall\, \kappa > 0, \exists\, c > 0$ *such that* $l(y) \geq \kappa \ln(y), \forall y \geq c$*) and*

$$\sup_m \Big\{ \mathsf{E}_{p_m}\left[l(|Y|)\right], \mathsf{E}_p\left[l(|Y|)\right] \Big\} \leq L, \tag{B.2}$$

for some positive (finite) value L.

Under these conditions, $h(p_m) \to h(p)$ *whenever the PDFs* $p_m(y) \to p(y)$ *point-wise.*

Before we prove the theorem, we highlight the importance of condition A2 by providing an example where it is not satisfied, and the theorem does not hold.

Example 1. Consider the sequence of PDFs $\{p_m(x)\}_{m\geq 3}$ defined on $\mathbb{R}$ as follows:

$$p_m(x) = \begin{cases} 1 - \dfrac{1}{\ln m} & x \in [0; 1] \\ \dfrac{1}{(\ln m)^2} \dfrac{1}{x} & x \in (1; m]. \end{cases}$$

This sequence of PDFs converges point-wise to $p(x)$, the uniform distribution on $[0, 1]$, and condition A1 is satisfied with a uniform upperbound $M = 1$. Computing the differential entropies,

$$\begin{aligned} h(p) &= 0 \\ h(p_m) &= -\left(1 - \frac{1}{\ln m}\right) \ln\left(1 - \frac{1}{\ln m}\right) \\ &\quad + \frac{2\ln(\ln m)}{(\ln m)^2} \int_1^m \frac{1}{x}\, dx + \frac{1}{(\ln m)^2} \int_1^m \frac{\ln x}{x}\, dx \\ &= -\left(1 - \frac{1}{\ln m}\right) \ln\left(1 - \frac{1}{\ln m}\right) + \frac{2\ln(\ln m)}{\ln m} + \frac{1}{2} \\ &\to \frac{1}{2} \text{ as } m \to \infty, \end{aligned}$$

and hence there is no convergence of differential entropies. This is explained by the fact that condition A2 is not satisfied. Indeed, consider any function $l(x)$ that is non-negative, non-decreasing and $l(x) = \omega(\ln x)$. By definition, for any $\kappa > 0$, there exists a $c > 0$ such that $l(x) \geq \kappa \ln x$ for $x \geq c$. Therefore, for any

$m \geq c$,

$$\begin{aligned}
\mathsf{E}_{p_m}\left[l(|X|)\right] &= \left(1 - \frac{1}{\ln m}\right)\int_0^1 l(x)\,dx + \frac{1}{(\ln m)^2}\int_1^m \frac{1}{x}\,l(x)\,dx \\
&= \left(1 - \frac{1}{\ln m}\right)\int_0^1 l(x)\,dx + \frac{1}{(\ln m)^2}\int_1^c \frac{1}{x}\,l(x)\,dx \\
&\qquad + \frac{1}{(\ln m)^2}\int_c^m \frac{1}{x}\,l(x)\,dx \\
&\geq \left(1 - \frac{1}{\ln m}\right)\int_0^1 l(x)\,dx + \frac{1}{(\ln m)^2}\int_1^c \frac{1}{x}\,l(x)\,dx \\
&\qquad + \frac{\kappa}{(\ln m)^2}\int_c^m \frac{1}{x}\,\ln x\,dx \\
&= \left(1 - \frac{1}{\ln m}\right)\int_0^1 l(x)\,dx + \frac{1}{(\ln m)^2}\int_1^c \frac{1}{x}\,l(x)\,dx \\
&\qquad + \kappa\,\frac{(\ln m)^2 - (\ln c)^2}{2(\ln m)^2} \\
&\geq \kappa\,\frac{(\ln m)^2 - (\ln c)^2}{2(\ln m)^2},
\end{aligned}$$

which is greater than $\frac{3}{8}\kappa$ whenever $m > c^2$. Since the inequality holds for any $\kappa > 0$ and m large enough then $\sup_m \left\{\mathsf{E}_{p_m}\left[l(|X|)\right]\right\}$ is unbounded which violates condition A2. We proceed next to the proof of Theorem 23.

Proof. We start by noting that the differential entropies $h(p)$ and $\{h(p_m)\}_{m\geq 1}$ exist and are finite by virtue of the fact that the PDFs are upperbounded and have a finite logarithmic moment [82, Proposition 1].

Assume now that the conditions of the theorem hold and that p_m converges to p point-wise. If the upperbound (B.1) M is larger than one, consider the change of variables, $Z = MY$ (for which $h(Z) = h(Y) + \ln M$,) or equivalently the PDFs,

$$d(y) \hat{=} \frac{1}{M}p\left(\frac{y}{M}\right), \qquad d_m(y) \hat{=} \frac{1}{M}p_m\left(\frac{y}{M}\right), m \geq 1.$$

These densities are upperbounded by one and the sequence $\{d_m(y)\}$ converges point-wise to $d(y)$. Furthermore, the function $l'(y) = l(y/M)$ is non-negative, non-decreasing and $l'(y) = \omega\left(\ln(y)\right)$. Additionally,

$$\mathsf{E}_{d_m}\left[l'(|Y|)\right] = \mathsf{E}_{p_m}\left[l'(|MY|)\right] = \mathsf{E}_{p_m}\left[l(|Y|)\right] \leq L.$$

The conditions of the theorem therefore hold for the laws $\{d_m, d\}$ and in what follows we assume without loss of generality that $M \leq 1$, and the differential entropies are all non-negative .

Let $\tilde{y}$ be any positive scalar such that $l(\tilde{y}) > 0$, and denote by $q(y) = \frac{1}{\pi}\frac{1}{1+y^2}$ the Cauchy density. Then, using the convention "$0 \ln 0 = 0$" and the fact that $y \ln y \geq -\frac{1}{e}$ for $y > 0$, we can write

$$\begin{aligned}
& -\int\limits_{|y|\geq\tilde{y}} p(y)\ln p(y)\,dy \\
&= -\int\limits_{|y|\geq\tilde{y}} p(y)\ln q(y)\,dy + \int\limits_{|y|\geq\tilde{y}} q(y)\frac{p(y)}{q(y)}\ln\frac{q(y)}{p(y)}\,dy \\
&\leq \ln\pi \int\limits_{|y|\geq\tilde{y}} p(y)\,dy + \int\limits_{|y|\geq\tilde{y}} \ln\left[1+y^2\right] p(y)\,dy + \frac{1}{e}\int\limits_{|y|\geq\tilde{y}} q(y)\,dy \\
&\leq \frac{\ln\pi}{l(\tilde{y})}\int\limits_{|y|\geq\tilde{y}} l(|y|)\,p(y)\,dy + \int\limits_{|y|\geq\tilde{y}} \ln\left[1+y^2\right] p(y)\,dy \\
&\qquad\qquad + \frac{1}{e}\frac{1}{\ln\left[1+\tilde{y}^2\right]}\int\limits_{|y|\geq\tilde{y}} \ln\left[1+y^2\right] q(y)\,dy,
\end{aligned} \tag{B.3}$$

where equation (B.3) is due to the fact that $l(\cdot)$ is non-decreasing. Hence,

$$\begin{aligned}
&- \int\limits_{|y|\geq\tilde{y}} p(y) \ln p(y)\, dy \\
&\leq \ln\pi \frac{\mathsf{E}_p\left[l(|Y|)\right]}{l(\tilde{y})} + 2 \int\limits_{|y|\geq\tilde{y}} \ln\left[1+|y|\right] p(y)\, dy + \frac{1}{e}\frac{\mathsf{E}_q\left[\ln\left[1+Y^2\right]\right]}{\ln\left[1+\tilde{y}^2\right]} && \text{(B.4)} \\
&\leq \frac{L\ln\pi}{l(\tilde{y})} + 2\sup_{|y|\geq\tilde{y}}\left\{\frac{\ln\left[1+|y|\right]}{l(|y|)}\right\} \int\limits_{|y|\geq\tilde{y}} l(y)\, p(y)\, dy + \frac{1}{e}\frac{\ln 4}{\ln\left[1+\tilde{y}^2\right]} && \text{(B.5)} \\
&\leq \frac{L\ln\pi}{l(\tilde{y})} + 2L\sup_{|y|\geq\tilde{y}}\left\{\frac{\ln\left[1+|y|\right]}{l(|y|)}\right\} + \frac{1}{e}\frac{\ln 4}{\ln\left[1+\tilde{y}^2\right]}, && \text{(B.6)}
\end{aligned}$$

where equation (B.4) is justified since $l(\tilde{y})$ is positive and $l(y)$ is non-negative. In order to write equation (B.5) we use the identity $\mathrm{E}_q\left[\ln\left(1+y^2\right)\right] = \ln 4$ [90, Sec.3.1.3, p.51]. The supremum in equations (B.5) and (B.6) is finite –and goes to 0– for $\tilde{y}$ large-enough because $l(y) = \omega\left(\ln y\right)$.

Since the upperbound (B.6) also holds for any $p_m(y)$, then for every $\delta > 0$, there exists a $\tilde{y} > 0$ such that for all $m \geq 1$:

$$\left| \int\limits_{|y|\geq\tilde{y}} p_m(y) \ln p_m(y)\, dy \right| < \delta \;\&\; \left| \int\limits_{|y|\geq\tilde{y}} p(y)\, \ln p(y)\, dy \right| < \delta.$$

It remains to show that

$$\lim_{m\to+\infty} - \int\limits_{|y|<\tilde{y}} p_m(y) \ln p_m(y)\, dy = - \int\limits_{|y|<\tilde{y}} p(y)\, \ln p(y)\, dy,$$

which is guaranteed by the Dominated Convergence Theorem (DCT) since

$$|p_m(y) \ln p_m(y)| \leq \frac{1}{e},$$

by virtue of the fact that $p_m(y) \leq 1$ for all m, which completes the proof. □

B.2 Finiteness of channel capacity

Sufficient conditions

We show in this section that conditions C1 to C6 stated in Chapter 3 represent sufficient conditions on the triplet $f(\cdot)$, $\mathcal{C}(\cdot)$, and $p_N(\cdot)$ that guarantee the finiteness and the achievability of the capacity of channel (3.1):

Theorem 24 (Achievability and Finiteness of Channel Capacity)**.** *Under conditions C1 through C6, the capacity of the average-cost constrained channel (3.3) is finite and achievable.*

Furthermore, the maximum is achieved by a unique F^* in $\mathcal{P}_A$ if and only if the output PDF is injective in F.

We point out that assumptions C1, C2, C5 and C6 are related to the channel model at hand and are not "conditions" per say. These assumptions are satisfied by the vast majority of common models found in the literature.

When thinking in terms of conditions on the input –controlled by the user, C3 and C4 are to be considered. Note that these conditions are also common to all cost functions found in the literature. While C3 is rather technical, the relevance of C4 may be seen in the following example.

Example 2. Consider the linear additive channel (3.3), where now the noise N is a uniformly distributed random variable on the interval $[0, 1)$.

Let X_1 and X_2 be two discrete random variables taking integer values $k \geq 2$, with respective probability mass functions:

$$p_{X_1}(k) = B_1 \frac{1}{k(\ln k)^2}, \quad p_{X_2}(k) = B_2 \frac{1}{k(\ln k)^3}, \qquad k \geq 2,$$

where B_1 & B_2 are the normalizing finite constants,

$$B_1 = \left[\sum_{k=2}^{\infty} \frac{1}{k(\ln k)^2} \right]^{-1} \quad B_2 = \left[\sum_{k=2}^{\infty} \frac{1}{k(\ln k)^3} \right]^{-1}.$$

Let Y_1 and Y_2 be the outputs of channel (3.3) whenever its inputs are X_1 and X_2 respectively. Given the placement of the mass points, X_1 may be perfectly inferred from Y_1 and $H(X_1|Y_1) = 0$. Similarly $H(X_2|Y_2) = 0$ and therefore the mutual informations

$$I(X_1; Y_1) = H(X_1) - H(X_1|Y_1) = H(X_1)$$
$$I(X_2; Y_2) = H(X_2).$$

Computing $H(X_1)$ and $H(X_2)$, we obtain:

$$\begin{aligned} H(X_i) &= -\sum_{k\geq 2} p_{X_i}(k) \ln p_{X_i}(k) \\ &= -\ln B_i + B_i \sum_{k\geq 2} \frac{\ln k + (1+i)\ln(\ln k)}{k(\ln k)^{1+i}} \qquad i = 1, 2, \end{aligned}$$

which diverges for $i = 1$ and converges for $i = 2$. Accordingly, the mutual information of channel (3.3) is infinite when the input is X_1 whereas it is finite for input X_2. Note that $\mathsf{E}[\ln X_1]$ is infinite while $\mathsf{E}[\ln X_2]$ is finite, and this example showcases the importance of condition C4 when it comes to the finiteness of mutual information. Whenever C4 is not enforced, the channel capacity might be infinite as X_1 yields an infinite mutual information. The theorem states that when the condition is enforced, the capacity will be finite.

An interesting observation is that both $\mathsf{E}[X_1^2]$ and $\mathsf{E}[X_2^2]$ are infinite, however as inputs to the channel they yield respectively an infinite and a finite mutual information. We proceed next to prove Theorem 24.

Proof. The first statement of the theorem is established using the extreme value principle (see Theorem 22 in Appendix A).

In order to apply this principle, we show in Section B.3 that the set $\mathcal{P}_A$ is *compact* (Theorem 25) and that the mutual information $I(F)$ is *finite* and *continuous* (Theorems 26 and 27). Therefore, the capacity of the average-cost constrained channel is finite and achievable.

When it comes to uniqueness, since $\mathcal{P}_A$ is *convex* (Theorem 25) whenever $I(\cdot)$ is *strictly concave*, then the maximum

$$C = \max_{F \in \mathcal{P}_A} I(F),$$

is achieved by a unique F^* in $\mathcal{P}_A$.

Knowing that $I(\cdot)$ is *concave* (Theorem 27), its strict concavity is equivalent to the strict concavity of the output differential entropy in $p_Y(\cdot)$. This is indeed the case if and only if $p_Y(\cdot)$ is injective in F. □

The next section is dedicated to the proofs of Theorems 25, 26 and 27.

B.3 Proofs of the Theorems

We use techniques that have been first developed in [57] and later adopted in various works on mutual information maximization as in [49]: denote by $\mathcal{F}$ the space of all probability distribution functions on $\mathbb{R}$. We adopt weak convergence [84, III-1, Def.2, p.311] on $\mathcal{F}$, and use the Levy metric to metrize this weak convergence [30, Th.3.3, p.25]. The optimization is carried out in this metric topology.

Optimization set properties

Theorem 25 (Convexity and Compactness of the Optimization Set)**.** *Whenever conditions C2, C3 and C4 are satisfied, the set $\mathcal{P}_A$ defined in (3.2) is convex and compact.*

Proof. We note first that the theorem was shown to hold for cost functions of the form $\mathcal{C}(|x|) = |x|^r$, for $r > 1$ in [49]. We adopt the same methodologies to generalize the results presented hereafter.

Convexity

Let F_1 and F_2 be two probability distribution functions in $\mathcal{P}_A$, and λ some scalar between 0 and 1. Define $F = \lambda F_1 + (1-\lambda)F_2$. It is clear that F is a probability distribution function because it is non-decreasing, right continuous, $F(-\infty) = 0$ and $F(+\infty) = 1$. Additionally,

$$\begin{aligned}
\int_{\mathbb{R}} \mathcal{C}(|x|)\, dF &= \int_{\mathbb{R}} \mathcal{C}(|x|)\, d\left(\lambda F_1 + (1-\lambda)F_2\right) \\
&= \lambda \int_{\mathbb{R}} \mathcal{C}(|x|)\, dF_1 + (1-\lambda)\int_{\mathbb{R}} \mathcal{C}(|x|)\, dF_2 \\
&\leq \lambda A + (1-\lambda)A = A.
\end{aligned}$$

Therefore, $F \in \mathcal{P}_A$ and $\mathcal{P}_A$ is convex.

Compactness

Consider a random variable X with probability distribution function $F \in \mathcal{P}_A$. Applying Markov's inequality to random variable $\mathcal{C}(|X|)$ yields,

$$\Pr\{\mathcal{C}(|X|) \geq \alpha\} \leq \frac{\mathsf{E}\left[\mathcal{C}(|X|)\right]}{\alpha}, \quad \forall \alpha > 0.$$

Now let

$$K = \inf\left\{x \in [0,\infty) \text{ s.t. } \mathcal{C}(x) \geq \alpha\right\} + 1,$$

which is always greater or equal to 1. For any finite value of α, such a K exists since $\mathcal{C}(x)$ increases to $+\infty$ as $x \to +\infty$ by virtue of properties C3 and C4.

Additionally, since $\mathcal{C}(\cdot)$ is non-decreasing,

$$\begin{aligned}\Pr\{\mathcal{C}\left(|X|\right) \geq \alpha\} &\geq \Pr\left\{|X| > K-1\right\} \geq \Pr\{|X| \geq K\} \\ &\geq F(-K) + \left[1 - F(K)\right].\end{aligned}$$

Hence, for all $F \in \mathcal{P}_A$, we obtain

$$F(-K) + [1 - F(K)] \leq \frac{\mathsf{E}\left[\mathcal{C}\left(|X|\right)\right]}{\alpha} \leq \frac{A}{\alpha}.$$

Therefore, for every $\epsilon > 0$, there exists a $K_\epsilon > 0$, namely

$$K_\epsilon = \min\left\{x \in [0, \infty) \text{ s.t. } \mathcal{C}(x) \geq (A/\epsilon)\right\} + 1,$$

such that

$$\sup_{F \in \mathcal{P}_A} \left[F(-K_\epsilon) + [1 - F(K_\epsilon)]\right] \leq \epsilon.$$

This implies that $\mathcal{P}_A$ is *tight* [84, III-2, Def.2, p.318]. By Phrokhorov's Theorem [84, III-2, Th.1, p.318], $\mathcal{P}_A$ is therefore relatively sequentially compact and every sequence $\{F_n\}$ of distribution functions in $\mathcal{P}_A$ has a convergent sub-sequence $\{F_{n_j}\}$ where the limit F^* does not necessarily belong to $\mathcal{P}_A$. If we prove that $F^* \in \mathcal{P}_A$, the latter will be sequentially compact and hence compact since the space is metrizable [112, Th.28.2, p.179]. In order to show that the limiting distribution function F^* is in $\mathcal{P}_A$, it must satisfy the cost constraint which is the case. In fact,

$$\int \mathcal{C}\left(|u|\right) \, dF^*(u) \leq \liminf_{n_j \to \infty} \int \mathcal{C}\left(|u|\right) \, dF_{n_j} \leq A,$$

where the first inequality holds because $\mathcal{C}(|u|)$ is lower semi-continuous and is bounded from below by $\mathcal{C}\left(0\right)$ for all $u \in \mathbb{R}$ (property C3) [113, Th. 4.4.4]. In addition, the second inequality is valid since the sub-sequence $\{F_{n_j}\}$ is in $\mathcal{P}_A$

and therefore satisfies the cost constraint $\forall\, n_j$. Finally, $F^* \in \mathcal{P}_A$ and $\mathcal{P}_A$ is compact. $\square$

Properties of the mutual information, $I(\cdot)$

We prove in what follows the finiteness, concavity and continuity of $I(\cdot)$ on $\mathcal{P}_A$ through Theorems 26 and 27.

Theorem 26 (Finiteness of the Mutual Information)**.** *Whenever conditions C4, C5 and C6 hold, the mutual information $I(F)$ between the input and output of channel (3.3) is finite for all input distribution functions F such that $\mathsf{E}[\mathcal{C}(|X|)]$ is* finite.

Proof. Since $Y = f(X) + N$,

$$\ln[1 + |Y|] \leq \ln[1 + |f(X)|] + \ln[1 + |N|],$$

and $\mathsf{E}[\ln[1 + |Y|]]$ is finite because both $\mathsf{E}[\ln[1 + |f(X)|]]$ and $\mathsf{E}[\ln[1 + |N|]]$ are finite (by properties C4 and C6).

Moreover, and since $p_Y(y)$ is upperbounded (by C5) by one, the differential entropy of Y, $h_Y(F) = -\int p(y;F) \ln p(y;F)\, dy$, is well defined [82, Proposition 1] and $0 \leq h_Y(y) < +\infty$.

The differential entropy h_N of the noise being finite (due to properties C5 and C6), the mutual information $I(F)$ in (3.18) can therefore be written as the difference of two terms:

$$I(F) = h_Y(F) - h_{Y|X}(F) = h_Y(F) - h_N, \tag{B.7}$$

both of which are finite and this completes the proof. $\square$

Theorem 27 (Concavity and Continuity of the Mutual Information)**.** *Assume* *that conditions C1 through C6 hold. Under a cost constraint*

$$\int \mathcal{C}\left(|X|\right)\, dF(x) \leq A \qquad A > 0,$$

the mutual information $I(F)$ between the input and the output of channel (3.3) is concave and continuous.

Before we proceed with the proof, we note that under the conditions of the theorem, the mutual information $I(F)$ between the input and the output of channel (3.3) is finite by virtue of Theorem 26.

Proof. **Concavity**

The output differential entropy $h_Y(F)$ is a concave function of F on $\mathcal{F}$. In fact,

$$h_Y(F) \; = \; -\int p_Y(y;F)\, \ln\, p_Y(y;F)\, dy$$

exists (by Theorem 26) and is a concave function of $p_Y(\cdot)$ because $-x \ln x$ is concave in x. Since $p_Y(F)$ is linear in F, $I(F) \; = h_Y(F) \; - \; h_N$ is concave on $\mathcal{P}_A$.

Continuity

To prove the continuity of $I(F)$, it suffices to show that $h_Y(F)$ is continuous by virtue of equation (B.7). To this end, we let $F \in \mathcal{P}_A$ and let $\{F_m\}_{m \geq 1}$ be a sequence of probability measures in $\mathcal{P}_A$ that converges weakly to F.

In order to apply Theorem 23 and show the convergence of $h_Y(F_m)$ to $h_Y(F)$ and hence the weak continuity of $h_Y(F)$ on $\mathcal{P}_A$, we establish that the appropriate conditions are satisfied:

- By definition of weak convergence, since $p_N(y-x)$ is bounded and continuous (property C5), then $p(y;F_m) \; = \; \int p_N(y - f(x))\, dF_m(x)$ converges point-wise to $p(y;F) = \int p_N(y - f(x))\, dF(x)$.

- The induced output PDF $p(y; F_m)$ is also bounded by one.

- It remains to find a non-negative and non-decreasing function, $l : [0, \infty) \to [0, \infty)$ such that $l(y) = \omega(\ln(y))$, and a scalar $L > 0$ such that equation (B.2) holds for $p(y; F_m)$, $m \geq 1$ and $p(y; F)$, a task which we fulfill in what follows.

For any $y \geq |f(0)|$, let $\mathcal{S} = f^{-1}([|f(0)|, y])$ be the inverse image by $f(\cdot)$ of the closed interval $[|f(0)|, y]$. Since $f(\cdot)$ is continuous (C1), the set $\mathcal{S}$ is closed. It is also bounded because $|f(x)|$ is non-decreasing in $|x|$ and tends to infinity (A5). Therefore any element in $\mathcal{S}$ is smaller than a positive t_u such that $|f(t_u)| = 2y$ and greater than a negative t_b such that $|f(t_b)| = 2y$. Such t_u and t_b exist because $f(\cdot)$ is continuous.

The set $\mathcal{S}$ is compact and has a maximal value that we denote $z(y) = \max\{z : z \in \mathcal{S}\}$. Note that $|f(z(y))| = y$.

Define the function $\mathcal{C}_{\min}(\cdot)$: $[0, \infty) \longrightarrow \mathbb{R}$ as follows:

$$\mathcal{C}_{\min}(y) = \begin{cases} \min\{\mathcal{C}(z(y)); \mathcal{C}_N(y)\} & y \geq |f(0)| \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathcal{C}_N(\cdot)$ is defined in C6. The function $\mathcal{C}_{\min}(y)$ is non-negative and non-decreasing on $[0, \infty)$ since both $\mathcal{C}(y)$ and $\mathcal{C}_N(\cdot)$ are non-negative and non-decreasing by properties C3 and C6 and $z(y)$ is non-decreasing for $y \geq |f(0)|$. Additionally, $\mathcal{C}_{\min}(y) = \omega(\ln y)$ because $\mathcal{C}(x) = \omega(\ln |f(x)|)$ (A3) and $\mathcal{C}_N(x) = \omega(\ln x)$ (C6).

Now, for any X with distribution $F \in \mathcal{P}_A$,

$$\begin{aligned}
&\mathsf{E}_Y \left[\mathcal{C}_{\min} \left[\frac{|Y|}{2} \right] \right] \\
&= \mathsf{E}_{X,N} \left[\mathcal{C}_{\min} \left[\frac{|f(X)+N|}{2} \right] \right] \\
&\le \mathsf{E}_{X,N} \left[\mathcal{C}_{\min} \left[\frac{|f(X)|+|N|}{2} \right] \right] && \text{(B.8)} \\
&= \mathsf{E}_{X,N} \left[\mathcal{C}_{\min} \left[\frac{|f(X)|+|N|}{2} \right] \Bigg|_{|f(X)| \le |N|} \right] \Pr\left(_{|f(X)| \le |N|}\right) \\
&\quad + \mathsf{E}_{X,N} \left[\mathcal{C}_{\min} \left[\frac{|f(X)|+|N|}{2} \right] \Bigg|_{|f(X)| > |N|} \right] \Pr\left(_{|f(X)| > |N|}\right) \\
&\le \mathsf{E}_{X,N} \left[\mathcal{C}_{\min} \left(|N| \right) \Bigg|_{|f(X)| \le |N|} \right] \Pr\left(_{|f(X)| \le |N|}\right) \\
&\quad + \mathsf{E}_{X,N} \left[\mathcal{C}_{\min} \left(|f(X)| \right) \Bigg|_{|f(X)| > |N|} \right] \Pr\left(_{|f(X)| > |N|}\right) && \text{(B.9)} \\
&\le \mathsf{E}_N \left[\mathcal{C}_{\min} \left(|N| \right) \right] + \mathsf{E}_X \left[\mathcal{C}_{\min} \left(|f(X)| \right) \right] && \text{(B.10)} \\
&\le \mathsf{E}_N \left[\mathcal{C}_N \left(|N| \right) \right] + \mathsf{E}_X \left[\mathcal{C} \left(|X| \right) \right] \\
&\le L_N + A = L < \infty. && \text{(B.11)}
\end{aligned}$$

where $0 \le L_N = \mathsf{E}_N \left[\mathcal{C}_N \left(|N| \right) \right] < \infty$ by property C6. Equations (B.8) and (B.9) are justified since $\mathcal{C}_{\min} \left(|x| \right)$ is non-decreasing in $|x|$ and (B.10) is due to the fact that $\mathcal{C}_{\min} \left(|x| \right)$ is non-negative. Since the value $0 \le L < \infty$ is independent of the input distribution function $F \in \mathcal{P}_A$, then (B.11) holds for any output variable Y, i.e for all $p(y; F)$ where $F \in \mathcal{P}_A$. Letting $l(y) = \mathcal{C}_{\min} \left(\frac{y}{2} \right)$, $y \in [0, \infty)$, then equation (B.2) is satisfied for $p(y; F_m)$, $m \ge 1$ and $p(y; F)$. Therefore, Theorem 23 holds and $h_Y(F_m)$ converges to $h_Y(F)$ and hence $h_Y(F)$ is continuous which concludes the proof. ☐

B.4 Related Publications

The results of this appendix are the subject of an article published in the IEEE Transactions on Communications [114].

Appendix C

Weak Differentiability of $I(\cdot)$ at F^*

Theorem 28 (Weak Differentiability of the Mutual Information)**.** *Let F^* be an optimal input distribution. Under a cost constraint $\int \mathcal{C}\left(|X|\right) dF(x) \leq A$, $A > 0$, the mutual information $I(F)$ between the input and the output of channel (3.3) is weakly differentiable at F^*.*

Before proceeding to the proof, we note that the existence of an optimal F^* and the finiteness of the solution are insured by the results of Appendix B.

Proof. Let θ be a number in $[0, 1]$, $(F^*, F) \in \mathcal{P}_A \times \mathcal{P}_A$ and define $F_\theta = (1 - \theta)F^* + \theta F$. The weak derivative of $I(.)$ at F^* in the direction of F is defined as,

$$I'(F^*, F) \triangleq \lim_{\theta \to 0^+} \frac{I(F_\theta) - I(F^*)}{\theta},$$

whenever the limit exists. For simplicity, we denote by

$$t(x) = i(x; F^*),$$

where $i(x;F)$ is given by equation (3.11), and we prove

$$\begin{aligned} I'(F^*,F) &= -\int p(y;F)\ln p(y;F^*)\,dy \;-\; h_Y(F^*) \\ &= \int t(f(x))\,dF(x) \;-\; h_Y(F^*), \end{aligned}$$

where by Tonelli, the interchange is valid as long as the integral term is finite which we prove next. Using L'Hôpital's rule,

$$\begin{aligned} I'(F^*,F) &= \lim_{\theta\to 0^+}\frac{I(F_\theta)-I(F^*)}{\theta} = \lim_{\theta\to 0^+}\frac{h_Y(F_\theta)-h_Y(F^*)}{\theta} \\ &= \lim_{\theta\to 0^+} -\left[\int p(y;F_\theta)\ln p(y;F_\theta)\,dy\right]', \end{aligned} \tag{C.1}$$

where the derivative is with respect to θ. In order to evaluate $\left[\int p(y;F_\theta)\ln p(y;F_\theta)\,dy\right]'$ we use the definition of the derivative

$$\begin{aligned} &\left[\int p(y;F_\theta)\ln p(y;F_\theta)\,dy\right]' \\ &= \lim_{h\to 0}\left[\frac{\int p(y;F_{\theta+h})\ln p(y;F_{\theta+h})\,dy}{h} - \frac{\int p(y;F_\theta)\ln p(y;F_\theta)\,dy}{h}\right], \end{aligned}$$

where by the limit we mean that both, the limit as h goes to 0^+ and the limit as h goes to 0^- exist and are equal. In what follows, we only provide detailed evaluations as h goes to 0^+ since those when h goes to 0^- are similar. Using the mean value theorem, for some $0 \le c(h) \le h$,

$$\begin{aligned} &\lim_{h\to 0+}\left[\frac{\int p(y;F_{\theta+h})\ln p(y;F_{\theta+h})\,dy}{h} - \frac{\int p(y;F_\theta)\ln p(y;F_\theta)\,dy}{h}\right] \\ &= \lim_{h\to 0^+}\int \left[p(y;F_\theta)\ln p(y;F_\theta)\right]'_{|\theta+c(h)}\,dy. \end{aligned}$$

Now, since $p(y;F_\theta) = p(y;F^*) + \theta\,[p(y;F) - p(y;F^*)]$,

$$
\begin{aligned}
&\lim_{h\to 0^+} \int [p(y;F_\theta)\ln p(y;F_\theta)]'_{|\theta+c(h)}\,dy \\
&= \lim_{h\to 0^+} \int [p(y;F) - p(y;F^*)] \ln p(y;F_{\theta+c(h)})\,dy + \int [p(y;F) - p(y;F^*)]\,dy \\
&= \int \lim_{h\to 0^+} [p(y;F) - p(y;F^*)] \ln p(y;F_{\theta+c(h)})\,dy \qquad \text{(C.2)} \\
&= \int [p(y;F) - p(y;F^*)] \ln p(y;F_\theta)\,dy, \qquad \text{(C.3)}
\end{aligned}
$$

where (C.3) is due to the fact that $c(h) \to 0$ as $h \to 0$ and that $p(y;F_\theta)$ is continuous in θ by virtue of its linearity, and (C.2) is due to DCT. Indeed,

$$
\left|[p(y;F) - p(y;F^*)] \ln p(y;F_{\theta+c(h)})\right| \le \big(p(y;F) + p(y;F^*)\big) \left|\ln p(y;F_{\theta+c(h)})\right|,
$$

and

$$
\begin{aligned}
p(y;F_{\theta+c(h)}) &= [1-\theta-c(h)]\,p(y;F^*) + [\theta+c(h)]\,p(y;F) \\
&\ge [1-\theta-c(h)]\,p(y;F^*) \ge \frac{1}{2}\,p(y;F^*),
\end{aligned}
$$

whenever $\theta + c(h) \le \frac{1}{2}$, which is true since both θ and $c(h)$ are arbitrarily small. Therefore, since $0 < p(y;F) < 1$ for all F

$$
\left|[p(y;F) - p(y;F^*)] \ln p(y;F_{\theta+c(h)})\right| \le -\big(p(y;F) + p(y;F^*)\big) \ln\left[\frac{1}{2}\,p(y;F^*)\right].
$$

Since $h_Y(F) = -\int p(y;F)\ln p(y;F)\,dy$ is finite for all F in $\mathcal{P}_A$ [115, Theorem 2], $-p(y;F^*)\ln p(y;F^*)$ is integrable. It remains to prove that $-p(y;F)\ln p(y;F^*)$ is integrable to justify (C.2) and hence (C.3). To this end, we will proceed by choosing first a specific $F(\cdot)$, namely

$$F_s(x) = \left(1 - \frac{B_s}{\mathcal{C}(x_s)}\right) u(x)^1 + \frac{B_s}{\mathcal{C}(x_s)} u(x - x_s),$$

for some $x_s > 0$ such that $\mathcal{C}(x_s) > 0$ and where $(0 <)\, B_s < \min\{A; \mathcal{C}(x_s)\}$. We note that $F_s \in \mathcal{P}_A$ since $\mathcal{C}(0) = 0$ and hence $\int \mathcal{C}(|x|) dF_s = B_s \leq A$. If F_s were the input distribution, it would induce the following output

$$p(y; F_s) = \left(1 - \frac{B_s}{\mathcal{C}(x_s)}\right) p_N(y) + \frac{B_s}{\mathcal{C}(x_s)} p_N(y - f(x_s)). \tag{C.4}$$

Equation (C.4) along with Lemma 2 and properties C7 and C8 (see Chapter 3) show that $-p(y; F_s) \ln p(y; F^*)$ is integrable and (C.3) is justified for $F \equiv F_s$. Hence,

$$\begin{aligned} & I'(F^*, F_s) \\ &= \lim_{\theta \to 0^+} - \left[\int p(y; F_\theta^*) \ln p(y; F_\theta^*)\, dy \right]' \\ &= \lim_{\theta \to 0^+} \int [p(y; F_s) - p(y; F^*)] \ln p(y; F_\theta^*)\, dy = \int t(f(x)) dF_s(x) \; - \; h_Y(F^*). \end{aligned}$$

where the interchange between the limit and integral sign is justified in an identical fashion as done to validate (C.3).

Now, since F^* is optimal, necessarily $I'(F^*, F_s) \leq 0$ (see Appendix C in [49]), which implies that

$$\int t(f(x))\, dF_s(x) \leq h_Y(F^*).$$

[1]where $u(x)$ denotes the Heaviside unit step function.

Plugging in the expression of $F_s(x)$ yields,

$$\left(1-\frac{B_s}{\mathcal{C}(x_s)}\right)t(f(0))+\frac{B_s}{\mathcal{C}(x_s)}t(f(x_s))\leq h_Y(F^*)$$
$$\Leftrightarrow \quad t(f(x_s))\leq\frac{h_Y(F^*)-t(f(0))}{B_s}\mathcal{C}(x_s)+t(f(0)). \tag{C.5}$$

The above equation is valid for any $x_s > 0$ (such that $\mathcal{C}(x_s) > 0$) and therefore for all $|x| \geq x_s$ since $\mathcal{C}(|x|)$ is non-decreasing in $|x|$. we proceed by writing

$$\int t(f(x))\,dF=\int_{|x|\leq x_s} t\left(f(x)\right)\,dF+\int_{|x|>x_s} t(f(x))\,dF.$$

As for the first integral term, we have:

$$\begin{aligned}
&\int_{|x|\leq x_s} t\left(f(x)\right)\,dF\\
&=-\int_{|x|\leq x_s}\int p_N(y-f(x))\,\ln\,p(y;F^*)\,dy\,dF\\
&=-\int_{|x|\leq x_s}\int_{|y|\geq y_0} p_N(y-f(x))\,\ln\,p(y;F^*)\,dy\,dF\\
&\quad-\int_{|x|\leq x_s}\int_{|y|\leq y_0} p_N(y-f(x))\,\ln\,p(y;F^*)\,dy\,dF
\end{aligned} \tag{C.6}$$

Using Lemma 2 and property C7 defined in Chapter 3, the first term of equation (C.6) is finite. As for the second term, it is finite by the fact that $p(y;F^*)$ is positive and continuous hence achieves a positive minimum on compact subsets of $\mathbb{R}$. When it comes to the range $|x| > x_s$, we use the upperbound in (C.5) which gives:

$$\begin{aligned}
\int_{|x|>x_s} t(f(x))\,dF &\leq \int_{|x|>x_s}\left(\frac{h_Y(F^*)-t(f(0))}{B_u}\mathcal{C}(|x|)+t(f(0))\right)\,dF\\
&\leq \frac{h_Y(F^*)-t(f(0))}{B_u}A+t(f(0)),
\end{aligned}$$

which is finite.

In conclusion,

$$-\int p(y;F)\ln p(y;F^*)\,dy = \int t(f(x))\,dF < \infty,$$

and $I'(F^*,F) = \int t(f(x))\,dF - h_Y(F^*)$, $\forall\, F \in \mathcal{P}_A$. $\square$

Cost

The mapping from $\mathcal{F}$ to $\mathbb{R}$:

$$\mathcal{T}(F) = \int \mathcal{C}\left(|x|\right)\,dF - A$$

is weakly differentiable on $\mathcal{P}_A$ as well. In fact,

$$\mathcal{T}'(F^*,F) = \mathcal{T}(F) - \mathcal{T}(F^*),$$

which is finite, since $-A < \mathcal{T}(F) \leq 0$ for all $F \in \mathcal{P}_A$.

Appendix D

Rate of Decay of alpha-stable density functions on the Horizontal Strip

We study in this appendix the rate of decay of alpha-stable distributions $S(\alpha, \beta, \gamma, \delta)$ on the horizontal strip $\mathcal{S}_\eta = \{z \in \mathbb{C} : |\Im(z)| < \eta\}$ where η is a small-enough positive number. The study is limited to the case: $\alpha \in [1, 2)$, $\beta \in]-1, 1[$, $\gamma \in \mathbb{R}^{+*}$ and $\delta \in \mathbb{R}$. We prove the following:

Theorem (Rate of Decay of Alpha-Stable PDFs on the Horizontal Strip)**.** *Let $N \sim S(\alpha, \beta, \gamma, \delta)$, $\alpha \in [1, 2)$, $\beta \in]-1, 1[$ and denote by $p_N(x)$ its PDF. Let $p_N(z)$ be the analytical extension of $p_N(x)$ to $\mathcal{S}_\eta$. Then $|p_N(z)| = O\left(\frac{1}{|\Re(z)|^{\alpha+1}}\right)$, for $z \in \mathcal{S}_\eta$ as $|\Re(z)| \to \infty$.*

Before we proceed, we first prove the following Lemma:

Lemma 10 (Extension to the Horizontal Strip)**.** *Whenever $N \sim S(\alpha, \beta, \gamma, \delta)$, where $\alpha \in [1, 2)$, $\beta \in]-1, 1[$, $\gamma \in \mathbb{R}^{+*}$ and $\delta \in \mathbb{R}$, $p_N(\cdot)$ can be formally extended on $\mathcal{S}_\eta = \{z \in \mathbb{C} : |\Im(z)| < \eta\}$ as*

$$p_N(z) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-izt} \phi(t) dt. \tag{D.1}$$

Proof. By definition,

$$p_N(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ixt}\phi(t)dt,$$

where

$$\begin{aligned}\phi(t) &= \exp\left[i\delta t - \gamma^\alpha \left[1 - i\beta\,\mathrm{sgn}(t)\Phi(t)\right]|t|^\alpha\right] \\ \Phi(t) &= \begin{cases} \tan\left(\dfrac{\pi\alpha}{2}\right) & \alpha \neq 1 \\ -\dfrac{2}{\pi}\ln|t| & \alpha = 1. \end{cases}\end{aligned}$$

Let $p_N(z)$ be the extension of $p_N(x)$ on $\mathbb{C}$. It is known that $p_N(z)$ is analytic on $\mathcal{S}_\eta$ (see [16] for example) . Now, define

$$q(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-izt}\phi(t)dt,$$

for all $z = (x + iy) \in \mathbb{C}$. If we establish that $q(z)$ is analytic on $\mathcal{S}_\eta$ then by the identity theorem, $p_N(z) = q(z)$, for all $z \in \mathcal{S}_\eta$. We start by proving the continuity of $q(z)$:

$$\begin{aligned}\lim_{z\to z_0} q(z) = &\lim_{z\to z_0}\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-izt}\phi(t)dt \\ =&\frac{1}{2\pi}\int \lim_{z\to z_0} e^{-izt}\phi(t)dt \qquad \text{(D.2)} \\ =&\frac{1}{2\pi}\int e^{-iz_0t}\phi(t)dt = q(z_0).\end{aligned}$$

where the interchange in (D.2) is justified by DCT since:

$$\left|e^{-izt}\phi(t)\right| \leq e^{yt-|\gamma t|^\alpha},$$

which is integrable on $\mathcal{S}_\eta$ since η is small-enough and chosen so that $|y| < \eta \leq \gamma^\alpha$. Now, let $\Delta \subset \mathcal{S}_\eta$ be a compact triangle and denote by $\partial\Delta$ its boundary and $|\Delta|$

its perimeter. We obtain

$$
\begin{aligned}
\int_{\partial\Delta} q(z)dz &= \frac{1}{2\pi}\int_{\partial\Delta}\int_{\mathbb{R}} e^{-izt}\phi(t)\,dt\,dz \\
&= \frac{1}{2\pi}\int_{\mathbb{R}}\int_{\partial\Delta} e^{-izt}\phi(t)\,dz\,dt \\
&= \int_{\mathbb{R}} \phi(t)\int_{\partial\Delta} e^{-izt}dz = 0,
\end{aligned}
\tag{D.3}
$$

where the last equation is due to the fact that e^{-izt} is entire. The interchange in (D.3) is valid by Fubini since

$$
\frac{1}{2\pi}\int_{\partial\Delta}\int_{\mathbb{R}}\left|e^{-izt}\phi(t)\right|\,dt\,dz \;\le\; \frac{1}{2\pi}\int_{\partial\Delta}\int_{\mathbb{R}} e^{yt-|\gamma t|^{\alpha}}\,dt\,dz \;<\; \frac{|\Delta|}{2\pi}\int_{\mathbb{R}} e^{yt-|\gamma t|^{\alpha}}|\,dt \;<\; \infty.
$$

By applying Morera's Theorem [87, sec. 53], $q(z)$ is analytic on $\mathcal{S}_\eta$ and the result is established. ☐

Note that equation (D.1) shows that $p_N(z) = p_{N'}(z-\delta)$ where $N' \sim S(\alpha,\beta,\gamma,0)$. Therefore, and without loss of generality, we restrict our analysis in the remainder of this section to $p_N(z)$, for $N \sim S(\alpha,\beta,\gamma,0)$.

For $z = (x+iy)$,

$$
\begin{aligned}
p_N(z) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-izt-\gamma^{\alpha}(1-i\beta\,\mathrm{sgn}(t)\Phi)|t|^{\alpha}}dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ixt+yt-\gamma^{\alpha}[1-i\beta\,\mathrm{sgn}(t)\Phi(t)]|t|^{\alpha}}dt \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ixt-\gamma^{\alpha}[1-i\beta\,\mathrm{sgn}(t)\Phi(t)]|t|^{\alpha}}\sum_{n=0}^{\infty}\frac{y^n}{n!}\,t^n\,dt \\
&= \frac{1}{2\pi}\sum_{n=0}^{\infty}\frac{y^n}{n!}\int_{-\infty}^{\infty} t^n e^{-ixt-\gamma^{\alpha}[1-i\beta\,\mathrm{sgn}(t)\Phi(t)]|t|^{\alpha}}\,dt.
\end{aligned}
\tag{D.4}
$$

The interchange in (D.4) is justified by DCT. Indeed,

$$
\left|\sum_{n=0}^{N}\frac{y^n}{n!}\,t^n e^{-ixt-\gamma^{\alpha}[1-i\beta\,\mathrm{sgn}(t)\Phi(t)]|t|^{\alpha}}\right| \le \sum_{n=0}^{\infty}\frac{|y|^n}{n!}\,|t|^n e^{-|\gamma t|^{\alpha}} = e^{|y||t|-|\gamma t|^{\alpha}},
$$

which is integrable for $|y| < \eta$ ($\le \gamma^{\alpha}$) and $\alpha \ge 1$. Now we proceed to studying the rate of decay in two separate cases.

D.1 Rate of Decay for $1 < \alpha < 2$:

In this case $\Phi(t)$ is a constant and it is equal to $\Phi(t) = \Phi = \tan\left(\frac{\pi\alpha}{2}\right)$. Then, using equation (D.4), we obtain by the change of variable $u = \gamma t$

$$\begin{aligned} p_N(z) &= \frac{1}{2\pi\gamma}\sum_{n=0}^{\infty}\frac{1}{n!}\left(\frac{y}{\gamma}\right)^n \int_{-\infty}^{\infty} t^n e^{-i\frac{x}{\gamma}t-[1-i\beta\,\mathrm{sgn}(t)\Phi]|t|^\alpha}\,dt \\ &= \frac{1}{2\pi\gamma}\sum_{n=0}^{\infty}\frac{1}{n!}\left(\frac{y}{\gamma}\right)^n T_n\left(-\frac{x}{\gamma};\beta\right), \end{aligned} \tag{D.5}$$

where $T_n(x;\beta)$ is a function defined as $T_n(x;\beta) \hat{=} \displaystyle\int_{-\infty}^{\infty} t^n e^{ixt-[1-i\beta\,\mathrm{sgn}(t)\Phi]|t|^\alpha}dt$.[1] Define $k_1 = (1 - i\beta\Phi)$ and denote by $\overline{k}_1 = (1 + i\beta\Phi)$ its conjugate. In what follows, we study the behavior of the function $T_n(x;\beta)$.

---

[1]Note that $T_n(-x;\beta) = (-1)^n T_n(x;-\beta)$ and that $p_N^{(n)}(x) = \frac{1}{2\pi}\frac{(-i)^n}{\gamma^{n+1}}T_n(-\frac{x}{\gamma};\beta) = \frac{1}{2\pi}\frac{i^n}{\gamma^{n+1}}T_n(\frac{x}{\gamma};-\beta)$, $n \in \mathbb{N}^*$.

For $n \geq 1$ and $x > 0$, we have

$$\begin{aligned}
&x^{n+\alpha+1} T_n(x;\beta) \\
&= x^{n+\alpha}\left[\int_0^\infty x t^n e^{ixt-k_1 t^\alpha} dt + (-1)^n \int_0^\infty x t^n e^{-ixt-\overline{k}_1 t^\alpha} dt\right] \\
&= -i\, x^{n+\alpha}\left[\int_0^\infty t^n \left(e^{ixt-k_1 t^\alpha}\right)' dt + k_1\,\alpha \int_0^\infty t^{n+\alpha-1} e^{ixt-k_1 t^\alpha} dt\right. \\
&\qquad \left. + (-1)^{n-1} \int_0^\infty t^n \left(e^{-ixt-\overline{k}_1 t^\alpha}\right)' dt + (-1)^{n-1}\overline{k}_1\,\alpha \int_0^\infty t^{n+\alpha-1} e^{-ixt-\overline{k}_1 t^\alpha} dt\right] \\
&= inx^{n+\alpha}\left[\int_0^\infty t^{n-1} e^{ixt-k_1 t^\alpha} dt + (-1)^{n-1}\int_0^\infty t^{n-1} e^{-ixt-\overline{k}_1 t^\alpha} dt\right] \\
&\quad - i\alpha x^{n+\alpha}\left[k_1 \int_0^\infty t^{n+\alpha-1} e^{ixt-k_1 t^\alpha} dt + (-1)^{n-1}\overline{k}_1 \int_0^\infty t^{n+\alpha-1} e^{-ixt-\overline{k}_1 t^\alpha} dt\right] \qquad \text{(D.6)} \\
&= i\, n\, x^{n+\alpha} T_{n-1}(x;\beta) - i\alpha\left[k_1 S_n(x;k_1) + (-1)^{n-1}\overline{k}_1 \overline{S_n}(x;\overline{k}_1)\right], \qquad \text{(D.7)}
\end{aligned}$$

where equation (D.6) is obtained by integration by parts and regrouping, and where $\overline{S_n}(\cdot;\cdot)$ is the complex conjugate of $S_n(\cdot;\cdot)$ defined as,

$$S_n(x;k_1) = x^{n+\alpha}\int_0^\infty t^{n+\alpha-1} e^{ixt-k_1 t^\alpha} dt = c\int_0^\infty e^{iv^c - k_1 \zeta v^{\alpha c}} dv,$$

where $c = \frac{1}{n+\alpha}\,(> 0)$, $\zeta = x^{-\alpha}\,(> 0)$ and the change of variable is $v = (xt)^{n+\alpha}$. As $x \to \infty$, $\zeta \to 0^+$ and hence

$$\begin{aligned}
\lim_{x\to+\infty} S_n(x;k_1) &= c\lim_{\zeta\to 0^+}\int_0^\infty e^{iv^c - k_1\zeta v^{\alpha c}}\, dv = c\lim_{\zeta\to 0^+}\int_0^\infty \lim_{\theta\to 0} e^{iv^c e^{ic\theta} - k_1 \zeta v^{\alpha c} e^{i\alpha c\theta} + i\theta}\, dv \\
&= c\lim_{\zeta\to 0^+}\lim_{\theta\to 0}\int_0^\infty e^{iv^c e^{ic\theta} - k_1 \zeta v^{\alpha c} e^{i\alpha c\theta} + i\theta}\, dv \qquad \text{(D.8)} \\
&= c\lim_{\theta\to 0}\lim_{\zeta\to 0^+}\int_0^\infty e^{iv^c e^{ic\theta} - k_1 \zeta v^{\alpha c} e^{i\alpha c\theta} + i\theta}\, dv \qquad \text{(D.9)} \\
&= c\lim_{\theta\to 0}\int_0^\infty e^{iv^c e^{ic\theta} + i\theta}\, dv \qquad \text{(D.10)} \\
&= c\lim_{\theta\to 0}\lim_{R\to\infty,\rho\to 0}\int_{\mathrm{L}_1} e^{iz^c}\, dz,
\end{aligned}$$

where $z = ve^{i\theta}$ and $\mathrm{L}_1 = \{z \in \mathbb{C} : z = ve^{i\theta},\, 0 < \rho \le v \le R\}$. Equation (D.8) is justified by DCT since:

$$\left| e^{iv^c e^{ic\theta} - k_1 \zeta v^{\alpha c} e^{i\alpha c\theta} + i\theta} \right| \le e^{-v^c \sin(c\theta) - \zeta v^{\alpha c}[\cos(\alpha c\theta) + \beta\Phi \sin(\alpha c\theta)]} \le e^{-\frac{\zeta}{2} v^{\alpha c}},$$

for small-enough θ, and the upper-bound is integrable since c and ζ are positive. The last inequality is justified by virtue of the fact that $\sin(c\theta) > 0$ and $\left[\cos(\alpha c\theta) + \beta \tan\frac{\alpha\pi}{2} \sin(\alpha c\theta)\right] > \frac{1}{2}$ for small positive θ. Similarly, (D.10) is justified because the integrand in (D.9) is $O(e^{-v^c \sin c\theta})$ as $\zeta \to 0^+$ which is also integrable. The interchange between the two limits in (D.9) is valid by the preceding argument as long as the result in (D.10) is finite. To evaluate the limit of $\int_{\mathrm{L}_1} e^{iz^c} dz$ as $R \to \infty$, $\rho \to 0$, we use contour integration over $\mathcal{C}$ shown in Figure D-1.

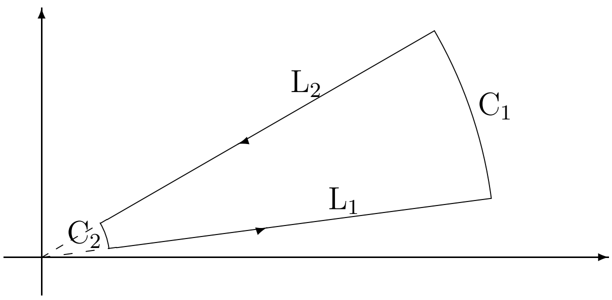

Figure D-1: The contour $\mathcal{C}$.

The arcs C_1 and C_2 are of radius R, and ρ respectively and are between angles θ and $\varphi \triangleq \frac{\pi}{2c} \bmod 2\pi$. Note that since we are interested in the limit as θ goes to zero, we can always choose it small enough in order to have the contour counter-clockwise. Finally, L_2 is a line connecting the extremities of the arcs.

Now since $f(z) = e^{iz^c}$ is analytic on and inside $\mathcal{C}$ (by choosing an appropriate branch cut in the plane), by Cauchy's Theorem [116, p.111 Sec.2.2],

$$0 = \oint_{\mathcal{C}} f(z)\, dz = \int_{\mathrm{L}_1} f(z) + \int_{\mathrm{C}_1} f(z) + \int_{\mathrm{L}_2} f(z) + \int_{\mathrm{C}_2} f(z).$$

On C_1, we have:

$$\lim_{R\to\infty}\left|\int_{\mathrm{C}_1} f(z)dz\right| = \lim_{R\to\infty}\left|\int_\theta^\varphi iRe^{i\phi}e^{iR^c e^{ic\phi}}\, d\phi\right|$$
$$\leq \lim_{R\to\infty}\int_\theta^\varphi Re^{-R^c\sin(c\phi)}d\phi = \int_\theta^\varphi \lim_{R\to\infty} Re^{-R^c\sin(c\phi)}d\phi = 0,$$

where the interchange is valid because $Re^{-R^c\sin(c\phi)}$ is decreasing as $0 < c\theta \leq c\phi \leq \frac{\pi}{2}$. Similarly, on C_2,

$$\lim_{\rho\to 0}\left|\int_{\mathrm{C}_2} f(z)dz\right| = \lim_{\rho\to 0}\left|\int_\theta^\varphi i\rho e^{i\phi}e^{i\rho^c e^{ic\phi}}\, d\phi\right|$$
$$\leq \lim_{\rho\to 0}\int_\theta^\varphi \rho e^{-\rho^c\sin(c\phi)}d\phi = \int_\theta^\varphi \lim_{\rho\to 0} \rho e^{-\rho^c\sin(c\phi)}d\phi = 0,$$

where we justify the interchange by virtue of the fact that $\rho e^{-\rho^c\sin(c\phi)}$ is bounded for small values of ρ. It remains to evaluate the integral on L_2 where $z = te^{i\frac{\pi}{2c}}$,

$$\lim_{R\to\infty,\rho\to 0}\int_{\mathrm{L}_2} f(z)dz = -\int_0^\infty e^{i\frac{\pi}{2c}}e^{it^c e^{i\frac{\pi}{2}}}\, dt = -e^{i\frac{\pi}{2c}}\int_0^\infty e^{-t^c}\, dt = -e^{i\frac{\pi}{2c}}\,\frac{1}{c}\,\Gamma\left(\frac{1}{c}\right).$$

In conclusion,

$$\lim_{R\to\infty,\rho\to 0}\int_{\mathrm{L}_1} f(z)\, dz = e^{i\frac{\pi}{2c}}\,\frac{1}{c}\,\Gamma\left(\frac{1}{c}\right),$$

which implies that

$$\lim_{x\to+\infty} S_n(x;k_1) = e^{i\frac{\pi}{2}(n+\alpha)}\Gamma(n+\alpha),$$

and by (D.7), we can write for $n \geq 1$

$$\lim_{x\to+\infty}\left[x^{n+\alpha+1}T_n(x;\beta) - i\, nx^{n+\alpha}T_{n-1}(x;\beta)\right]$$
$$= W_n(\beta) \hat{=}\ -i\alpha\Gamma(n+\alpha)\left[k_1 e^{i\frac{\pi}{2}(n+\alpha)} + (-1)^{n-1}\overline{k}_1 e^{-i\frac{\pi}{2}(n+\alpha)}\right],$$

which implies that $U_n(\beta) \hat{=} \lim\limits_{x\to+\infty} x^{n+\alpha+1}T_n(x;\beta)$ is a well defined quantity because

$$U_0(\beta) = \lim_{x\to+\infty}\left[x^{\alpha+1}T_0(x;\beta)\right] = 2\pi\gamma \lim_{x\to+\infty}\left[x^{\alpha+1}p_N(-\gamma x)\right],$$

exists –and is non zero for $\beta \neq 1$ and $U_0(1) = 0$, and

$$U_n(\beta) = \; inU_{n-1}(\beta) + W_n(\beta) = n!\left[i^n\, U_0(\beta) + \sum_{k=0}^{n-1}\frac{i^k}{(n-k)!}W_{n-k}(\beta)\right].$$

Furthermore, for $n \geq 0$,

$$\begin{aligned}
|U_n(\beta)| &\leq n!\left[|U_0(\beta)| + \sum_{k=0}^{n-1}\frac{|W_{n-k}(\beta)|}{(n-k)!}\right] \\
&\leq n!\left[|U_0(\beta)| + 2\alpha|k_1|\sum_{k=0}^{n-1}\frac{\Gamma(n+\alpha-k)}{(n-k)!}\right] \\
&\leq n!\left[|U_0(\beta)| + 4|k_1|\sum_{k=0}^{n-1}\frac{\Gamma(n+2-k)}{(n-k)!}\right] \qquad \text{(D.11)} \\
&= n!\left[|U_0(\beta)| + 4|k_1|\sum_{k=0}^{n-1}(n+1-k)\right] \\
&= 2\,n!\left(|k_1|n^2 + 3|k_1|n + \frac{|U_0(\beta)|}{2}\right),
\end{aligned}$$

where equation (D.11) is justified using the fact that $0 < \alpha < 2$ and $\Gamma(\alpha + l)$ is increasing in $\alpha > 0$ for $l \in \mathbb{N}^*$.

Now using equation (D.5),

$$\begin{aligned}
\lim_{x\to\infty} x^{\alpha+1}\, |p_N(z)| &= \frac{1}{2\pi\gamma} \lim_{x\to\infty} x^{\alpha+1} \left| \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{y}{\gamma}\right)^n T_n\left(-\frac{x}{\gamma};\beta\right) \right| \\
&= \frac{1}{2\pi\gamma} \left| \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{y}{\gamma}\right)^n \lim_{x\to\infty} x^{\alpha+1} T_n\left(-\frac{x}{\gamma};\beta\right) \right| \qquad \text{(D.12)} \\
&\le \frac{1}{2\pi\gamma} \sum_{n=0}^{\infty} \frac{1}{n!} \left|\frac{y}{\gamma}\right|^n \lim_{x\to\infty} x^{\alpha+1} \left| T_n\left(\frac{x}{\gamma};-\beta\right) \right| \\
&\le \frac{1}{2\pi\gamma} \sum_{n=0}^{\infty} \frac{1}{n!} \left|\frac{y}{\gamma}\right|^n \lim_{x\to\infty} x^{n+\alpha+1} \left| T_n\left(\frac{x}{\gamma};-\beta\right) \right| = \frac{1}{2\pi\gamma} \sum_{n=0}^{\infty} \frac{1}{n!} \left|\frac{y}{\gamma}\right|^n \gamma^{n+\alpha+1} |U_n(-\beta)| \\
&\le \frac{\gamma^\alpha}{\pi} \sum_{n=0}^{\infty} |y|^n \left(|k_1| n^2 + 3|k_1| n + \frac{|U_0(-\beta)|}{2} \right),
\end{aligned}$$

which is finite because $|y| < \eta$ which is small-enough (and assumed to be less than one), and where we used the fact that $f(x) = |x|$ is continuous. The interchange in (D.12) is valid because the end result is finite.

In conclusion, $\lim_{x\to+\infty} x^{\alpha+1}\, |p_N(z)| < \infty$ which concludes our proof.

D.2 Rate of Decay for $\alpha = 1$:

In this case, $\Phi(t) = -\frac{2}{\pi}\log|t|$ is a function of t. According to equation (D.4) and for $z = x + iy$,

$$p_N(z) = \frac{1}{2\pi} \sum_{n=0}^{\infty} \frac{y^n}{n!} \int_{-\infty}^{\infty} t^n e^{-ixt-\gamma\left[1-i\beta\,\mathrm{sgn}(t)\Phi(t)\right]|t|}\, dt. \qquad \text{(D.13)}$$

Once more, we study the behavior of the integral

$$I_n(x) = \int_{-\infty}^{\infty} t^n e^{-ixt-\gamma(1-i\beta\,\mathrm{sgn}(t)\Phi(t))|t|}\, dt \qquad \text{(D.14)}$$

as $x \to \infty$ for $n \ge 0$. We note that $I_0(x) = 2\pi p_N(x;1,\beta,\gamma,0)$ which is $\Theta\left(\frac{1}{x^2}\right)$.

For $n \geq 1$,

$$
\begin{aligned}
& I_n(x) \\
&= \int_{-\infty}^{+\infty} t^n e^{-ixt-\gamma(1-i\beta\,\mathrm{sgn}(t)\Phi(t))|t|}\,dt \\
&= \int_0^{+\infty} t^n e^{-ixt-\gamma\left(1+i\frac{2}{\pi}\beta\log(t)\right)t}\,dt + \int_{-\infty}^0 t^n e^{-ixt-\gamma\left(1-i\frac{2}{\pi}\beta\log(-t)\right)(-t)}\,dt \\
&= \int_0^{+\infty} e^{-ixt}\, t^n e^{-\gamma\left(1+i\frac{2}{\pi}\beta\log(t)\right)t}\,dt + (-1)^n \int_0^{+\infty} e^{ixt}\, t^n e^{-\gamma\left(1-i\frac{2}{\pi}\beta\log(t)\right)t}\,dt \\
&= \left[-\frac{1}{ix}e^{-ixt}\, t^n e^{-\gamma\left(1+i\frac{2}{\pi}\beta\log(t)\right)t}\right]_0^{+\infty} + (-1)^n\left[\frac{1}{ix}e^{ixt}\, t^n e^{-\gamma\left(1-i\frac{2}{\pi}\beta\log(t)\right)t}\right]_0^{+\infty} \\
&\quad + \frac{1}{ix}\int_0^{+\infty} e^{-ixt}\left[nt^{n-1}-\gamma t^n - i\frac{2}{\pi}\beta\gamma t^n - i\frac{2}{\pi}\beta\gamma t^n\log(t)\right] e^{-\gamma\left(1+i\frac{2}{\pi}\beta\log(t)\right)t}\,dt \\
&+ \frac{(-1)^{n+1}}{ix}\int_0^{+\infty} e^{ixt}\left[nt^{n-1}-\gamma t^n + i\frac{2}{\pi}\beta\gamma t^n + i\frac{2}{\pi}\beta\gamma t^n\log(t)\right] e^{-\gamma\left(1-i\frac{2}{\pi}\beta\log(t)\right)t}\,dt
\end{aligned}
\tag{D.15}
$$

$$
\begin{aligned}
&= \frac{1}{ix}\int_0^{+\infty} e^{-ixt}\left[nt^{n-1}-\gamma t^n - i\frac{2}{\pi}\beta\gamma t^n - i\frac{2}{\pi}\beta\gamma t^n\log(t)\right] e^{-\gamma\left(1+i\frac{2}{\pi}\beta\log(t)\right)t}\,dt \\
&+ \frac{(-1)^{n+1}}{ix}\int_0^{+\infty} e^{ixt}\left[nt^{n-1}-\gamma t^n + i\frac{2}{\pi}\beta\gamma t^n + i\frac{2}{\pi}\beta\gamma t^n\log(t)\right] e^{-\gamma\left(1-i\frac{2}{\pi}\beta\log(t)\right)t}\,dt \\
&= \left[\frac{1}{x^2}e^{-ixt}\left[nt^{n-1}-\gamma t^n - i\frac{2}{\pi}\beta\gamma t^n - i\frac{2}{\pi}\beta\gamma t^n\log(t)\right] e^{-\gamma\left(1+i\frac{2}{\pi}\beta\log(t)\right)t}\right]_0^{+\infty} \\
&\qquad - \frac{1}{x^2}\int_0^{+\infty} e^{-ixt} g_n(t)\,dt \\
&+ (-1)^{n+1}\left[-\frac{1}{x^2}e^{ixt}\left[nt^{n-1}-\gamma t^n + i\frac{2}{\pi}\beta\gamma t^n + i\frac{2}{\pi}\beta\gamma t^n\log(t)\right] e^{-\gamma\left(1-i\frac{2}{\pi}\beta\log(t)\right)t}\right]_0^{+\infty} \\
&\qquad + \frac{(-1)^{n+1}}{x^2}\int_0^{+\infty} e^{ixt} h_n(t)\,dt
\end{aligned}
\tag{D.16}
$$

$$
= \frac{1}{x^2}\left((-1)^{n+1}\int_0^{+\infty} e^{ixt}\, h_n(t)\,dt - \int_0^{+\infty} e^{-ixt} g_n(t)\,dt\right)
\tag{D.17}
$$

where equations (D.15) and (D.16) are due to integration by parts. The functions $g_n(\cdot)$ and $h_n(\cdot)$, $n \geq 1$ are defined on $\mathbb{R}^{+*}$ and are given by:

$$
\begin{aligned}
g_n(t) = \Bigg[& n(n-1)t^{n-2} - 2n\gamma t^{n-1} + (\gamma^2 - \frac{4}{\pi^2}\beta^2\gamma^2 + i\frac{4}{\pi}\beta\gamma^2)t^n \\
& +(-\frac{8}{\pi^2}\beta^2\gamma^2 + i\frac{4}{\pi}\beta\gamma^2)t^n \log(t) - i\frac{2}{\pi}(2n+1)\beta\gamma t^{n-1} \\
& -i\frac{4}{\pi}n\beta\gamma t^{n-1}\log(t) - \frac{4}{\pi^2}\beta^2\gamma^2 t^n \log^2(t)\Bigg] e^{-\gamma\left(1+i\frac{2}{\pi}\beta\log(t)\right)t}. \qquad \text{(D.18)}
\end{aligned}
$$

The term $n(n-1)t^{n-2}$ is equal to zero when $n = 1$ and $h_n(t)$ is deduced from $g_n(t)$ by replacing β by $-\beta$. The functions $g_n(t)$, $h_n(t)$ are $\mathbb{L}^1(\mathbb{R}^+)$ functions and hence by Riemann-Lebesgue [117, p.3 sec.2 th.1] their $\mathbb{L}^1(\mathbb{R}^+)$ Fourier transforms are $o(1)$. Therefore equation (D.17) is $o(\frac{1}{x^2})$. Equivalently, $I_n(x) = o(\frac{1}{x^2})$ as $x \to \infty$ for all $n \geq 1$. Now using equation (D.13) we obtain:

$$
\begin{aligned}
& \lim_{x\to\infty} 2\pi x^2 |p_N(z)| \\
&= \lim_{x\to\infty} \left| \sum_{n=0}^{\infty} \frac{y^n}{n!} x^2 I_n(x) \right| \\
&= \lim_{x\to\infty} \left| 2\pi x^2 p_N(x) + \sum_{n=1}^{\infty} \frac{y^n}{n!} \left((-1)^{n+1} \int_0^{+\infty} e^{ixt} h_n(t)\, dt - \int_0^{+\infty} e^{-ixt} g_n(t)\, dt \right) \right| \\
&\leq \lim_{x\to\infty} 2\pi x^2 p_N(x) + \lim_{x\to\infty} \sum_{n=1}^{\infty} \frac{|y|^n}{n!} \left(\int_0^{+\infty} |h_n(t)|\, dt + \int_0^{+\infty} |g_n(t)|\, dt \right) \\
&= \lim_{x\to\infty} 2\pi x^2 p_N(x) + \sum_{n=1}^{\infty} \frac{|y|^n}{n!} \int_0^{+\infty} (|h_n(t)| + |g_n(t)|)\, dt \\
&= \lim_{x\to\infty} 2\pi x^2 p_N(x) + \int_0^{+\infty} \sum_{n=1}^{\infty} \frac{|y|^n}{n!} (|h_n(t)| + |g_n(t)|)\, dt \qquad \text{(D.19)}
\end{aligned}
$$

The interchange in (D.19) is valid since:

$$\begin{aligned}
&\sum_{n=1}^{\infty} \frac{|y|^n}{n!}(|h_n(t)| + |g_n(t)|) \\
&\leq \sum_{n=1}^{\infty} \frac{|y|^n}{n!} \left[A_1 n(n-1)t^{n-2} + A_2 t^n + A_3 t^n |\log(t)| + A_4(2n+1)t^{n-1}\right. \\
&\qquad\qquad\qquad\qquad \left. + A_5 n t^{n-1}|\log(t)| + A_6 t^n \log^2(t)\right] e^{-\gamma t} \\
&\leq e^{-\gamma t} \sum_{n=1}^{\infty} \frac{|y|^n}{n!} \left[A_1 n(n-1)t^{n-2} + (A_2 + A_3|\log(t)| + A_6 \log^2(t))t^n\right. \\
&\qquad\qquad\qquad\qquad \left. + n(3A_4 + A_5|\log(t)|)t^{n-1}\right] \qquad\qquad \text{(D.20)} \\
&\leq e^{-\gamma t} \left[A_1 y^2 e^{|y|t} + (A_2 + A_3|\log(t)| + A_6 \log^2(t))(e^{|y|t} - 1) + |y|(3A_4 + A_5)e^{|y|t}\right] \\
&\leq e^{-(\gamma - |y|)t} \left[A_2 + (3A_4 + A_5)|y| + A_1 y^2 + A_3|\log(t)| + A_6 \log^2(t)\right]
\end{aligned}$$

which is integrable on $[0, +\infty[$ since $|y| < \eta \,(< \gamma)$. The A_is, $1 \leq i \leq 6$ are positive constants function of β, γ and can be derived from the expression of $g_n(t)$ (equation D.18) and from that of $h_n(t)$ accordingly after taking the norm of each term in those expressions. To write equation (D.20), we used the obvious inequality $2n + 1 \leq 3n$ whenever $n \geq 1$. Back to (D.19),

$$\begin{aligned}
\lim_{x\to\infty} 2\pi x^2 |p_N(z)| &\leq \lim_{x\to\infty} 2\pi x^2 p_N(x) + \int_0^{+\infty} \sum_{n=1}^{\infty} \frac{|y|^n}{n!}(|h_n(t)| + |g_n(t)|)\, dt \\
&\leq \lim_{x\to\infty} 2\pi x^2 p_N(x) + \int_0^{+\infty} l(t)\, dt,
\end{aligned}$$

where $l(t) = e^{-(\gamma - |y|)t} \left[A_2 + (3A_4 + A_5)|y| + A_1 y^2 + A_3|\log(t)| + A_6 \log^2(t)\right]$. Since $\lim_{x\to\infty} 2\pi x^2 p_N(x)$ and $\int_0^{+\infty} l(t)\, dt$ are both finite and non zero when $|y| < \eta \,(< \gamma)$, then $0 \leq \lim\limits_{x\to\infty} 2\pi x^2 |p_N(z)| < \infty$ and $|p_N(z)| = O\left(\frac{1}{|\Re(z)|^2}\right)$ as $\Re(z) \to \infty$ whenever $z \in \mathcal{S}_\eta$.

D.3 Rate of Decay of the Derivative Functions of Alpha-Stable PDFs

It is already know that alpha-stable PDFs are infinitely differentiable (Property 2 in Chapter 2) and that their tail behaviour is $\Theta\left(\frac{1}{|x|^{\alpha+1}}\right)$ whenever $\beta \neq 1$ (Theorem 1 in Chapter 2). In what follows, we state a theorem concerning the tail behaviour of the derivatives of any order of these density functions.

Theorem 29 (Rate of Decay of the Derivative Functions of Alpha-Stable PDFs)**.** *Let $N \sim \mathcal{S}(\alpha, \beta, \gamma, \delta)$, $|\beta| \neq 1$, be a non-totally skewed alpha-stable RV and let $p_N(x)$ be its corresponding PDF. Denote by $p_N^{(n)}(x)$ the n-th derivative of $p_N(x)$. Then, for $n \geq 1$*

$$p_N^{(n)}(x) = \begin{cases} O\left(\dfrac{1}{|x|^{n+\alpha+1}}\right) & \alpha \neq 1 \\ o\left(\dfrac{1}{x^2}\right) & \alpha = 1, \beta \neq 0 \\ \Theta\left(\dfrac{1}{|x|^{n+2}}\right) & \alpha = 1, \beta = 0 \end{cases}$$

Proof. We assume WLOG that $\delta = 0$ since this corresponds to a translation of the PDF which does not affect the tail behaviour. The case $\alpha = 1$, $\beta = 0$ corresponds to the Cauchy density function and the result is straight-forward. In general, the derivative $p_N^{(n)}(x)$, $n \geq 1$ is given by:

$$p_N^{(n)}(x) = \frac{(-i)^n}{2\pi} \int_{-\infty}^{+\infty} t^n e^{-ixt - \gamma^\alpha [1 - i\beta \operatorname{sgn}(t) \Phi(t)] |t|^\alpha} \, dt. \tag{D.21}$$

It has been previously seen that equation (D.21) is identical to the function $\frac{1}{\gamma^{n+1}} T_n\left(-\frac{x}{\gamma}; \beta\right)$ when $\alpha \neq 1$ (see footnote in Section D.1) and to the function $I_n(x)$ when $\alpha = 1$ (see equation (D.14) in Section D.2). The tail behaviours of $T_n(x; \beta)$ and $I_n(x)$ established respectively in Section D.1 and D.2 yields the result of the theorem. We note that though the study in Section D.1 was restricted to the range $1 < \alpha < 2$, the adopted methodology to characterize the tail behaviour of $T_n\left(x;\beta\right)$ holds true for any $\alpha \neq 1$. $\square$

Appendix E

Evaluation of the function $f(x;\xi)$

We prove in this Appendix that for all $\xi \in \mathbb{R}^{+*}$, $x \in \mathbb{R}$,

$$f(x;\xi) = \frac{\xi}{\pi}\int_{-\infty}^{\infty} \ln(1+u^2)\,\frac{1}{1+(\xi u - x)^2}\,du = \ln\left(\left[\frac{\xi+1}{\xi}\right]^2 + \left[\frac{x}{\xi}\right]^2\right).$$

Proof. First, we consider the special case $x = 0$, $\xi = 1$ for which $f(0;1) = \frac{1}{\pi}\int_{-\infty}^{\infty} \ln(1+u^2)\,\frac{1}{1+u^2}\,du = \ln 4$ [90] and the above assertion is true. Whenever $\xi \neq 1$, taking the partial derivative of $f(x;\xi)$ with respect to x and integrating by parts yields,

$$\begin{aligned}
\frac{\partial f}{\partial x}(x;\xi) &= \frac{\partial}{\partial x}\left[\frac{\xi}{\pi}\int_{-\infty}^{\infty} \ln(1+u^2)\,\frac{1}{1+(\xi u - x)^2}\,du\right] \\
&= \frac{\xi}{\pi}\int_{-\infty}^{\infty} \ln(1+u^2)\,\frac{\partial}{\partial x}\left[\frac{1}{1+(\xi u - x)^2}\right] du \qquad \text{(E.1)} \\
&= -\frac{1}{\pi}\int_{-\infty}^{\infty} \ln(1+u^2)\,\frac{\partial}{\partial u}\left[\frac{1}{1+(\xi u - x)^2}\right] du \\
&= -\frac{1}{\pi}\left[\ln(1+u^2)\frac{1}{1+(\xi u - x)^2}\right]_{-\infty}^{+\infty} + \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{2u}{1+u^2}\frac{1}{1+(\xi u - x)^2}\,du \\
&= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{2u}{1+u^2}\frac{1}{1+(\xi u - x)^2}\,du \qquad \text{(E.2)} \\
&= \frac{1}{\pi}\int_{-\infty}^{\infty} \left[\frac{C_1 u + C_2}{1+u^2} + \frac{C_3 u + C_4}{1+(\xi u - x)^2}\right] du,
\end{aligned}$$

where

$$C_1 = \frac{2(1 + x^2 - \xi^2)}{(x^2 + (\xi - 1)^2)\,(x^2 + (\xi + 1)^2)}$$
$$C_2 = \frac{-4x\xi}{(x^2 + (\xi - 1)^2)\,(x^2 + (\xi + 1)^2)}$$
$$C_3 = -\xi^2 C_1 \qquad C_4 = -(x^2 + 1)C_2.$$

The interchange in (E.1) is justified by the fact that $\int_{-\infty}^{\infty} \ln(1+u^2)\,\frac{\partial}{\partial x}\left[\frac{1}{1+(\xi u - x)^2}\right] du$ is absolutely convergent.

We note that whenever for a real integrable function $r(\cdot)$ the integral $\int_{-\infty}^{\infty} r(x)\,dx$ exists, then $\int_{-\infty}^{\infty} r(x)\,dx = \lim_{l\to\infty} \int_{-l}^{l} r(x)\,dx$. In what follows we will implicitly assume the preceding equality in our evaluations whenever the case. For instance, it is the case for $\frac{\partial f}{\partial x}(x;\xi)$ since the integral exists for all $\xi \neq 1$, $x \in \mathbb{R}$ as seen in (E.2). Hence,

$$\begin{aligned}
\frac{\partial f}{\partial x}(x;\xi) &= \frac{1}{\pi}\left[\frac{C_1}{2}\ln(1+u^2) + \frac{C_3}{2\xi^2}\ln\left(1 + (\xi u - x)^2\right)\right.\\
&\quad \left. + C_2 \arctan u + \left(\frac{x}{\xi^2}C_3 + \frac{1}{\xi}C_4\right)\arctan(\xi u - x)\right]_{-\infty}^{+\infty}\\
&= C_2 - xC_1 + \frac{1}{\xi}\,C_4 = \frac{2x}{x^2 + (\xi+1)^2}.\\
\Leftrightarrow \quad f(x;\xi) &= \ln\left(x^2 + (\xi+1)^2\right) + \beta(\xi). \qquad \text{(E.3)}
\end{aligned}$$

Now taking the partial derivative with respect to ξ:

$$\frac{\partial f}{\partial \xi}(x;\xi) = \frac{\partial}{\partial \xi}\left[\frac{\xi}{\pi}\int_{-\infty}^{\infty}\frac{1}{1+(\xi u - x)^2}\ \ln(1+u^2)\,du\right]$$
$$= \frac{1}{\xi}f(x;\xi) + \frac{\xi}{\pi}\int_{-\infty}^{\infty}\ln(1+u^2)\,\frac{\partial}{\partial \xi}\left[\frac{1}{1+(\xi u - x)^2}\right]du$$
$$= \frac{1}{\xi}f(x;\xi) + \frac{1}{\pi}\int_{-\infty}^{\infty}u\ln(1+u^2)\frac{\partial}{\partial u}\left[\frac{1}{1+(\xi u - x)^2}\right]du$$
$$= \frac{1}{\xi}f(x;\xi) + \frac{1}{\pi}\left[u\ \ln(1+u^2)\frac{1}{1+(\xi u - x)^2}\right]_{-\infty}^{+\infty}$$
$$-\frac{1}{\pi}\int_{-\infty}^{\infty}\left[\ln(1+u^2)+\frac{2u^2}{1+u^2}\right]\frac{1}{1+(\xi u - x)^2}\,du$$
$$= -\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{2u^2}{1+u^2}\frac{1}{1+(\xi u - x)^2}\,du$$
$$= -\frac{1}{\pi}\int_{-\infty}^{\infty}\left[\frac{C_5 u + C_6}{1+u^2}+\frac{C_7 u + C_8}{1+(\xi u - x)^2}\right]du,$$

where

$$C_5 = \frac{-4x\xi}{(x^2+(\xi-1)^2)\,(x^2+(\xi+1)^2)}$$
$$C_6 = \frac{-2(1+x^2-\xi^2)}{(x^2+(\xi-1)^2)\,(x^2+(\xi+1)^2)}$$
$$C_7 = -\xi^2 C_5 \qquad C_8 = -(x^2+1)C_6,$$

which yields

$$\frac{\partial f}{\partial \xi}(x;\xi) = -\frac{1}{\pi}\left[\frac{C_5}{2}\ln(1+u^2) + C_6\arctan u\right.$$
$$\left.+\left[\frac{xC_7}{\xi^2}+\frac{C_8}{\xi}\right]\arctan(\xi u - x) + \frac{C_7}{2\xi^2}\ln\left[1+(\xi u - x)^2\right]\right]_{-\infty}^{+\infty}$$
$$= -\left(C_6 - xC_5 + \frac{1}{\xi}C_8\right)$$
$$= \frac{-2(x^2+1)^2 + 2(x^2+1)\xi + 2(1-x^2)\xi^2 - 2\xi^3}{\xi\,(x^2+(\xi-1)^2)\,(x^2+(\xi+1)^2)}. \tag{E.4}$$

On the other hand, using (E.3) we obtain

$$\frac{\partial f}{\partial \xi}(x;\xi) = \frac{2(\xi+1)}{x^2+(\xi+1)^2} + \beta^{'}(\xi),$$

which yields along with (E.4)

$$\beta^{'}(\xi) = -\frac{2}{\xi} \Longrightarrow \beta(\xi) = -2\ln\xi + C_9$$
$$\Longrightarrow f(x;\xi) = \ln\left(\left[\frac{\xi+1}{\xi}\right]^2 + \left[\frac{x}{\xi}\right]^2\right) + C_9, \tag{E.5}$$

which is valid $\forall\, x \in \mathbb{R}$, $\xi \in \mathbb{R}^{+*} \setminus \{1\}$. In order to find the constant term C_9, we restate (4.20)

$$f(x;\xi) = \mathsf{E}_N\left[\ln\left(1 + \left[\frac{x+N}{\xi}\right]^2\right)\right].$$

Since $f(x;\xi)$ is continuous in $\xi = 1$ by virtue of MCT, by taking the limit as $\xi \to 1$, equation (E.5) gives,

$$f(0;1) = \lim_{x\to 1} \ln\left(\left[\frac{\xi+1}{\xi}\right]^2 + \left[\frac{x}{\xi}\right]^2\right) + C_9 = \ln 4 + C_9.$$

Finally, since $f(0;1) = \mathsf{E}_N\left[\ln\left(1+N^2\right)\right] = \ln 4$, equating the two equations gives $C_9 = 0$. □

Appendix F

Sufficient Conditions for Existence of $J_\alpha(X + \sqrt[\alpha]{\eta}N)$

In his technical report [98, sec. 6], Barron states that the de Bruijn's identity for Gaussian perturbations (equation (5.4)) holds for any RV having a finite variance. In this appendix, we follow Barron's steps as we prove the existence of $J_\alpha\left(X + \sqrt[\alpha]{\eta}N\right)$, $\eta > 0$ for any RV $X \in \mathcal{L}$ where

$$\mathcal{L} = \left\{ \text{RVs } U : \int \ln\left(1 + |U|\right)\, dF_U(u) \text{ is finite } \right\},$$

and where $N \sim \mathcal{S}(\alpha; 1)$ is independent of X, $0 < \alpha < 2$. According to the definition, $J_\alpha(X + \sqrt[\alpha]{\eta}N)$ is the derivative of the entropy with respect to the dispersion η of the added stable variable. Therefore, the problem boils down to proving differentiability of $h(X + \sqrt[\alpha]{\eta}N)$.

First let $q_\eta(y) = \mathrm{E}\left[p_\eta(y - X)\right]$ be the PDF of $Y = X + \sqrt[\alpha]{\eta}N$ where $p_\eta(\cdot)$ is the density of the alpha-stable variable with dispersion η. Note that since $p_\eta(\cdot)$ is bounded then so is $q_\eta(\cdot)$ and since $X \in \mathcal{L}$ then so is Y. Then $h(Y)$ is finite and is defined as

$$h(Y) = -\int q_\eta(y) \ln q_\eta(y)\, dy.$$

We list and prove next two technical lemmas.

Lemma 11 (Technical Result)**.**

$$\frac{d}{d\eta}q_\eta(y) = E\left[\frac{d}{d\eta}p_\eta(y-X)\right]$$

Lemma 12 (Existence of the Generalized Fisher Information)**.**

$$\frac{d}{d\eta}h(X+\sqrt[\alpha]{\eta}N) = -\int \frac{d}{d\eta}\left(q_\eta(y)\right)\ln q_\eta(y)\,dy$$

exists and is finite. Also,

$$J_\alpha(X+\sqrt[\alpha]{\eta}N) = -\int \frac{d}{d\eta}\left(q_\eta(y)\right)\ln q_\eta(y)\,dy.$$

Proof. We start by proving lemma 11. The interchange holds whenever $|\frac{d}{d\eta}p_\eta(t)|$ is bounded uniformly by an integrable function in a neighbourhood of η by virtue of the MVT and the Lebesgue DCT. To prove boundedness, we start by evaluating the derivative. Since

$$p_\eta(t) = \frac{1}{\sqrt[\alpha]{\eta}}p_N\left(\frac{t}{\sqrt[\alpha]{\eta}}\right),$$

then

$$\frac{d}{d\eta}p_\eta(t) = -\frac{1}{\alpha}\frac{1}{\eta^{1+\frac{1}{\alpha}}}p_N\left(\frac{t}{\sqrt[\alpha]{\eta}}\right) - \frac{1}{\alpha}\frac{t}{\eta^{1+\frac{2}{\alpha}}}\frac{dp_N}{d\eta}\left(\frac{t}{\sqrt[\alpha]{\eta}}\right),$$

which gives

$$\left|\frac{dp_\eta}{d\eta}(t)\right| \le \frac{1}{\alpha}\frac{1}{\eta^{1+\frac{1}{\alpha}}}p_N\left(\frac{t}{\sqrt[\alpha]{\eta}}\right) + \frac{1}{\alpha}\frac{|t|}{\eta^{1+\frac{2}{\alpha}}}\left|\frac{dp_N}{du}\right|_{u=\frac{t}{\sqrt[\alpha]{\eta}}}. \tag{F.1}$$

For the purpose of finding the uniform bound on the derivative, we define b as a positive number chosen such that $b < \eta < 2b$. Concerning the first term of the bound in (F.1), we consider two separate ranges of the variable t to find the uniform upperbound . On compact sets, we have

$$\frac{1}{\alpha}\frac{1}{\eta^{1+\frac{1}{\alpha}}}p_N\left(\frac{t}{\sqrt[\alpha]{\eta}}\right) \leq \frac{1}{\alpha}\frac{1}{b^{1+\frac{1}{\alpha}}}\max_{u\in\mathbb{R}} p_N(u) \tag{F.2}$$

where the maximum exists since alpha-stable variables are unimodal [71] and thus their PDF is upperbounded. As for large values of $|t|$, we use the fact that there exists some $k > 0$ such that $p_N(t) \leq k\frac{1}{|t|^{1+\alpha}}$ [71] which gives

$$\frac{1}{\alpha}\frac{1}{\eta^{1+\frac{1}{\alpha}}}p_N\left(\frac{t}{\sqrt[\alpha]{\eta}}\right) \leq \frac{k}{\alpha}\frac{1}{|t|^{1+\alpha}}, \tag{F.3}$$

an integrable upperbound independent of η. Equations (F.2) and (F.3) insures that the first term of the right-hand side (RHS) of equation (F.1) is uniformly upperbounded by an integrable function for $b < \eta < 2b$. When it comes to the second term of the RHS of (F.1), we have for $n \geq 0$ (see [15, p.183])

$$\frac{d^n p_N}{du^n}(u) = \frac{(-i)^n}{2\pi}\int \omega^n \phi_N(\omega) e^{-i\omega u}\, d\omega, \tag{F.4}$$

and

$$\left|\frac{d^n p_N}{du^n}(u)\right| \leq \frac{1}{\pi\alpha}\Gamma\left(\frac{n+1}{\alpha}\right) \tag{F.5}$$

where $\phi_N(\omega) = e^{-|\omega|^\alpha}$ is the characteristic function of $\mathcal{S}(\alpha; 1)$. Hence, on compact sets, equation (F.5) gives a uniform integrable upperbound on the second term of the RHS of the form

$$\frac{1}{\alpha}\frac{|t|}{\eta^{1+\frac{2}{\alpha}}}\left|\frac{dp_N}{du}\right|_{u=\frac{t}{\sqrt[\alpha]{\eta}}} \leq \frac{1}{\pi\alpha^2}\frac{|t|}{b^{1+\frac{2}{\alpha}}}\Gamma\left(\frac{2}{\alpha}\right), \tag{F.6}$$

which is integrable and independent of η. Therefore, we only consider next the integral term in equation (F.4) at large values of u. To this end, we make use of the results proven in Appendix D. It was shown in this appendix (Section D.3 Theorem 29) that $\frac{d^n p_U}{du^n}(u) = O\left(\frac{1}{|u|^{n+\alpha+1}}\right)$ when $\alpha \neq 1$, $|\beta| \neq 1$. When $\alpha = 1$, the symmetric alpha-stabled variable is Cauchy distributed and it is clear that $\frac{d^n p_U}{du^n}(u) = \Theta\left(\frac{1}{|u|^{n+2}}\right)$. Since $N \sim \mathcal{S}(\alpha, 1)$, then for $0 < \alpha < 2$

$$\left|\frac{d^n p_N}{du^n}(u)\right| = \frac{1}{2\pi}\left|T_n(-u;0)\right| \leq \frac{\kappa_n}{|u|^{n+\alpha+1}}$$

and

$$\frac{1}{\alpha}\frac{|t|}{\eta^{1+\frac{2}{\alpha}}}\left|\frac{dp_N}{du}\right|_{u=\frac{t}{\sqrt[\alpha]{\eta}}} \leq \frac{1}{\alpha}\frac{\kappa_1}{|t|^{1+\alpha}} \tag{F.7}$$

is uniformly bounded at large values of $|t|$ by an integrable function. Equations (F.6) and (F.7) imply that the second term in the RHS of equation (F.1) is uniformly upperbounded by an integrable function for $b < \eta < 2b$. This proves Lemma 11.

When it comes to Lemma 12, we have the following:

$$\begin{aligned}
\frac{d}{d\eta}h(Y) &= -\int \frac{d}{d\eta}\left(q_\eta(y)\ln q_\eta(y)\right)\,dy && \text{(F.8)}\\
&= -\int \frac{dq_\eta}{d\eta}(y)\ln q_\eta(y)\,dy - \int \frac{dq_\eta}{d\eta}(y)\,dy \\
&= -\int \frac{dq_\eta}{d\eta}(y)\ln q_\eta(y)\,dy - \frac{d}{d\eta}\int q_\eta(y)\,dy && \text{(F.9)}\\
&= -\int \frac{dq_\eta}{d\eta}(y)\ln q_\eta(y)\,dy. && \text{(F.10)}
\end{aligned}$$

Equation (F.10) is true since $q_\eta(y)$ is a PDF and integrates to 1. Next, we start by justifying equation (F.9). In fact,

$$\begin{aligned}
\left|\frac{dq_\eta}{d\eta}(y)\right| &= \left|\mathrm{E}\left[\frac{dp_\eta}{d\eta}(y-X)\right]\right| \\
&\leq \mathrm{E}\left|\frac{dp_\eta}{d\eta}(y-X)\right| \\
&\leq r_b(y),
\end{aligned}$$

where the first equation is due to Lemma 11 and the second is justified by the fact that the absolute value function is convex. When it comes to the last equation, it has been shown in the proof of Lemma 11 that $\left|\frac{dp_\eta}{d\eta}(t)\right|$ is uniformly upper-bounded in a neighbourhood of η by an integrable function $s_b(t)$. Note that the upperbound can be written as follows by virtue of equations (F.1), (F.2), (F.3), (F.6) and (F.7):

$$s_b(t) = \begin{cases} A(b) + B(b)|t| & |t| \leq t_0 \\ C\, p_N(t) & |t| \geq t_0, \end{cases} \tag{F.11}$$

where $A(b)$, $B(b)$, C and t_0 are some positive values chosen in order to write the bound. Then

$$\mathrm{E}\left|\frac{dp_\eta}{d\eta}(y-X)\right| \leq \mathrm{E}\left|s_b(y-X)\right| = r_b(y),$$

which is integrable since $s_b(t)$ is integrable and by using Fubini's theorem. This completes the justification of equation (F.9). As for equation (F.8), instead of finding a uniform integrable upperbound to $\frac{d}{d\eta}\left(q_\eta(y)\ln q_\eta(y)\right)$, an equivalent task is to find such one to $\frac{dq_\eta(y)}{d\eta}\ln q_\eta(y)$ which we show next. Since $p_N(t) = \Theta\left(\frac{1}{|t|^{\alpha+1}}\right)$ (see for example [71]), there exist positive T and K such that $p_N(t)$ is greater than $K\frac{1}{|t|^{\alpha+1}}$ for some K whenever $|t| \geq T$. Now let $y > 0$ be any scalar is large enough and define $\tilde{y} > 0$ such that $\Pr(|X| \leq \tilde{y}) \geq \frac{1}{2}$. Then

$$\begin{aligned} q_\eta(y) &= \frac{1}{\sqrt[\alpha]{\eta}}\int p_N\left(\frac{y-u}{\sqrt[\alpha]{\eta}}\right) dF_X(u) \\ &\geq \frac{1}{\sqrt[\alpha]{\eta}}\int_{-\tilde{y}}^{+\tilde{y}} p_N\left(\frac{y-u}{\sqrt[\alpha]{\eta}}\right) dF_X(u) \\ &\geq \frac{1}{2\sqrt[\alpha]{\eta}} p_N\left(\frac{y+\tilde{y}}{\sqrt[\alpha]{\eta}}\right) \\ &\geq \frac{1}{2\sqrt[\alpha]{2b}} p_N\left(\frac{y+\tilde{y}}{\sqrt[\alpha]{b}}\right) \\ &\geq \frac{bK}{2\sqrt[\alpha]{2}|y+\tilde{y}|^{\alpha+1}} \\ &\geq \frac{b\tilde{K}}{|y|^{\alpha+1}}, \end{aligned}$$

where $b < \eta < 2b$ and $\tilde{K}$ is some positive constant. A similar derivation may be carried for the case $y \leq -T$ large enough. Now, since at large values of $|y|$, $q_\eta(y) \leq 1$, then $|\ln q_\eta(y)| \leq \ln\left(\frac{|y|^{\alpha+1}}{b\tilde{K}}\right)$. Furthermore since $q_\eta(y)$ is continuous and positive, then it achieves a positive minimum on compact subsets of $\mathbb{R}$. Let $y_0 > 0$ be large enough, then on $|y| \leq y_0$, we have

$$\left|\frac{dq_\eta(y)}{d\eta}\ln q_\eta(y)\right| \leq \max_{y\in\mathbb{R}} r_b(y)\left|\ln \min_{|y|\leq y_0} q_\eta(y)\right| \tag{F.12}$$

$$\leq \max_{y\in\mathbb{R}} s_b(y)\left|\ln \min_{|y|\leq y_0} p_\eta(y)\right| \tag{F.13}$$

$$\leq \max_{y\in\mathbb{R}} s_b(y)\left|\ln \frac{1}{\sqrt[\alpha]{\eta}} p_N\left(\frac{y_1}{\sqrt[\alpha]{\eta}}\right)\right|$$

$$\leq \max_{y\in\mathbb{R}} s_b(y)\left|\ln \frac{1}{\sqrt[\alpha]{2b}} p_N\left(\frac{y_1}{\sqrt[\alpha]{b}}\right)\right| < \infty$$

which is independent of η. We choose y_0 large enough in order to guarantee that $\min_{|y|\leq y_0} q_\eta(y) \leq 1$ and that $\max_{|y|\leq y_0} |\ln q_\eta(y)| \leq \left|\ln \min_{|y|\leq y_0} q_\eta(y)\right|$. This justifies equations (F.12). The same reasoning applies to the justification of equation (F.13) by virtue of the fact that $\min_{|y|\leq y_0} p_\eta(y) \leq \min_{|y|\leq y_0} q_\eta(y)$ since $q_\eta(y) = \mathrm{E}\left[p_\eta(y - X)\right]$. Now for $|y| > y_0$, we have

$$\left|\frac{dq_\eta(y)}{d\eta}\ln q_\eta(y)\right| \leq r_b(y)\left(\ln \frac{|y|^{\alpha+1}}{b\tilde{K}}\right)$$

which is a uniform integrable upperbound. The integrability is justified since:

$$\begin{aligned}
&\int \ln\left(1+|y|\right) r_b(y)\, dy && \text{(F.14)}\\
&= \int\int \ln\left(1+|y|\right) s_b(y-x)\, dF_X(x)\, dy \\
&= \int\int \ln\left(1+|y|\right) s_b(y-x)\, dy\, dF_X(x) && \text{(F.15)}\\
&\leq \int\int \left(\ln(1+|x|) + \ln(1+|y|)\right) s_b(y)\, dy\, dF_X(x) \\
&= S_b \int \ln(1+|x|) dF_X(x) + L_b \\
&< \infty, && \text{(F.16)}
\end{aligned}$$

where

$$S_b = \int s_b(y)\, dy < \infty,$$

and

$$L_b = \int \ln(1+|y|) s_b(y)\, dy < \infty.$$

Note that S_b and L_b are finite by virtue of (F.11). Equation (F.15) is due to Fubini and equation (F.16) is justified by the fact that $X \in \mathcal{L}$. By this, equation (??) is true and Lemma 12 is proven. □

Appendix G

Multivariate Alpha-Stable Distributions, Riesz Potentials and Hypergeometric Fucntions

G.1 Multivariate Alpha-Stable Distributions

Definition 8 (Sub-Gaussian SαS). [18, p.78 Definition 2.5.1]

Let $0 < \alpha < 2$ and let $\mathcal{A} \sim \mathcal{S}\left(\frac{\alpha}{2}, 1, \left(\cos\left(\frac{\pi\alpha}{4}\right)\right)^{\frac{2}{\alpha}}, 0\right)$ be a totally skewed one sided alpha-stable distribution. Define $\mathbf{G} = (G_1, \cdots, G_d)$ be a zero mean Gaussian vector in $\mathbb{R}^d$. Then the random vector $(A^{\frac{1}{2}}G_1, \cdots, A^{\frac{1}{2}}G_d)$ is called a sub-Gaussian SαS random vector in $\mathbb{R}^d$ with underlying vector $\mathbf{G}$. In particular, each component $A^{\frac{1}{2}}G_i$, $1 \leq i \leq d$ is a symmetric alpha-stable variable with characteristic exponent α.

Proposition 1. *[18, p.79 Proposition 2.5.5]*

Let $\mathbf{N} = (N_1, \cdots, N_d)$ be a sub-Gaussian SαS with an underlying Gaussian vector having IID zero-mean components with variance $2\gamma^2$, $\gamma > 0$. Then,

$$\phi_{\mathbf{N}}(\boldsymbol{\omega}) = e^{-\gamma^\alpha|\boldsymbol{\omega}|^\alpha}. \tag{G.1}$$

The RVs N_is, $1 \leq i \leq d$, are dependent and each distributed according to $\mathcal{S}(\alpha, \gamma)$.

In this dissertation, we uniquely use stable vectors such as in Theorem 1, that we refer to as sub-Gaussian SαS and will be denoted by $\boldsymbol{\mathcal{S}}(\alpha, \gamma)$. We note that in the scalar case, $\boldsymbol{\mathcal{S}}(\alpha, \gamma)$ is identically distributed to $\mathcal{S}(\alpha, \gamma)$.

G.2 Riesz Potentials

Definition 9 (Riesz Potentials)**.** [101, p.117 Section 1] Let $0 < \nu < 1$. The Riesz potential $I_\nu(f)(x)$ for a sufficiently smooth $f : \mathbb{R}^d \to \mathbb{R}$ having a sufficient decay at ∞ is given by:

$$I_\nu(f)(\mathbf{x}) = \frac{1}{\kappa(\nu)} \int_{\mathbb{R}^d} \|\mathbf{x} - \mathbf{y}\|^{-d+\nu} f(\mathbf{y})\, d\mathbf{y}, \tag{G.2}$$

where $\kappa(\nu) = \pi^{\frac{d}{2}} 2^\nu \frac{\Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{d}{2} - \frac{\nu}{2}\right)}$.

Proposition 2. *Among other properties, $I_\nu(f)$ satisfies the following:*

- $\mathcal{F}\left(I_\nu(f)\right)(\boldsymbol{\omega}) = \|\boldsymbol{\omega}\|^{-\nu} \mathcal{F}(f(\mathbf{x}))(\boldsymbol{\omega})$ *in the distributional sense.*
- $I_0(f)(\mathbf{x}) \hat{=} \lim_{\nu \to 0} I_\nu(f)(\mathbf{x}) = f(\mathbf{x})$.
- *Whenever $\int |f|(\mathbf{x}) I_\nu(|g|)(\mathbf{x})\, d\mathbf{x}$ is finite, we have:*

$$\int f(\mathbf{x}) I_\nu(g)(\mathbf{x})\, d\mathbf{x} = \int I_\nu(f)(\mathbf{x}) g(\mathbf{x})\, d\mathbf{x}.$$

G.3 Hypergeometric Functions

We only consider in this appendix the Gauss hypergeometric function ${}_2F_1(a, b; c; z)$.

Definition 10 (Gauss Hypergeometric Functions)**.** The Gauss hypergeometric function is defined as the following power series:

$$ {}_2F_1(a, b; c; z) = \sum_{i=0}^{+\infty} \frac{(a)_i (b)_i}{(c)_i i!} z^i, $$

for $|z| < 1$ and generic parameters a,b,c. Outside of the unit circle $|z| < 1$, the function is defined as the analytic continuation of this sum with respect to z, with the parameters a, b and c held fixed. The notation $(d)_i$ is defined as:

$$ (d)_i = \begin{cases} 1 & i = 0 \\ d(d+1)\dots(d+i-1) & i > 0. \end{cases} $$

Proposition 3. *The Gauss hypergeometric function* ${}_2F_1(a, b; c; z)$ *satisfies the following property:*

$$ {}_2F_1(a, b; c; z) = (1-z)^{-a}_2 F_1 \left(a, c-b; c; \frac{z}{z-1} \right), $$

for $z \notin (1, +\infty)$.

Bibliography

[1] J. G. Gonzalez, *Robust Techniques for Wireless Communications in non-Gaussian Environments.* PhD thesis, University of Delaware, 1997.

[2] D. Middleton, "Statistical-physical models for electromagnetic interference," *Electromagnetic Compatibility, IEEE Transactions on*, vol. 19, pp. 106–127, Aug. 1977.

[3] D. Middleton, "Non-Gaussian noise models in signal processing for telecommunications: New methods and results for class A and class B noise models," *Information Theory, IEEE Transactions on*, vol. 45, pp. 1129–1149, May 1999.

[4] K. Gulati, B. L. Evans, J. G. Andrews, and K. R. Tinsley, "Statistics of co-channel interference in a field of Poisson and Poisson-Poisson clustered interferers," *Signal Processing, IEEE Transactions on*, vol. 58, pp. 6207–6222, Dec. 2010.

[5] R. S. Blum, R. J. Kozick, and B. M. Sadler, "An adaptive spatial diversity receiver for non-Gaussian interference and noise," *Signal Processing, IEEE Transactions on*, vol. 47, pp. 2100–2111, Aug. 1999.

[6] A. Nasri, and R. Schober, "Performance of BICM-SC and BICM-OFDM systems with diversity reception in non-Gaussian noise and interference," *Communications, IEEE Transactions on*, vol. 57, pp. 3316–3327, Nov. 2009.

[7] J. H. Miller, J. B. Thomas, "Detectors for discrete-time signals in non-Gaussian noise," *Communications, IEEE Transactions on*, vol. 18, pp. 241–250, Mar. 1972.

[8] M. P. Shinde and S. N. Gupta, "Signal detection in the presence of atmospheric noise in tropics," *Communications, IEEE Transactions on*, vol. 22, pp. 1055–1063, Aug. 1974.

[9] M. Bouvet and S. C. Schwartz, "Comparison of adaptive and robust receivers for signal detection in ambient underwater noise," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 37, pp. 621–626, May 1989.

[10] K. L. Blackard, T. S. Rappaport, and C. W. Bostian, "Radio frequency noise measurments and models for indoor wireless communications at 918 MHz, 2.44 GHz, AND 4.0 GHz," *ICC*, pp. 28–32, 1991.

[11] M. Shao and C. Nikias, "Signal processing with fractional lower order moments: stable processes and their applications," *Proceedings of the IEEE*, vol. 81, pp. 986 –1010, Jul. 1993.

[12] H. M. Hall, "A new model for "impulsive" phenomena: Application to atmospheric-noise communication channels," *Stanford Electronics Lab., Stanford University, Stanford California, Tech. Rep. 3412-8 and 7050-7*, Aug. 1966.

[13] S. A. Kassam, *Signal Detection in Non-Gaussian Noise.* Springer-Verlag, 1988.

[14] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. 2. Wiley, New York, 1966.

[15] B. V. Gnedenko and A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables.* Reading Massachusetts: Addison-Wesley Publishing Company, 1968.

[16] V. M. Zolotarev, *One-dimensional Stable Distributions*, vol. 65. American Mathematical Society, 1983.

[17] V. V. Uchaikin and V. M. Zolotarev, *CHANCE and STABILITY: Stable Distributions and their Applications.* Utrecht, Netherlands: VSP, 1999.

[18] G. Samoradnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance.* New York: Chapman and Hall, June 1994.

[19] P. Lévy, *Calcul des probabilités.* Gauthier-Villards, Paris, 1925.

[20] A. Ya. Khintchine, *Limit Theorems for Sums of Independent Random Variables.* Moscow-Leningrad: Gonti, 1938.

[21] B. Mandelbrot, *The Fractal Geometry of Nature.* New York: Freeman, 1983.

[22] B. W. Stuck and B. Kleiner, "A statistical analysis of telephone noise," *Bell Syst. Tech. J.*, vol. 53, no. 7, pp. 1263–1320, 1974.

[23] P. G. Georgiou, P. Tsakalides, and C. Kyriakakis, "Alpha-stable modeling of noise and robust time- delay estimation in the presence of impulsive noise," *Multimedia, IEEE Transactions on*, vol. 1, no. 3, pp. 291–301, 1999.

[24] E. S. Sousa, "Performance of a spread spectrum packet radio network link in a Poisson field of interferers," *Information Theory, IEEE Transactions on*, vol. 38, pp. 1743–1754, Nov. 1992.

[25] J. Ilow and D. Hatzinakos, "Analytic alpha-stable noise modeling in a Poisson field of interferers or scatterers," *Signal Processing, IEEE Transactions on*, vol. 46, pp. 1601–1611, Jun. 1998.

[26] B. L. Hughes, "Alpha-stable models of multiuser interference," in *IEEE International Symposium on Information Theory*, (Sorrento, Italy), 2000.

[27] M. Win, P. Pinto, and L. Shepp, "A mathematical theory of network interference and its applications," *Proceedings of the IEEE*, vol. 97, pp. 205 –230, feb. 2009.

[28] A. Chopra, *Modeling and mitigation of interference in wireless receivers with multiple antennas.* PhD thesis, University of Texas at Austin, December 2011.

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

[30] P. J. Huber, *Robust Statistics.* John Wiley & Sons, 1981.

[31] S. Ambike, J. Ilow, and D. Hatzinakos, "Detection for binary transmission in a mixture of Gaussian noise and impulsive noise modeled as an alpha-stable process," *IEEE Signal Processing Letters*, vol. 1, pp. 55–57, Mar. 1994.

[32] P. Tsakalides and C. L. Nikias, "Maximumm likelihood localization of sources in noise modeled as a stable process," *Signal Processing, IEEE Transactions on*, vol. 43, pp. 2700–2713, Nov. 1995.

[33] X. Ma and C. L. Nikias, "Joint estimation of time delay and frequency delay in impulsive noise using fractional lower order statistics," *Signal Processing, IEEE Transactions on*, vol. 44, pp. 2669–2687, Nov. 1996.

[34] G. A. Tsihrintzis and C. L. Nikias, "Performance of optimum and suboptimum receivers in the presence of impulsive noise modeled as an alpha-stable process," *Communications, IEEE Transactions on*, vol. 43, pp. 904–914, Feb./Mar./Apr. 1995.

[35] N. Beaulieu and D. Young, "Designing time-hopping ultrawide bandwidth receivers for multiuser interference environments," *Proceedings of the IEEE*, vol. 97, pp. 255 –284, Feb. 2009.

[36] M. Nassar, K. Gulati, A. Sujeeth, N. Aghasadeghi, B. Evans, and K. Tinsley, "Mitigating near-field interference in laptop embedded wireless transceivers," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pp. 1405 –1408, 2008.

[37] H. El Ghannudi, L. Clavier, N. Azzaoui, F. Septier, and P. Rolland, "Stable interference modeling and Cauchy receiver for an IR-UWB ad hoc network," *Communications, IEEE Transactions on*, vol. 58, pp. 1748 –1757, Jun. 2010.

[38] A. Rajan and C. Tepedelenlioglu, "Diversity combining over Rayleigh fading channels with symmetric alpha-stable noise," *Communications, IEEE Transactions on*, vol. 9, pp. 2968–2976, Sep. 2010.

[39] J. G. Gonzalez, D. W. Griffith, and G. R. Arce, "Zero-order statistics: A signal processing framework for very impulsive processes," *Proceedings of the IEEE*, pp. 254–258, 1997.

[40] T. C. Chuah, B. S. Sharif, and O. R. Hinton, "Robust signal processing for CDMA communications in non-Gaussian noise," *5th Int. Conf. Mathematics in Signal Process., Warwick, U.K.*, 2000.

[41] T. C. Chuah, "Robust iterative decoding of turbo codes in heavy-tailed noise," *Inst. Elect. Eng. Proc.—Commun.*, vol. 152, pp. 29–38, S 2005.

[42] J. G. Gonzalez, J. L. Paredes, and G. R. Arce, "Zero-order statistics: A mathematical framework for the processing and characterization of very impulsive signals," *Signal Processing, IEEE Transactions on*, vol. 54, pp. 3839–3851, Nov. 2006.

[43] Alfred S. Carasso, "The APEX method in image sharpening and the use of low exponent Levy stable laws," *Siam J. Appl. Math.*, vol. 63, no. 2, pp. 593–618, 2002.

[44] H. Belkacemi, S. Marcos, "Robust subspace-based algorithms for joint angle/Doppler estimation in non-Gaussian clutter," *Signal Processing*, vol. 87, pp. 1547–1558, Nov. 2007.

[45] L. A. Jaeckel, "Robust estimates of location: Symmetry and asymmetric contamination," *Ann. Math. Stat.*, vol. 42, pp. 1020–1034, 1971.

[46] S. A. Kassam, G. Moustakides, and J. G. Shin, "Robust Detection of Known Signals in Asymmetric Noise," *Signal Processing, IEEE Transactions on*, vol. 28, pp. 84–91, Jan. 1982.

[47] J. Wang, E. Kuruoglu, and T. Zhou, "Alpha-stable channel capacity," *Communications Letters, IEEE*, vol. 15, pp. 1107 –1109, Oct. 2011.

[48] H. Ben Maad, A. Goupil, L. Clavier and G. Gelle, "Asymptotic performance of LDPC codes in impulsive non-Gaussian channel," in *IEEE SPAWC*, 2010.

[49] J. Fahs, I. Abou-Faycal, "Using Hermite bases in studying capacity-achieving distributions over AWGN channels," *Information Theory, IEEE Transactions on*, vol. 58, August 2012.

[50] S. T. Mittnik, T. Doganoglu and D. Chenyao, "Computing the probablity density function of the stable paretian distribution," *Mathematical and Computer Modelling*, vol. 29, pp. 235–240, May–June 1999.

[51] D. Middleton and A. D. Spaulding, "A tutorial review of elements of weak signal detection in non-Gaussian EMI environments," NTIA Rep. 86-194, U.S. Dept. of Commerce, 1986.

[52] Y. Kim and G. T. Zhou, "The Middleton class B model and its mixture representation," Tech. Rep. CSIP TR-98-01, Center for Signal and Image Processing, Georgia Institute of Technology, Atlanta, GA, May 1998.

[53] S. G. Bobkov and G. P. Chistyakov, "Entropy power inequality for the renyi entropy," *IEEE Transactions on Information Theory*, vol. 61, pp. 708–714, February 2015.

[54] C. E. Shannon, "A mathematical theory of communication, part i," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 1948.

[55] C. E. Shannon, "A mathematical theory of communication, part ii," *Bell Syst. Tech. J.*, vol. 27, pp. 623–656, 1948.

[56] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* Wiley, 2nd ed., 2006.

[57] J. G. Smith, "The information capacity of peak and average power constrained scalar Gaussian channels," *Inform. Contr.*, vol. 18, pp. 203–219, 1971.

[58] S. Shamai and I. Bar-David, "The capacity of average and peak-power-limited quadrature Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 41, pp. 1060–1071, July 1995.

[59] I. Abou-Faycal, M. D. Trott and S. Shamai, "The capacity of discrete-time memoryless Rayleigh-fading channels," *Information Theory, IEEE Transactions on*, vol. 47, pp. 1290–1301, May 2001.

[60] C. Luo, *Communication for Wideband Fading Channels: on Theory and Practice*. PhD thesis, Massachusetts Institute of Technology, February 2006.

[61] M. Katz and S. Shamai, "On the capacity-achieving distribution of the discrete-time noncoherent and partially coherent AWGN channels," *Information Theory, IEEE Transactions on*, vol. 50, pp. 2257–2270, October 2004.

[62] L. Zhang and D. Guo, "Capacity of Gaussian channels with duty cycle and power constraints," in *IEEE International Symposium on Information Theory*, (Saint Petersburg, Russia), pp. 424–428, 2011.

[63] R. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, Nov. 1968.

[64] W. Hirt and J. Massey, "Capacity of the discrete-time Gaussian channel with intersymbol interference," *Information Theory, IEEE Transactions on*, vol. 34, pp. 380–388, May 1988.

[65] J. Fahs and I. Abou-Faycal, "On the detrimental effect of assuming a linear model for non-linear AWGN channels," in *IEEE International Symposium on Information Theory*, (Saint Petersburg, Russia), pp. 1693–1697, 2011.

[66] A. Tchamkerten, "On the discreteness of capacity-achieving distributions," *IEEE Trans. Inf. Theory*, vol. 50, pp. 2773–2778, November 2004.

[67] A. Das, "Capacity-achieving distributions for non-Gaussian additive noise channels," *in Proc. IEEE International Symposium on Information Theory*, p. 432, June 2000. Sorrento, Italy.

[68] J. Fahs, N. Ajeeb, and I. Abou-Faycal, "The capacity of average power constrained additive non-Gaussian noise channels," *in Proc. IEEE International Conference on Telecommunications*, April 2012. Beirut, Lebanon.

[69] Y. Tsai, C. Rose, R. Song, and I. S. Mian, "An additive exponential noise channel with a transmission deadline," in *IEEE International Symposium on Information Theory*, (Saint Petersburg, Russia), pp. 598–602, 2011.

[70] E. Lukacs, *Characteristics Functions.* Griffin London, 2nd ed., 1970.

[71] J. P. Nolan, *Stable Distributions - Models for Heavy Tailed Data.* Boston: Birkhauser, 2012. In progress, Chapter 1 online at academic2.american.edu/∼jpnolan.

[72] W. H. Lee, *Continuous and discrete properties of stochastic processes.* PhD thesis, University of Nottingham, July 2009.

[73] A. V. Skorohod, "On a theorem concerning stable distributions," *Selected translations in mathematical statistics and probability*, vol. 1, pp. 169–172, 1961.

[74] A. V. Skorohod, "Asymptotic formulas for stable distribution laws," *Selected translations in mathematical statistics and probability*, vol. 1, pp. 157–161, 1961.

[75] I. A. Ibragimov and Y. V. Linnik, *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff, Groningen: J.F.C. Kingman, 1971.

[76] A. J. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," *Information and Control*, vol. 2, no. 2, pp. 101–112, 1959.

[77] N. M. Blachman, "The convolution inequality for entropy powers," *IEEE Trans. Inf. Theory*, vol. 11, pp. 267–271, Apr. 1965.

[78] S. Artstein, K. M. Ball, F. Barthe, and A. Naor, "Solution of Shannon's problem on the monotonicity of entropy," *American Mathematical Society*, vol. 17, pp. 975–982, May 2004.

[79] E. H. Lieb, "Proof of an entropy conjecture of Wehrl," *Comm. Math. Phys.*, vol. 62, no. 1, pp. 35–41, 1978.

[80] I. Abou-Faycal, J. Fahs, "On the capacity of some deterministic non-linear channels subject to additive white gaussian noise," in *IEEE 17th International Conference on Telecommunications (ICT)*, (Doha, Qatar), pp. 63–70, April 2010.

[81] V. Anantharam and S. Verdu, "Bits through queues," *IEEE Transactions on Information Theory*, vol. 42, pp. 4–18, January 1996.

[82] O. Rioul, "Information theoretic proofs of entropy power inequality," *IEEE Trans. Inf. Theory*, vol. 57, pp. 33–55, January 2011.

[83] B. V. Bahr and C. Esseen, "Inequalities for the rth Absolute Moment of a Sum of Random Variables, $1 \leq r \leq 2$," *The Annals of Mathematical Statistics*, vol. 36, pp. 299–303, February 1965.

[84] A. N. Shiryaev, *Probability.* Springer-Verlag, 2nd ed., 1996.

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

[86] E. Agrell, "The channel capacity increases with power," *CoRR*, vol. abs/1108.0391, 2011.

[87] R. V. Churchill, J. W. Brown and R. F. Verhey, *Complex Variables and Applications.* McGraw-Hill, 3rd ed., 1976.

[88] J. Fahs, I. Abou-Faycal, "On the single-user capacity of some multiple access channels," in *The Eleventh International Symposium on Wireless Communication Systems*, (Barcelona, Spain), pp. 622–627, August, 26-29 2014.

[89] J. Fahs and I. Abou-Faycal, “On the capacity of additive white alpha-stable noise channels,” in *IEEE International Symposium on Information Theory*, (Cambridge, MA, USA), pp. 294–298, 1-6 July 2012.

[90] J. N. Kapur, *Maximum Entropy Models In Science And Engineering.* Wiley, 1989.

[91] J. Fahs, I. Abou-Faycal, “A Cauchy input achieves the capacity of a Cauchy channel under a logarithmic constraint,” in *IEEE International Symposium on Information Theory*, (Honolulu, HI, USA), pp. 3077–3081, June 29 - July 4 2014.

[92] A. Dembo, T. M. Cover and J. A. Thomas, “Information theoretic inequalities,” *Information Theory, IEEE Transactions on*, vol. 37, pp. 1501–1518, November 1991.

[93] L. Miclo, “Notes on the speed of entropic convergence in the central limit theorem,” *Progress in Probability*, vol. 56, pp. 129–156, 2003.

[94] G. Toscani, “Entropy inequalities for stable densities and strengthened central limit theorems.” arXiv:1512.05874v1, December 2015.

[95] H. Heyer, *Structural Aspects in the Theory of Probability: A Primer in Probabilities on Algebraic-Topological Structures*, vol. 7 of *Multivariate Analysis.* World Scientific, 2004.

[96] M. H. M. Costa, “A new entropy power inequality,” *IEEE Transactions on Information Theory*, vol. 31, pp. 751–760, Nov. 1985.

[97] S. Kullback, *Information Theory and Statistics.* New York: Dover, 1968.

[98] A. R. Barron, “Monotonic central limit theorem for densities,” Tech. Rep. 50, Stanford University, Stanford California, March 1984.

[99] R. Zamir, "A proof of the Fisher Information Inequality via a Data Processing Argument," *IEEE Transactions on Information Theory*, vol. 44, pp. 1246–1250, May 1998.

[100] W. A. Strauss, *Partial Differential Equation: An Introduction.* Wiley, second ed., 2008.

[101] Elias M. Stein, *Singular Integrals and Differentiability Properties of Functions.* New Jersey: Princeton University Press, third ed., 1970.

[102] I. Z. Ruzsa, "Sumsets and entropy," *Random Structures & Algorithms*, vol. 34, no. 1, pp. 1–10, 2009.

[103] T. Tao, "Sumset and inverse sumset theory for Shannon entropy," *Combinatorics, Probability and Computing*, vol. 19, pp. 603–639, 2010.

[104] I. Kontoyiannis and M. Madiman, "Sumset and inverse sumset inequalities for differential entropy and mutual information," *IEEE Transactions on Information Theory*, vol. 60, pp. 4503–4514, August 2014.

[105] M. Madiman, "On the entropy of sums," in *Proceedings of the 2008 IEEE Information Theory Workshop*, (Porto, Portugal), May 2008.

[106] T. M. Cover and Z. Zhang, ""on the maximum entropy of the sum of two dependent random variables"," *IEEE Transactions on Information Theory*, vol. 40, no. 4, pp. 1244–1246, 1994.

[107] E. Ordentlich, ""maximizing the entropy of a sum of independent bounded random variables"," *IEEE Transactions on Information Theory*, vol. 52, pp. 2176–2181, May 2006.

[108] S. Bobkov and M. Madiman, *"On the problem of reversibility of the entropy power inequality". Limit Theorems in Probability, Statistics and Number Theory, Festschrift in honor of F. Götze's 60th birthday, P. Eichelsbacher*

et al. Springer Proceedings in Mathematics and Statistics 42, Springer-Verlag, 2013.

[109] J. Fahs and I. Abou-Faycal, "A new tight upper bound on the entropy of sums," *Entropy*, vol. 17, no. 12, pp. 8312–8324, 2015.

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

[111] A. H. M. Godavarti, "Convergence of differential entropies," *IEEE Transactions on Information Theory*, vol. 50, pp. 1–6, January 2004.

[112] J. R. Munkres, *Topology.* Prentice Hall, 2nd ed., 2000.

[113] K. L. Chung, *A Course in Probability Theory.* San Diego: Academic Press, 3rd ed., 2001.

[114] J. Fahs and I. Abou-Faycal, "On the finiteness of the capacity of continuous channels," *IEEE Transactions on Communications*, vol. 64, pp. 166–173, Jan 2016.

[115] J. Fahs, I. Abou-Faycal, "On the finiteness and achievability of channel capacity," *http://arxiv.org/pdf/1504.03814v1.pdf*, 2015.

[116] J. E. Marsden and M. J. Hoffman, *Basic Complex Analysis.* W. H. Freeman and Company, 3rd ed., 1999.

[117] S. Bochner and K. Chandrasekharan, *Fourier Transforms.* Annals of mathematics studies, Princeton University Press, 1949.

[118] 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 [cs.IT]*, 2016.

[119] A. K. Basu, *Measure Theory and Probability.* PHI Learning, second ed., 2012.

[120] A. M. Zapala, "Unbounded mappings and weak convergence of measures," *Statistics and Probability Letters*, vol. 78, pp. 698–706, 2008.

[121] N. H. Bingham, C. M. Goldie, J. L. Teugels, *Regular Variation.* Cambridge University Press, 1989.

[122] N. Beaulieu, H. Shao, and J. Fiorina, "P-order metric uwb receiver structures with superior performance," *Communications, IEEE Transactions on*, vol. 56, pp. 1666–1676, October 2008.

[123] *A Simple IR-UWB Receiver Adapted to Multi-User Interferences*, Proceedings of Globecom 2006, 2006.

[124] G. Folland, *Real Analysis - Modern Techniques and Their Applications.* New York: Wiley-Interscience, 1984.

[125] M. R. Fortet and E. Mourier, "Convergence de la répartition empirique vers la répartition théorique," *Anales scientifiques de l'E.N.S*, vol. 70, no. 3, pp. 267–285, 1953.

[126] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes.* San Francisco, CA: Holden-Day, 1964.

[127] M. Shao and C. Nikias, "Signal processing with fractional lower order moments: stable processes and their applications," in *Proceedings of the IEEE*, vol. 81, pp. 986 –1010, July 1993.

[128] F. W. Steutel and K. V. Harn, *Infinite Divisibility of Probability Distributions on the Real Line.* Marcel Dekker Inc., 2006.

[129] S. Verdú, "On channel capacity per unit cost," *IEEE Transactions on Information Theory*, vol. 36, pp. 1019 – 1030, Sep. 1990.

[130] F. Luisier, T. Blu and M. Unser, "Image denoising in mixed poisson-gaussian noise," *IEEE Transactions on Image Processing*, vol. 20, pp. 696–708, March 2011.

[131] C. P. Niculescu and L. E. Persson", *Convex Functions and their Applications: A Contemporary Approach*. Springer-Verlag, 2006.

[132] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *Information Theory, IEEE Transactions on*, vol. 51, pp. 1261–1282, April 2005.

[133] D. Guo, S. Shamai, and S. Verdú, "Additive non-Gaussian noise channels: mutual information and conditional mean estimation," in *IEEE International Symposium on Information Theory*, (Adelaide, Australia), pp. 719–723, 2005.

[134] D. W. Griffith, J. G. Gonzales, and G. R. Arce, "Robust time-frequency representations for signals in alpha-stable noise," *Signal Processing, IEEE Transactions on*, vol. 44, pp. 2669–2687, Nov. 1996.

[135] A. I. Lapin, *On some properties of stable laws*. PhD thesis, Rutgers University, 1947.

[136] I. Kontoyiannis and M. Madiman, "Sumset Inequalities for Differential Entropy and Mutual Information," in *IEEE International Symposium on Information Theory*, (Cambridge, MA, USA), 2012.

[137] M. Madiman and A. Barron, "Generalized entropy power inequalities and monotonicity properties of information," *Information Theory, IEEE Transactions on*, vol. 53, pp. 2317–2328, Jul. 2007.

[138] G. Pólya and G. Szegö, *Problems and Theorems in Analysis I*. Springer-Verlag, 1978.

[139] W. Rudin, *Principles of Mathematical Analysis.* McGraw-Hill, 3rd ed., 1976.

[140] H. Silverman, *Complex Variables.* Houghton Mifflin Company, 1975.

[141] E. M. Stein and R. Shakarchi, *Complex Analysis.* Princeton University Press, April 2003.

[142] I. Abou-Faycal, M. D. Trott, and S. Shamai, "Correction to the capacity of discrete-time memoryless rayleigh-fading channels," *Information Theory, IEEE Transactions on*, vol. 47, pp. 1290–1301, May 2002.

[143] E. E. Kuruoglu, "Density parameter estimation of skewed α-stable distributions," *Signal Processing, IEEE Transactions on*, vol. 49, pp. 2192–2201, Oct. 2001.

[144] R. B. Ash, *Information Theory.* New York: Interscience Publishers, 1965.

[145] M. Loeve, *Probability Theory.* New York: Van Nostrand, 1955.

[146] P. A. Moran, *Introduction to Probability Theory.* Oxford: Clanrendon Press, 1968.

[147] A. W. Roberts and D. E. Varberg, *Convex Functions*, vol. 57. Academic Press, 1973.